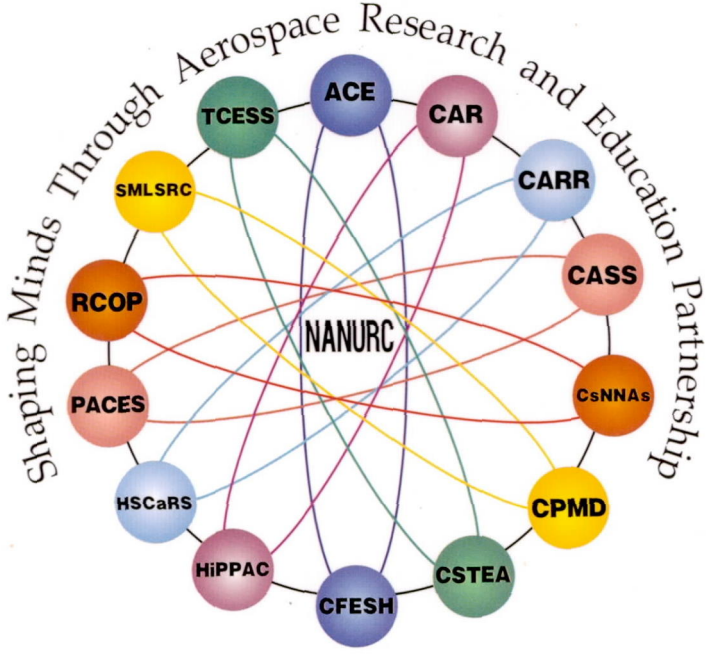
Sponsored by
Office of Equal Opportunity Programs
NASA Headquarters
Washington, DC

NASA
UNIVERSITY RESEARCH CENTERS

TECHNICAL ADVANCES IN AERONAUTICS, SPACE SCIENCES AND TECHNOLOGY, EARTH SYSTEMS SCIENCES, GLOBAL HYDROLOGY, AND EDUCATION

VOLUME III



National Alliance of NASA University Research Centers

EDITORS
T. L. Coleman, B. White, and S. Goodman

ASSOCIATE EDITORS
P. Sakimoto, L. Randolph and D. Rickman

Alabama A&M University
Center for Hydrology, Soil Climatology
and Remote Sensing

NASA UNIVERSITY RESEARCH CENTERS TECHNICAL ADVANCES IN AERONAUTICS, SPACE SCIENCES AND TECHNOLOGY, EARTH SYSTEMS SCIENCES, GLOBAL HYDROLOGY, AND EDUCATION

VOLUME III

URC-TC'98

NASA University Research Centers' Technical Conference on Technical Advances in Aeronautics, Space Sciences and Technology, *Earth Systems Sciences, Global Hydrology,* and Education

February 22 - 25, 1998
Huntsville Hilton Hotel
Von Braun Center - Huntsville, Alabama

Honorary Chairs
The Honorable Richard Shelby, R-Alabama
The Honorable Robert Cramer, D-Alabama
The Honorable Daniel Goldin, NASA - Administrator

General Chairs
Dr. John T. Gibson, President, Alabama A&M University
Ms. Carolyn Griner, Deputy Director, Marshall Space Flight Center (MSFC)
Mr. George Reese, NASA - Associate Administrator

Program Chairs
Ms. Bettie White, NASA Headquarters (HQ)
Dr. Tommy L. Coleman, Alabama A&M University

Program Co-chairs
Dr. Philip Sakimoto, NASA HQ
Mr. Willie Love, NASA MSFC
Dr. Doug Rickman, NASA MSFC

Members
Dr. M. Aggarwal, Alabama A&M University
Ms. Alicia Bean, NASA MSFC
Dr. G. Brown, Alabama A&M University
Dr. M. Busby, Tennessee State University
Mr. M. Estes, Universities Space Research Association (USRA)
Dr. F. Ferguson, North Carolina A&T University
Dr. R. Femandez-Scin, University of Puerto Rico
Mr. R. Fletcher, USDA Agricultural Research Service
Dr. M. Floyd, Alabama A&M University
Dr. T. Fogarty, Prarie View A&M University
Dr. E. Gooding, Alabama A&M University
Dr. R. Greenwood, *Global Hydrology and Climate Change* (GHCC) Center - Huntsville, AL
Dr. W. Hill, Tuskegee University
Dr. M. Jamshidi, University of New Mexico
Dr. J. Johnson, Florida A&M University
Ms. K. Jordan, K Corporation, Huntsville, AL
Dr. A. Manu, Alabama A&M University
Dr. E. Mintz, Clark Atlanta University
Dr. 0. Montgomery, Alabama A&M University
Dr. L. Randolph, NASA Headquarters
Dr. B. Reddy, Alabama A&M University
Mr. C. Scales, NASA MSFC
Dr. Z. Senwo, Alabama A&M University
Dr. A. Sharma, Alabama A&M University
Dr. E. Silberman, Fisk University
Dr. K. Sistani, Alabama A&M University
Dr. J. Shuford, Alabama A&M University
Dr. S. Starks, University of Texas at El Paso
Mr. W. Tadesse, Alabwna A&M University
Dr. R. Taylor, Alabama A&M University
Dr. D. Temple, Hampton University
Dr. M. Thierry-Palmer, Morehouse School of Medicine
Dr. A. Thorpe, Howard University
Dr. T. Tsegaye, Alabama A&M University
Dr. G. Wheelock, Alabama A&M University

Workshops Chair
Dr. A. Fahsi, Alabama A&M *University*

Conference Editors
Dr. T. Coleman, Alabama A&M University
Ms. B. White, NASA HQ
Dr. S. Goodman, NASA/MSFC

Associate Editors
Dr. C. Beyl, Alabama A&M University
Dr. W. Crosson, USRA GHCC
Dr. D. Perkey, USRA
Dr. P. Sakimoto, NASA HQ

Registration/Publicity Committee
Mr. H. Anthony, Alabama A&M University
Ms. P. Campbell, Alabama A&M University
Ms. M. Chandler, Alabama A&M University
Dr. G. DaBai, Alabama A&M University
Mr. M. *Estes,* Alabama A&M *University*
Ms. C. Hardin-, Alabama A&M University
Dr. A. Manu, Alabama A&M University
Ms. W. Porter, Alabama A&M University
Dr. M. Schamschula, Alabama A&M University
Mr. J. Surrency, Alabama A&M University
Ms. D. Rashid, Alabwna A&M University
Ms. C. Strother, Alabama A&M University
Mr. W. Tadesse, Alabama A&M University
Dr. T. Tsegaye, Alabama A&M University
Ms..V. Williams-Gaines, Alabama A&M University

FORWORD

This volume chronicles the proceedings of the 1998 NASA University Research Centers Technical Conference (URC-TC '98), held on February 22-25, 1998, in Huntsville, Alabama.

The University Research Centers (URCS) are multidisciplinary research units established by NASA at 1. I Historically Black Colleges or Universities (HBCU's) and 3 Other Minority Universities (OMU's) to conduct research work in areas of interest to NASA. The URC Technical Conferences bring together the faculty members and students from the URC's with representatives from other Universities, NASA, and the aerospace industry to discuss recent advances in their fields.

The achievements of the URC's during Academic Year 1996-97 and Summer 1997 were remarkable: 262 professional-level investigators and 549 students participated in URC research work; 451 refereed papers or book chapters were accepted or published, including 152 students as authors or co-authors; and 642 technical presentations were given at national and international conferences, including 210 presentations given by students. Perhaps most importantly, 190 degrees, including 106 bachelor's degrees, 76 master's degrees, and 8 doctoral degrees, were awarded to URC students, launching them into science and technology careers.

The URC program would not be possible without the continued strong support of the NASA Offices of Aeronautics and Space Transportation Technology, Life and Mircogravity Sciences and Applications, Mission to Planet Earth, Safety and Mission Assurance, Space Flight, and Space Science, as well as that of the NASA Field Installations. Their support is visible evidence of NASA's commitment to bringing minority institutions into the forefront of our agenda of exploration and discovery.

George E. Reese
Associate Administrator for
Equal Opportunity Programs
NASA Headquarters
Washington. DC

Bettie L. White
Director, Minority University
Research & Education Programs
NASA Headquarters
Washington, DC

NASA UNIVERSITY RESEARCH CENTERS (URC)
at Minority Institutions

HISTORICALLY BLACK COLLEGES AND UNIVERSITIES

Alabama A&M University
Center for Hydrology, Soil Climatology and Remote Sensing (HSCARS) Dr. Tommy Coleman, Director
http://hscarsl.saes.aamu.edu

Clark Atlanta University
High Performance Polymers and Composites Research Center (I-EPPAC)
Dr. Eric Mintz, Director

Fisk University
Center for Photonic Materials and Devices (CPMD) Dr. Enrique Silberman, Director
http://dubois.fisk.edu/~physics/nasa.html

Florida A&M **University**
Center for Nonlinear and Nonequilibrium Aerosciences (CeNNAs)
Dr. Joseph Johnson 111, Director
http://www.cennas.nhmfl.gov

Hampton University
Research Center for Optical Physics (RCOP)
Dr. Doyle Temple, Director
http://phy.hamptonu.edu/physics/rcop.html

Howard University
Center for the Study of Terrestrial & Extraterrestrial Atmospheres (CSTEA)
Dr. Arthur'lborpe, Director
http://www.cstea.howard.edu

Morehouse School Of Medicine
Space Medicine and Life Sciences Research Center (SMLSRC)
Dr. Myrtle Ilierry-Palmer, Director
http://www.msm.edu/nasa/nasa.htm

North Carolina A&T State University
Center for Aerospace Research (CAR)
Dr. Frederick Ferguson, Director
bttp://www.ncat.edu/~research/

Prairie View A&M University Center for Applied Radiation Research (CARR)
Dr. Thomas Fogany, Director
http://www.carr.pvamu.edu

Tennesee State University
Center for Automated Space Science (CASS)
Dr. Michael Busby, Director
http://coe.tnstate.edu

Tuskegee University
Center for Food and Environmental Systems for Human Exploration of Space (CFESH)
Dr. Walter Hill, Director
http://agriculture.tusk.edu/AgHe%20Website/Pages/NASA/CFESH.html

OTHER MINORITY UNIVERSITIES

University of New Mexico
Center for Autonomous Control Engineering (ACE)
Dr. Mo Jamshidi, Director
http://ace.unm.edu

University of Puerto Rico at Mayagiiez
Tropical Center for Earth and Space Studies (TCESS)
Dr. Rafael Femandez Sein, Director
http://exodo.upr.clu.edu/~tcess

University of Texas **at El Paso**
Pan-American Center for Earth and Environmental Studies (PACES)
Dr. Scott StarksDirector
http://paces.geo.utep.edu/~paces/index.html

For further information on NASA Minority University Research and Education Programs, please visit

http://mured.gsfc.nasa.gov

or contact a NASA Installation or Jet Propulsion Laboratory Program Coordinator:

Mr. Geoffrey Lee
Ames Research Center
Mail Code 204-2
Moffett Field, CA 94035
(415) 604-6406
FAX (415) 604-3869

Ms. Erma Cox
Dryden Flight Research Center
Mail Stop D-2160
P.O. Box 273
Edwards, CA 93523
(805) 258-3033
FAX (805) 258-2298

Mr. Dillard Menchan
Goddard Space Flight Center
Mail Code 120
Greenbel@ MD 20771
(301) 286-7348
FAX (301) 286-0298

W. Richard Ashe
Jet Propulsion Laboratory
Mail Stop 72-109
4800 Oak Grove Drive
Pasadena, CA 91109-8099
(818) 354-0122
FAX (818) 393-4977

Dr. Joseph D. Atkinson
Johnson Space Center
Mail Code AJ
Houston, TX 77058
(281) 4834831
FAX (281) 483-0609

Ms. Ember Smith
Kennedy **Space Center**
Mail Code AJ
KSC, FL 32899
(407) 867-8025
FAX (407) 867-1066

Dr. Samuel E. Massenberg
Langley **Research Center**
Mail Code 400
Hampton, VA 23665
(757) 864-3312
FAX (757) 864-8835

Mr. Roger Hathaway
Langley Research Center
Office of Education
Mail **Stop** 400
Hampton, VA 23681
(757) 864-3312
FAX (757) 864-8835

Dr. Julian Earls
Lewis Research Center
Mail Stop 0100
21000 Brookpark Road
Cleveland, OH 44135
(216) 433-3014
FAX (216) 433-5266
For HBCU's

Mr. Robert Lawrence
Lewis Research Center
Mail Stop 3-16
21 000 Brookpark Road
Cleveland, OH 44135-3191
(216) 433-2921
FAX (216) 433-2348
For USAR. GEM, and NAFEO Student Award

Ms. Jo Ann Charleston
Lewis Research Center
Mail Stop 7-4
2 1 000 Brookpark Road
Cleveland, OH 44135
(216) 433-2957
FAX (216) 433-8000
For OMUs

Mr. Willie Love
Marshall Space Flight Center
Mail Code CEOI
MSFC, AL 35812
(205) 544-0088
FAX (205) 544-2411

Dr. Armond Joyce
Stennis Space Center
Mail Code MAOO
Stennis Space Center, MS 39529-6000
(601) 688-3830
FAX (601) 688-7499

MAIN MENU

TABLE OF CONTENTS

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

PAGE 1 OF INDEX

Modeling the Effect of Vegetation on Passive Microwave Remote Sensing of Soil Moisture

Y.P. Liu[1], R. Inguva[2], W. L. Crosson[3], T.L. Coleman', C. Laymon[3], A. Fahsi[1]

'Alabama A&M University, Normal, AL 35762; 'Physics Department, University of Wyoming, Laramie, WY 82071 and Science and Technology Consultants, Huntsville, AL 35805; [3]Global I Iydrology and Climate Center.

Abstract

The effect of vegetation on passive microwave remote sensing of soil moisture is studied. The radiative transfer modeling work of Njoku and Kong (1977) is applied to a stratified medium of which the upper layer is treated as a layer of vegetation. An effective dielectric constant for this vegetation layer is computed using estimates of the dielectric constant of individual components of the vegetation layer. The horizontally-polarized brightness temperature is then computed as a function of the incidence angle. Model predictions are used to compare with the data obtained in the Huntsville'96, remote sensing of soil moisture experiment, and with predictions obtained using a correction procedure of Jackson and Schmugge (1991).

Introduction

Moisture content in the surface layers of the soil is of great importance in hydrological research. Precipitation and irrigation are partitioned into runoff and infiltration components, depending on the moisture content at the surface. Water that infiltrates through the soil surface layer recharges groundwater. The quantity of subsurface water resources, therefore, is greatly influenced by the moisture distribution at the soil surface. Research has shown that passive microwave remote sensing at low frequencies provides a direct measurement of surface soil moisture (Engman & Chauhan, 1995), The dielectric constant of the surface soil-water mixture, which is a function of soil moisture content, is related to microwave brightness temperature through the emissivity or reflectivity of the soil. One of the research goals of the Center for Hydrology, Soil Climatology and Remote Sensing (HSCaRS) at Alabama A&M University is to develop a computer model to predict soil moisture profile using remotely-sensed brightness temperature. The forward radiative transfer model of Njoku and Kong (1977) forms the basis of such a model. In this investigation, we conduct further tests of the model developed by Njoku and Kong (1977) by adding a vegetation layer overlying the soil to the soil profile.

The vegetation canopy has significantly effects on microwave brightness temperature. According to Wang et al. (1982), the emission from covered soils consists of three components: radiation from soils, upwelling radiation from the vegetation medium, and downwelling radiation from the vegetation reflected from the soil surface. Therefore, the presence of vegetation canopy attenuates brightness temperature radiation from the soil surface. The forward radiative transfer model of Njoku and Kong (1977) predicts brightness temperature for a soil surface using measured soil moisture and temperature profiles. To include effects of a vegetation canopy on brightness temperature, Jackson and Schmugge (199 l) developed a simplified vegetation correction to the Njoku and Kong model, by using the vegetation water content and a constant based on the relationship between optical thickness and vegetation water content. An attempt is made in this paper to include the effects of vegetation canopy on brightness temperature in Njoku's and Kong's model by adding a vegetation layer to the soil profile. Prediction of this model is compared with that of Njoku and Kong model with the vegetation correction of Jackson and Schmugge (1991) and with data from the Huntsville '96 experiment.

Model Description

(1) Vegetation Layer

The soil profile with a vegetation layer is considered as a stratified medium of N layers. The top layer consists of the vegetation canopy. The maximum height of this vegetation layer is denoted by d_v. This vegetation layer has an air interface above and a soil interface below. Each layer is characterized by its thickness, dielectric constant and temperature. Each layer is a heterogeneous mixture of several constituents treated as (Ulaby et al., 1986):

$$\varepsilon_R^{\alpha} = \sum_{\beta=1}^{N} f_{\beta}\, \varepsilon_{R\beta}^{\alpha} \qquad (1)$$

$$\varepsilon_I^\alpha = \sum_{\beta=1}^{N} f_\beta \varepsilon_{I\beta}^\alpha \quad (2)$$

$$\varepsilon = \varepsilon_R + i\varepsilon_I \quad (3)$$

where f_β is the volume fraction of the βth constituents, ε_R is the real part of the dielectric constant, ε_I is the imaginary part of the dielectric constant, and α is a power constant describing the relationship between the dielectric constant and volumetric water content. The vegetation layer is composed of air (a), vegetation mass (vm), and moisture content (vmw). Therefore, Eq. (1) to (3) are rewritten as (Ulaby et al., 1986):

$$\varepsilon_{RV}^\alpha = f_a + f_{vm}\varepsilon_{Rvm}^\alpha + f_{vwm}\varepsilon_{Rvmw}^\alpha \quad (4)$$

$$\varepsilon_{IV}^\alpha = f_{vm}\varepsilon_{Ivm}^\alpha + f_{vmw}\varepsilon_{Ivmw}^\alpha \quad (5)$$

$$\varepsilon_V = \varepsilon_{RV} + i\varepsilon_{IV} \quad (6)$$

The data for volume fractions are available from the Huntsville'96 field experiment.

(2) The Radiative Transfer Model

The horizontally-polarized brightness temperature, T_B, is computed using the stratified medium approach developed by Njoku and Kong (1 977). In this model, the brightness temperature is given by:

$$T_B = \frac{k}{\cos\theta}\int_{-\infty}^{0} T(z)dz\varepsilon_I(z)\left|\psi(z)\right|^2 \quad (7)$$

where $T(z)$ is the temperature at soil depth z, $\varepsilon_I(z)$ is the imaginary part of dielectric constant at soil depth z, k ($=2\pi/\lambda$) is the wave number, and θ is the look angle. The quantity of $\psi(z)$ satisfies the differential equations as follows:

$$\frac{d^2\psi(z)}{dz^2} + \left[\varepsilon(z)k^2 - k_x^2\right]\psi(z) = 0 \quad (8)$$

$$\left\{\frac{d\psi(z)}{dz} + ik\cos\theta\left[2 - \psi(z)\right]\right\}_{z=0} = 0 \quad (9)$$

where $k_x = k \sin 6?$ Detailed expressions for brightness temperature are given by Eq. (11) in Njoku and Kong (1977). This expression is used in this study after adding a vegetation layer.

(3) Dielectric Constant of Soil

The dielectric constant for a typical soil layer is defined as:

$$\varepsilon = \varepsilon_R + i\varepsilon_I \quad (10)$$

$$\varepsilon_R^\alpha = f_a + f_b\,\varepsilon_{Rb}^\alpha + f_s\,\varepsilon_{Rs}^\alpha + f_w\,\varepsilon_{Rw}^\alpha \quad (11)$$

$$\varepsilon_I^\alpha = f_b\,\varepsilon_{Ib}^\alpha + f_s\,\varepsilon_{Is}^\alpha + f_w\,\varepsilon_{Iw}^\alpha \quad (12)$$

$$f_a + f_b + f_s + f_w = 1 \quad (13)$$

where f is the volume fraction, subscribes a, b, s, and w designate air, bound water, soil solid, and soil water, respectively. The dielectric constant of the soil solids is (Ulaby et al., 1986):

$$\varepsilon_{Rs} = (1 + 0.44\rho_b)^2 \quad (14)$$

where ρ_b is the soil bulk density, while the imaginary part ε_{Rs} is typically <0.05. The dielectric constant of soil moisture is (Ulaby et al., 1986):

$$\varepsilon_{Rw} = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w\infty}}{1+\left(2\pi f \tau_w\right)^2} \tag{15}$$

$$\varepsilon_{Iw} = \frac{2\pi f \tau_w \left(\varepsilon_{w0} - \varepsilon_{w\infty}\right)}{1+\left(2\pi f \tau_w\right)^2} \tag{16}$$

$$2\pi\tau_w = 1.1\,109 \times 10^{-10} - 3.824 \times 10^{-12} T + 6.938 \times 10^{-14} T^2 - 5.096 \times 10^{-16} T^3 \tag{17}$$

$$\varepsilon_{w0} = 88.045 - 0.4147T + 6.295 \times 10^{-4} T' + 1.075 \times 10^{-5} T^3 \tag{18}$$

where ε_{w0} is the static dielectric constant of pure water, $\varepsilon_{w\infty}$ (=4.9) is the high frequency (or optical) limit of ε_w, f is the electromagnetic frequency, Hz, and T is the soil temperature. The dielectric constant of the bound water, ε_b, is (Dobson et al., 1985):

$$\varepsilon_b = \varepsilon_{Rb} + i\varepsilon_{Ib} = 35.0 + i14.5 \tag{19}$$

where ε_{Rb} and ε_{Ib} are real and imaginary part of ε_b. The volume fractions of air, free water, and soil solid are determined from the Huntsville'96 experimental measurements. The volume ii-action the bound water (f_b) is assumed to be 0.08.

(4) Jackson and Schmugge Vegetation Correction

Jackson and Schmugge (1991) studied vegetation effects on the microwave emission of soils from a different direction. They developed a vegetation correction procedure to estimate soil moisture under vegetation canopy by microwave remote sensing. By using the electromagnetic relationships of a two-layer (vegetation and soil layers) incoherent medium, they relate the brightness temperature above vegetation canopy to that on the soil surface by following equation:

$$T_{Bv} = T_{Bs} \cdot I - \Delta T_v \tag{20}$$

$$\Delta T = \left(T_{Bs} - T_s\right)\left(\gamma^2 - 1\right) \tag{21}$$

$$\gamma^2 = e^{-2\tau \sec(\theta)} \tag{22}$$

$$\tau = W_v b \tag{23}$$

where T_{Bv} *is* the brightness temperature of the ground with a vegetation layer, T_{Bs} is the brightness temperature of ground surface without considering a vegetation layer, T_s is the temperature of the ground surface, θ is the look angle, W, *is* the vegetation water content, and b is an empirical constant (the regression slope between vegetation water content and optical depth). Jackson and Schmugge correction indicates that the brightness temperature above the vegetation canopy is higher than that below the canopy.

Huntsville'96 Experiment

Huntsville'96, an experiment in microwave remote sensing of soil moisture, was conducted in Huntsville, AL from July 1-14, 1996. A research test-bed with four plots about 50 x 60 m was constructed at Alabama A&M University's Winfred Thomas Agricultural Research Station. Two plots were bare of vegetation (bare-smooth and bare rough plots) and two others were vegetated (grass and mixed-vegetation plots). Brightness temperatures of four plots were measured by L and S band radiometers installed in a boom truck. The remote sensing measurements were supported by *in situ* measurement of soil and vegetation properties, which include moisture, density, temperature, etc. Soil moisture and temperature were continuously measured by sensors in four pits in each plot at five depths to 90 cm. In addition, gravimetric soil moisture at soil surface was measured daily by the oven-dry method at random locations in each plot. Detailed information about the Huntsville'96 experiment can be found in the paper of Laymen et al. (1997).

Results and Discussion

Data used as model inputs were collected at the grass plot during the Huntsville'96 experiment (Laymen et al., 1997). The brightness temperature is first computed by the forward radiative transfer model without considering vegetation canopy effects (solid circles in Figure 1). Then, the vegetation canopy effect on the brightness temperature is added by applying Jackson and Schmugge correction to the forward radiative transfer model (open triangles in Figure 1). Figure 1 shows that correcting for the vegetation layer (W, = 0.3, b = 0.3 from Wang et al., 1982) has the effect of increasing the modeled brightness temperature above the canopy. This effect is greater at higher soil moisture values because the contrast between the physical temperature of the soil and its brightness temperature is also greater. Figure 2 shows a comparison among observed brightness temperatures (L-band, 1.413 GHz) and modeled brightness temperature above the canopy using the Jackson and Schmugge (199 1) correction procedure and explicitly modeling the vegetation as a discrete layer with our radiative transfer model (called Radiative Transfer Model or RTM correction in Figure 2). Neither vegetation correction procedure does a very good job of reproducing the observed brightness temperatures. Explicit treatment of the vegetation layer with the radiative transfer model, however, resulted in significantly more scatter.

Figure 1. The predicted brightness temperature by Njoku and Kong model with and without a vegetation correction.

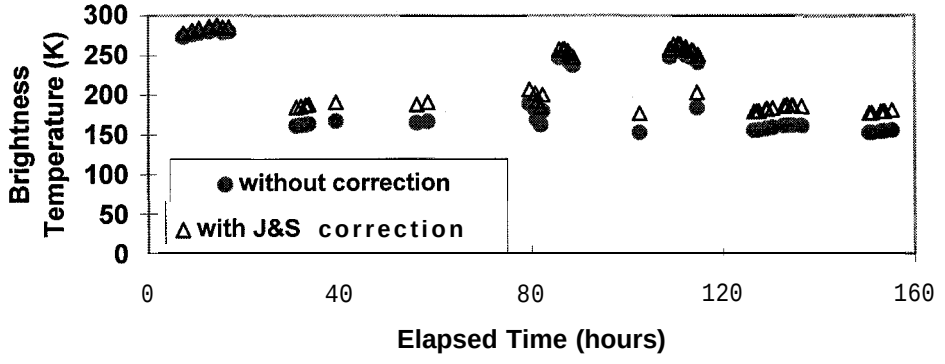
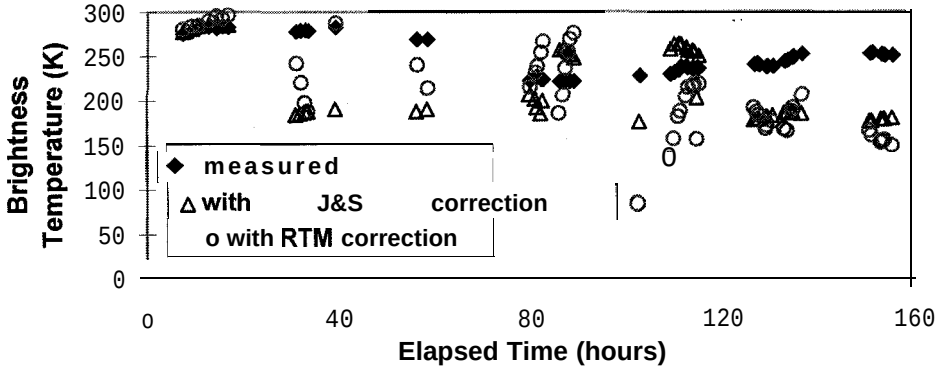


Figure 2. Comparison between the predicted and measured brightness temperature.



Preliminary results show that both models give some unreasonable results with some input data. The test of our radiative transfer model is preliminary. Testing *of the* model with RTM correction would benefit greatly from direct measurements of the vegetation dielectric constant. Currently, values of dielectric constant of vegetation are used based on the work of Ulaby and El-Rayes (1987). The relationship between vegetation water content and dielectric constant is presented only for corn leaves at specific microwave frequencies (for example: 1 GHz). Sensitivity of the model to input data is another area of further study. In addition, the modified model should be tested with experimental data under different vegetation types. Huntsville '96 was limited to two vegetation types, one of grass and one of a mixed vegetation. Only results from the grass plot are presented here.

Conclusion

A modified radiative transfer model is developed to study the effects of a vegetation cover on brightness temperature. Results using an existing vegetation correction procedure are compared with an explicit treatment of the vegetation layer with a radiative transfer model are compared. Results of this preliminary test are encouraging, but require further study. Specifically, test of the sensitivity of the model to the input data and characterization of the effective vegetation layer. In addition, the model needs to be tested against experimental data of different vegetation types. This model, after complete testing, will be used for an inverse study of soil moisture,

Reference

Dobson, M. C., F.T. Ulaby, M. Hallikainen, and M. El-Rayes. 1985. Microwave dielectric behavior of we t soil - Part II: four-component dielectric mixing models. IEEE Trans. Geosci. Remote Sensing. GE-23. Pp.35-46.

Engman, E.T. and N. Chauhan. 1995. Status of microwave soil moisture measurements with remote sensing. Remote Sens. Environ. 51:189-198.

Jackson, T.J. and T.J. Schmugge. 1991. Vegetation effects on the microwave emission of soils. Remote Sens. Environ. 36:203-212.

Laymen, C., W. Belisle, T. Coleman, W. Crosson, A. Fahsi, R. Hood, T. Jackson, J. Luvall, A. Manu, P. O'Neill, Z. Senwo, T. Tsegaye, and S. Wu. 1997. Huntsville '96: a plot-scale experiment in passive and active microwave remote sensing of soil moisture. International Journal of Remote Sensing. (submitted).

Njoku, E.G. and J.A. Kong. 1977. Theory for passive microwave remote sensing of near surface moisture. J. Geophys. Res. 82:3108-3117.

Ulaby, F. T., R.K. Moore, and A.K. Fung. 1986. Microwave Remote Sensing Active and Passive. Vol. III. From Theory to Applications. Artech house. Dedham. M.A.

Ulaby, F. T., M. Razani, and M.C. Dobson. 1983. Effects of vegetation cover on the microwave radiometric sensitivity to soil moisture. IEEE Trans. Geosci. Remote Sens. GE-21:51-61.

Ulaby, F.T. and M.A. El-Rayes. 1987. Microwave dielectric spectrum of vegetation - Part II: Dual-dispersion model. IEEEE trans. Geosci. Remote Sens. GE25-5:550-557.

Wang, J. P. O'Neill, E. Engman, et al. 1982. Remote measurements of soil moisture by microwave radiometers at BARC test site II. NASA Tech. Memorandum 83954.

Acknowledgments: Data for testing the model are based on experimental data of Huntsville' 96. The authors of this paper want to thank all scientists, collaborators, and students who conducted the experiment for their hard work in collecting the field data.

Effects of Grass Vegetation on Land Surface Brightness Temperature: Testing of the Numerical Solution of a Forward Radiative Transfer Model

M.D. Ruark[1], Y.P. Liu[2], R. Inguva[3]

[1] Michigan Technological University, [2] Alabama A&M University, and [3] University of Wyoming.

Abstract

The effects of grass vegetation on L-band passive microwave determination of brightness temperature at land surface are studied, using the numerical code of a forward radiative transfer model. This model, based on soil moisture and temperature profiles, first predicts the brightness temperature at a bare soil surface, then adjusts the brightness temperature using a correction for vegetative cover on soil surface. This correction, characterized by a vegetation water content and a correction parameter, attempts to compensate for changes which occur in radiative transfer when vegetation is present. Data from the HSCaRS Huntsville`96 experiment, specifically the measured brightness temperatures of plots with a smooth bare soil surface and a grass-covered soil surface, is used for comparison with the model predictions. Additionally, the sensitivity of the model prediction to soil moisture and the sensitivity of the vegetation correction to the vegetation water content and correction parameter is examined.

Introduction

The determination of soil moisture through microwave remote sensing has been a subject of research focus in both remote sensing and hydrology fields (Njoku and Kong, 1977, Jackson et al., 1982, Vyas et al., 1990). In theory, passive microwave remote sensing can be used to measure the brightness temperature of the soil surface, which is directly related to soil moisture. One of the research goals of the Center for Hydrology, Soil Climatology and Remote Sensing (HSCaRS) at Alabama A&M University is to develop a computer model which uses observed brightness temperatures to predict the soil moisture profile. The numerical code of a forward radiative transfer model (Njoku and Kong, 1977) recently developed at HSCaRS is the first step in the genesis of such a model. This model uses measured soil density, temperature and moisture profiles to calculate the land surface brightness temperature and a dielectric constant profile.

In this work, the numerical code of the forward radiative transfer model is initially tested using experimental data from a bare-smooth soil plot. The model predictions are then compared with observed brightness temperature values. The sensitivity of the model to soil moisture is examined. The model prediction is next modified by the addition of a vegetation correction, which represents the effects of vegetative cover on radiative transfer. The model is then tested using data from a grass-covered soil plot. The observed brightness temperature from the grass plot is compared with model predictions, with and without a vegetation correction.

Model Description

The numerical code of the forward radiative transfer model developed is based on the theory of Njoku and Kong (1977) and uses vertical profiles of soil density, temperature, and moisture to predict the land surface brightness temperature. A vegetation canopy will absorb the emission of the soil and add its own emission to the total radiative flux (Jackson and Schmugge, 1991). Thus, a vegetation correction is needed for improved prediction of brightness temperature when a vegetative cover is present. The brightness temperature of soil covered with a vegetation canopy is typically described as (Ulaby et al., 1983):

$$T_{B(v)} = (1 + R_S\gamma)(1 - \gamma)(1 - \alpha)\,T_v + (1 - R_S)\gamma\,T_S \qquad (1)$$

where $T_{B(v)}$ = brightness temperature of the vegetation and soil (K)
T_v = physical temperature of the vegetation (K)
T_s = physical temperature of the soil (K)
R_s = reflectivity of the air-soil interface
α = single scattering albedo of vegetation
γ = transmissivity of the vegetation layer.

It is assumed that the scattering albedo is zero (α = O), and that the physical temperature of the soil and vegetation is approximately the same (T_v = T,). Thus:

$$T_{B(v)} = (1 - R_S \gamma^2) T_S \quad (2)$$

Transmissivity is dependent upon the optical depth, τ, and the look angle, θ, so that (Kirdiashev et al., 1979):

$$\gamma^2 = e^{-2\tau \sec(\theta)} \quad (3)$$

where optical depth, τ, is a product of the vegetation water content, WV, and vegetation parameter, b (Shutko, 1986, Jackson and O'Neill, 1990):

$$\tau = W_v b \quad (4)$$

Emissivity and reflectivity of bare soil are related, so that:

$$e = \frac{T_{B(s)}}{T_s} = 1 - R_s \quad (5)$$

where $T_{B(s)}$ is brightness temperature of bare soil, or:

$$T_{B(s)} = (1 - R_s) T_s \quad (6)$$

By subtracting Eq. (6) from Eq. (2), the change in brightness temperature, ΔT_B, due to a vegetation canopy is equivalent to:

$$\Delta T_B = T_{B(v)} - T_{B(s)} = (T_{B(s)} - T_s)(\gamma^2 - 1) \quad (7)$$

The model used in this work is based upon these principles of radiative transfer and brightness temperature determination.

Model Testing

The Huntsville'96 experiment was conducted on four different plots: bare-smooth, bare-rough, grass-covered and mixed vegetation (Laymen et al., 1997). The experiment lasted two weeks. Data from the bare-smooth soil plot was used for initial testing of the model, and data from the grass-covered plot was used for testing the vegetation correction. This data includes vertical profiles of soil density, temperature and volumetric moisture content, and a look angle value. Soil density was measured once during the experiment, while soil temperature and moisture profiles were measured continuously, and averaged every twenty minutes. Surface soil moisture was measured daily. Observed brightness temperature values are used for comparison with the model predictions (see Figure 1: L-band Brightness Temperature). Since brightness temperature values from the elapsed time of 80 to 160 hours seemed to correlate well, values from this time range were chosen for model testing. In doing this, it is assumed that the vegetation correction should be 10°/0 or less. Thus, the brightness temperature differences observed between the bare smooth-soil plot and the grass-covered soil plot from the elapsed time of 220 hours to the finish are not accurate,

Bare and smooth plot Thirty-eight data files were built from the experimental data of two different observation pits on the bare-smooth soil plot. Also, twenty-four data files were built consisting of four different temperature profiles and six different moisture profiles. Temperature profiles were characterized by surface temperatures measuring 30, 27,24 and 21 'C and decreasing 3°C through the depth of the profile, while moisture profiles were characterized by surface moisture measuring 10-30%, and moisture increasing 5% through the depth of the profile. These data files were used to examine the sensitivity of the model to soil moisture.

Grass plot Thirty-two data files were built from experimental data of two different observation pits on the grass-vegetated soil plot. Additionally, twenty-four different data files were built consisting of four different values for the vegetation water content, three different values for the correction parameter and two different temperature profiles. Vegetation water content varied from 0.1, 0.25,0.35, and 0.5 kg/m^2, correction parameter varied from 0.1, 0.3 and 0.5, and temperature profiles were characterized by surface temperatures of 24 and 27°C, decreasing 3°C through the depth of profile. These data files were used to examine the sensitivity of the model to the vegetation correction.

Figure 1: L-band Brightness Temperature

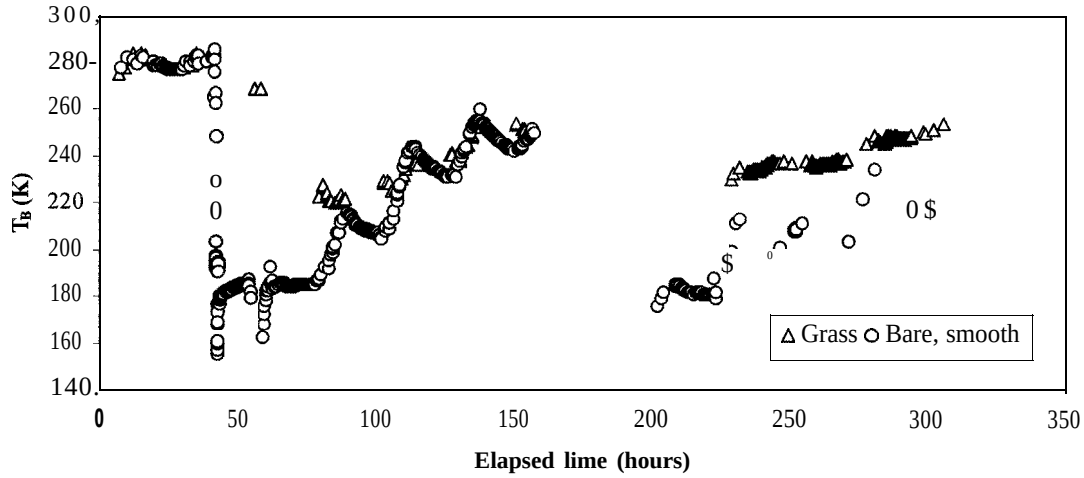
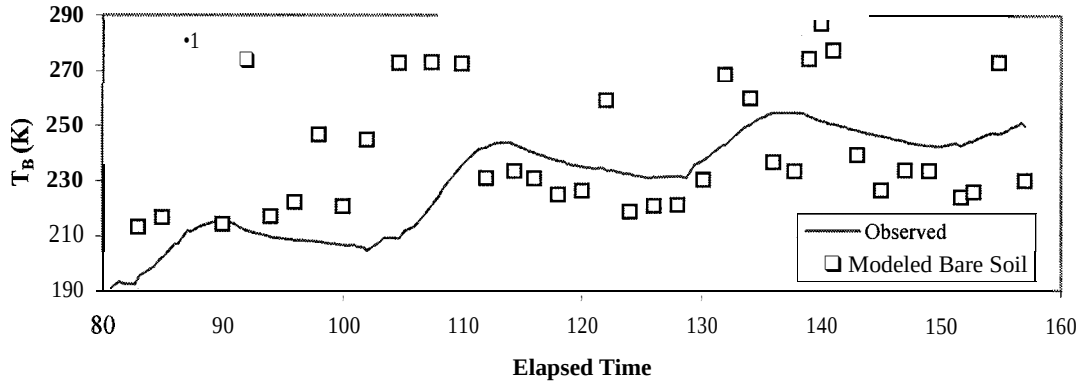


Figure 2: L-band Brightness Temperature of Bare Soil



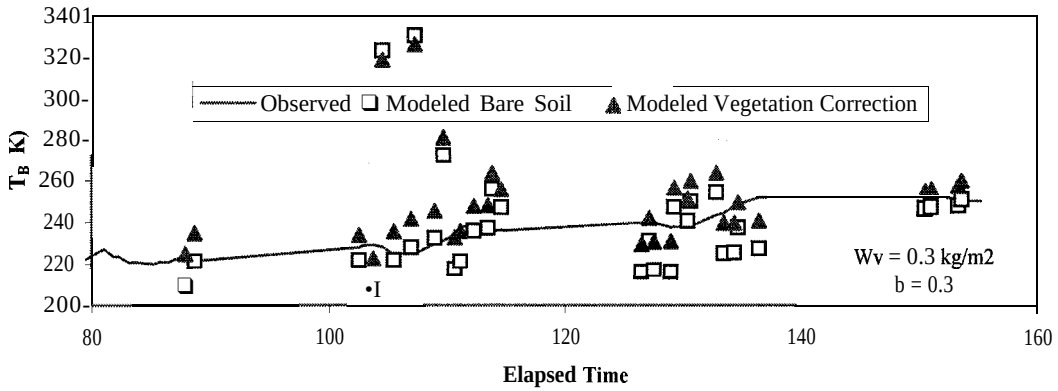
Discussion

Bare and smooth plot The observed brightness temperature of the bare, smooth plot shows an increase of approximately 60 K over the elapsed time (see Figure 2: L-band Brightness Temperature of Bare Soil). Local peaks exist at hours 90, 114 and 138. The model produces results which lie both above and below the observed trendline. The difference between the observed and predicted is, on average, 9.7%. Also, as the observed brightness temperature increases, the model seems more likely to produce a value that is lower than the observed brightness temperature. The sensitivity analysis seems to indicate that a moisture content profile of greater than 20% greatly impairs the ability of the model to accurately predict brightness temperature, especially when coupled with a higher temperature profile. At very high moisture contents, such as 30 and 35%, the model predicts brightness temperature values of several thousand Kelvin.

Grass plot The observed brightness temperature of the grass plot shows an increase of approximately 30 K over the elapsed time (see Figure 3: L-band Brightness Temperature of Grass Vegetation). While the observed values for the bare and smooth soil plot reach daily peaks, the observed values for this plot have plateaus. The observed brightness temperature is far less variable for this plot. This indicates that the grass plot resists changes in brightness temperature. In theory, the vegetation will act as a buffer zone, reducing the impact of changes in soil moisture or air temperature. The observed difference between the two plots supports this. The model without a vegetation correction produces results which are often too low, with two especially high temperatures occurring at hours 104 and 107. These low values are expected, since vegetation should attenuate the radiation better than bare soil. That is, when all other variables are the same, the observed brightness temperature for a vegetated plot should be slightly higher than

that of a bare plot. The difference between the observed and the predicted value without a correction is 7.5% when the two high values are considered, and 5.1 % after this anomaly occurs. The model with a vegetation correction produces results which appear better grouped around the observed trendline. The difference between observed and predicted values is 7.1% including the anomaly and 4.8% after the two high values occur. Thus, use of the vegetation correction resulted in improved prediction of brightness temperature. Examination of the data tiles of the two anomalous values did not yield an explanation for the high brightness temperatures which were predicted. Initially, high soil moisture was suspected. However, these two records had soil moistures that were typical of the entire data set. Values of the look angle and temperature profile were also within reasonable limits. Further study is necessary to explain this puzzle.

Figure 3: L-band Brightness Temperature of Grass Vegetation



The sensitivity analysis indicated a change of only 20% in the vegetation correction within the range of values considered. Since the vegetation correction is typically less than 10% of the total brightness temperature, the impact of changing these parameters is slight. Median values of vegetation water content of 0.3 kg/m^2 and a correction parameter of 0.3 were used in testing the grass-vegetation data set, and have been supported by other research (Wang et al., 1982). These values produced an average correction of 10.8K, or 4.5°/0 of the total brightness temperature.

Conclusion

Research has shown that a relationship exists between brightness temperature measured by microwave remote sensing and soil moisture. A vegetation canopy attenuates radiation from the soil surface, making a vegetation correction is necessary when a model is used to predict the brightness temperature of a vegetated plot. In this study, the ability of a forward radiative transfer model to numerically predict the brightness temperature is tested, both with and without a vegetation correction. The model demonstrated improved prediction of brightness temperature through use of the vegetation correction. The model is sensitive to soil moisture values above 20°/0. Also, the model has the capacity to consider surface roughness and variations in soil moisture. The next step in the development and testing of this model should include an examination of the ability of the model to respond to these variables, as well as an inquiry into the two anomalous values that were produced.

Acknowledgements

Experimental data was acquired during the HSCaRS Huntsville`96 field experiment. The authors would like to extend their thanks to all scientists and students who were there collecting all data.

References

JACKSON, T. J., and SCHMUGGE, T.J. (1991), Vegetation effects on the microwave emission of soils, Remote Sens. Environ., 36:203-212.

JACKSON, T.J., and O'NEILL, P.E. (1990), Attenuation of soil microwave emissivity by corn and soybeans at 1.4 and 5 GHz, IEEE Trans. Geosci. Remote Sens., GE-27:225-235.

JACKSON, T. J., SCHMUGGE, T. J., and WANG, J.R. (1982), Passive microwave remote sensing of soil moisture under vegetation canopies, Water Resources Res., 18:1137-1142.

KIRDIASHEV, K. P., CHUKLANTSEV, A. A., and SHUTKO, A.M. (1979), Microwave radiation of the Earth's surface in the presence of vegetation cover, Radio Eng. Electron Phys., 24:256-264 (Engl. transl.).

LAYMON, C., BELISLE, W., COLEMAN, T., CROSSON, W. FAHSI, A, HOOD, R, JACKSON, T., LUVALL, J., MANU, A., O'NEILL, P., SENWO, Z, TSEGAYE, T., and WU, S. 1997. Huntsville '96: a plot-scale experiment in passive and active microwave remote sensing of soil moisture. International Journal of Remote Sensing. (submitted).

NJOKU, E. G., and KONG, J.A. (1977), Theory for passive microwave remote sensing of near-surface moisture, J. Geophys. Res., 82:3108-3117.

SHUTKO, A.M. (1986), Microwave radiometry of water surface and grounds. Nauka, Moscow, 192 pp. (Engl. transl.).

ULABY, F. T., RAZANI, M., and DOBSON, M.C. (1983), Effects of vegetation cover on the microwave radiometric sensitivity so soil moisture, IEEE Trans. Geosci. Remote Sens., GE-21:51-61.

VYAS, A. D., TRIVEDI, A. J., CALLA, O.P.N., RANA, S. S., SHARMA, S. B., and VORA, A.B. (1990), Experimental data for separation of vegetation and soil estimation of soil moisture using passive microwaves, Int. J. Remote Sens., 11(8):1421-1438.

WANG, J., O'NEILL, P., ENGMAN, E., et al. (1982). Remote measurements of soil moisture by microwave radiometers at BARC test site II. NASA Tech. Memorandum 83954.

Development of an OPO-Based Lidar System and Differential Absorption Measurements of Methane

Sang W. Lee, T. Zenker, T. H. Chyba

Hampton University

Corresponding Author: Sang W. Lee

Research Center for Optical Physics

Department of Physics

Hampton University

Hampton, VA 23668

Sangwool@gprc.hamptonu.edu

An optical parametric oscillator (OPO) ground-based lidar system in the wavelength region of 1.45-4 pm for the remote measurement of methane will be described. The laser transmitter consists of an injection-seeded Nd:YAG laser which pumps an OPO. The OPO output is tunable from 1.45-4 pm, with a bandwidth less than 500 MHz, and a pulse energy of 1 to 3 mJ at 3.29 μm. The receiver is cart-mounted and consists of a 14" telescope with 1.57 and 3.29 μm detector channels. A fast oscilloscope is used for data acquisition. The system performance will be tested through measurements of sources of atmospheric methane.

1. Introduction

There is a scientific need to know the source strength and the global distribution of the tropospheric trace gases relevant for global warming and tropospheric chemistry [1-3] as well as a practical need to be able to measure gases related to the petrochemical industry. Lidar (Light Detection And Ranging), an active remote sensing technique, has several advantages over passive remote sensing techniques because of the laser's unique properties of high power, monochromaticity, short pulse duration, and beam collimation [4,5].

The differential absorption lidar (DIAL) technique [5-10] is widely used for range resolved measurements of the concentration of atmospheric constituents. In this method, the backscattered pulse energies at two wavelengths, λ_{on} and λ_{off}, are measured. A comparison is made between the scattered signal monitored when the wavelength (λ_{on}) of the laser is tuned to the peak of an absorption line of the molecule being investigated and when it (λ_{off}) is detuned to lie in the wing of the line. The path-averaged concentration of the absorbing molecule in the range cell R_2-R_1 is given by

$$N_A = [\ln\{P_{on}(R_1)P_{off}(R_2)/P_{off}(R_1)P_{on}(R_2)\}]/[2\{R_2-R_1\}\{\sigma_A(\lambda_{on})-\sigma_A(\lambda_{off})\}]$$

where σ_A is the absorption cross section of the absorbing molecule at λ_{on} and λ_{off}. P is the backscattered signal power normalized to the transmitted power, and R is the range to the target. The advantage of the DIAL technique is that only differences or ratios in the various parameters need to be considered.

2. Technical Approach

Figure 1 shows the overall lidar system schematic diagram. The lidar system consists of four parts : the laser source, laser wavelength diagnostics, the receiver, and the data acquisition electronics. Each of these will now be briefly described.

A. Transmitter

A frequency-doubled injection-seeded Nd:YAG laser pumps the Continuum Mirage 3000 OPO [12-16] that provides the tunable mid-infrared laser pulse. The OPO consists of three conversion stages which are a master oscillator, a nonresonant oscillator, and an amplifier. The master oscillator generates a single longitudinal mode. The nonresonant oscillator and amplifier stages convert the wavelengths to 1.5-4 pm. The OPO gives up to 3 mJ of energy at 3 μm with a pulse repetition rate of 10 Hz and a bandwidth less than 500 MHz. The five KTP crystals for the three stages and the rear mirror in the master oscillator are mounted on the motorized optic mounts. Wavelength tuning is controlled by a computer. The OPO simultaneously produces a 3.29 μm pulse and a 1.57 μm pulse colinearly from its output port. Since sequential pulses are not on-line and off-line pulse pairs, this is not a DIAL system in the usual sense. Rather, it is a tunable single frequency lidar which can be slowly tuned across an absorption feature in order to make DIAL measurements.

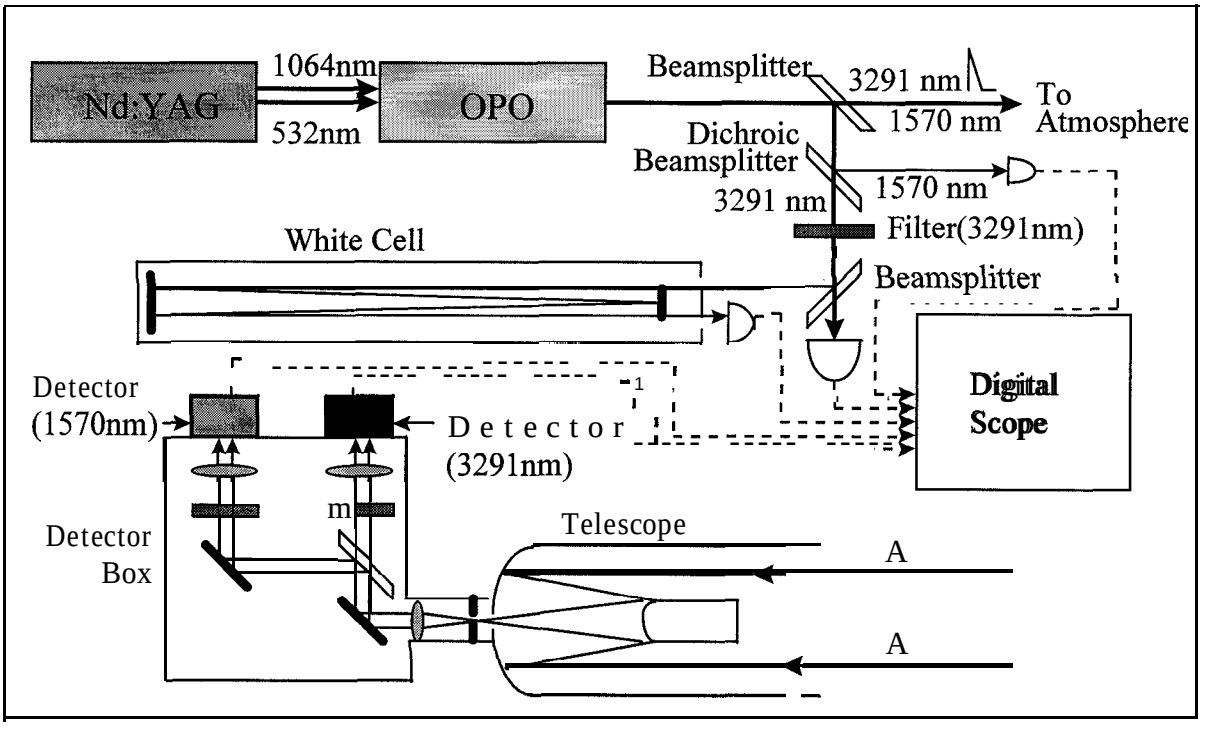


Figure 1. Lidar System Schematic Diagram

B. Laser Wavelength Diagnostics

Before the 1.57 μm and 3.29 μm beams are transmitted into the atmosphere, a portion of the 3.29 μm beam is diverted into a White cell. The 3.29 μm pulse energy before and after the absorption cell is measured and the ratio computed. This is used to verify that the laser is centered on the absorption line of interest. This technique is used to calculate the spectral purity of the laser source by comparing the theoretical cell absorption with the measured absorption.

C. Receiver

As shown in Figure 2, the receiver package for the mid-IR DIAL consists of a 14 inch telescope joined to a detector box containing a dichroic beam splitter, focusing lenses, and detectors for 1.57 and 3.291 μm wavelengths. This assembly will be cart-mounted for portability. The elevation angle of the system can be adjusted. The transmitted laser beam is made parallel to the telescope axis by beam steering optics.

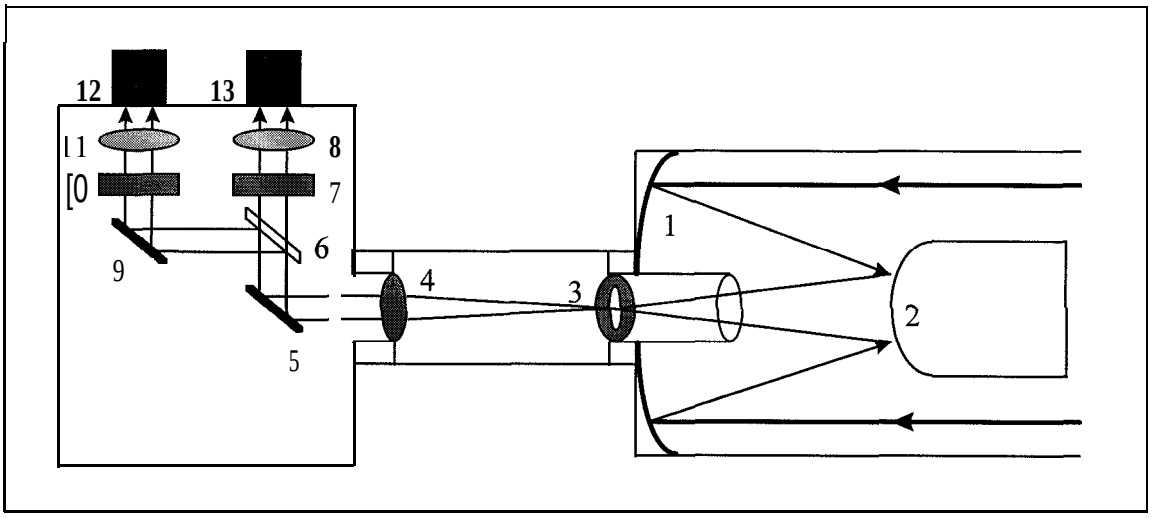


Figure 2. The Receiver Layout. 1. Mirror(14") 2. Mirror(4") 3. Variable Field Stop 4. Lens (Planoconvex, CaF_2) 5. IR Mirror 6. Dichroic Beamsplitter 7. Filter(3.29μm) 8. Lens(Planoconvex,CaF_2) 9. IR Mirror 10. Filter (1.57mm) 11. Lens(Planoconvex, CaF_2) 12. Detector(InGaAs) 13. Detector(InSb)

After passing through the telescope and an aperture which sets the field of view, the 1.57 μm return signal is reflected 99 % by a 45° dichroic beamsplitter and the 3.3 μm return signal is transmitted by the dichroic with less than 10 % loss. The energy passing through the beamsplitter is focused by a CaF_2 lens onto a LN_2 cooled 0.25 mm diameter InSb photodiode detector for the 3.29 μm channel and a room temperature 2 mm diameter InGaAs detector for the 1.57 μm channel.

D. Data Acquisition

The signals collected by the receiver will be recorded by a 4-channel 1 GHz digitizer. Data from the digitizer will be read into the computer using a GPIB interface.

3. **Preliminary Results**

A. Laser properties

The expected performance of our Research Center for Optical Physics (RCOP) system can be compared with a 3.3 μm DIAL system which has been recently demonstrated by SRI International [6], and with an external cavity laser diode-injection seeded OPO system developed by the National Physical Laboratory (NPL) [7]. The properties of these three systems are listed in Table 1. The SRI system is capable of making DIAL methane measurements with good accuracy at the range of 600-1000 m. Table 1 shows that our laser transmitter has a much narrower linewidth than that of the SRI laser. The FWHM linewidth

of the methane absorption at 3.29 μm is approximately 0.3 nm. The SRI laser linewidth is eight time greater than this value, but our laser linewidth is more than 15 time narrower than it. Therefore, our entire on-line pulse sees the peak absorption cross-section and has less interference from other gases. Our laser transmitter produces comparable output energy but at half of the repetition rate. The performance of our laser is comparable to that of the NPL system without its additional OPA stage.

	RCOP System	SRI System	NPL System
Wavelength	Tunable from 1.45 to 4 pm*	Fixed at 3.309 and 3.313 pm	Tunable from 3.26 to 3.51 μm
Energy at 3.3 pm	1 to 3 mJ*	2 mJ	3-6 mJ(OPO), 15-30 mJ(OPO+OPA)
Puke width	5 ns*	10 ns	8 ns
Line width	15 pm	2.5 nm	6.5 pm
Repetition frequency	10 Hz	20 Hz	10 Hz
Beam diameter	4 mm*		4 mm
Beam divergence (full angle)	2 mrad*	1 mrad	5 mrad
Spectral purity	99.95%*		

*indicates parameters verified in the lab.

Table 1. Comparison of RCOP system with the SRI and NPL system

Figure **3 shows the results of methane absorption measurements made with our laser and diagnostic White cell.** Very good agreement is obtained with the corresponding values from the Hitran database. The White cell is filled with 1 atm of ambient room air and is set for a total path length of 90 m. We have used the absorption cell and ambient room air to measure the laser spectral purity to be 99.95°/0 within a 160 pm bandwidth. The preliminary measurement will be repeated with an unbroadened absorption feature.

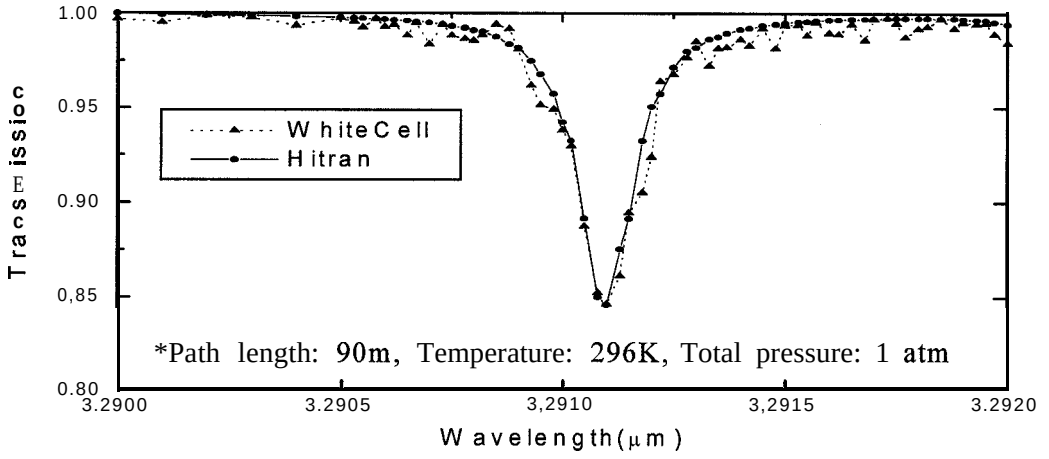


Figure3. Methane absorption measurement at 3 .291μm

We have calculated the expected performance of the lidar system. Figure 4(a) is plot of optical return power of the two wavelengths, λ_{on} and λ_{off}, vs. range. The on-line signal attenuation is faster than that of the off-line signal because of the absorption of methane. Figure 4(b) is an expanded picture of figure 4(a). As shown in the figure, the optical power corresponding to the detector amplifier noise level is about 6.3* 10-8 W. Therefore, range resolved measurements with our lidar system are possible out to 1 km. Parameters used in this calculation are: pulse energy = 3 mJ, detection bandwidth= 10 MHz, receiver throughput = 0.3, detector responsivity = 2.0 A/W, and input amplifier noise voltage = 2 $nV/(Hz)^{1/2}$.

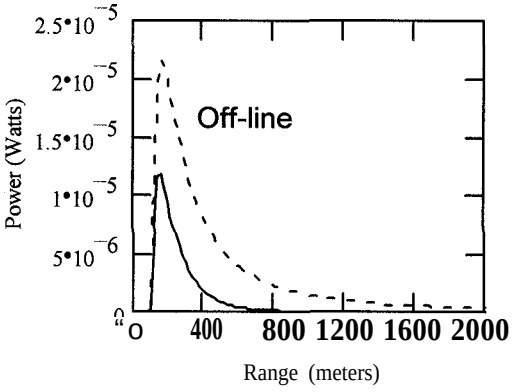
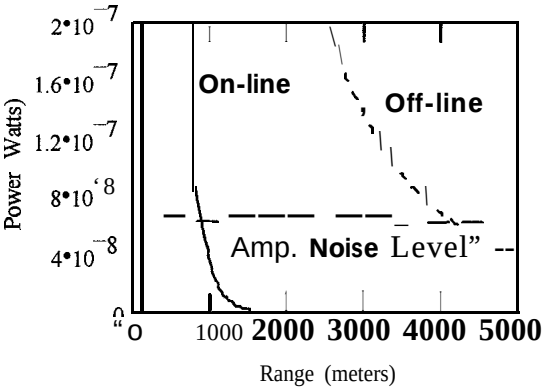


Figure 4. (a) Calculated lidar return power vs. Range



(b) Enlargement of(a)

B. Status of the receiver

The telescope received from NASA LaRC is now being modified for near-IR applications. The mount, telescope optics, and IR detector for the receiver have been ordered.

4. Future Work

Once the lidar system is completed, we will make lidar measurements and model the system to understand its capabilities and to determine its limitations.

5. Summary

Development of an OPO-based tunable near IR lidar system is described. The laser produces high spectral purity (99.95% for ambient water vapor) but relatively low output energy for long range measurements. We will investigate possible modification to our system to increase the laser pulse energy with an additional OPA stage. Total optical efficiency of the lidar receiver is about 0.3. The system can make range resolved measurements out to 1 km.

Acknowledgments

This work was supported by NASA grant NAGW2929 and NCC-1-214. We thank Dr. Edward Browell at NASA LaRC for use of the 14" telescope.

References

1. Scientific Assessment of Ozone Depletion: 1994, World Meteorological Organization Global Ozone Research and Monitoring Project-Report No. 37,2.16-2.20, (1994).

2. Mission to Planet Earth (MTPE) Strategic Enterprise Plan 1996-2002, NASA, (1996).

3. Climate Change, The IPCC (Intergovernmental Panel on Climate Change) Scientific Assessment, Cambridge University Press: New York, Port Chester, Melbourne, Sydney, (1990).

4. Dennis K. Killinger, and Norman Menyuk, "Laser remote sensing of the atmosphere", Science, 235, 37-45 (1987).

5. R. M. Measures, Laser Remote Sensing, Malabar: Krieger Publishing Company, (1992).

6. N. Nielsen, J. Van der Laan, E. Uthe, W. Bosenberg, R. Kaiser, and C. Carlisle, "Development and testing of a compact airbone 3-μm wavelength DIAL", Applied Optics, submitted.

7. M. J. T. Milton, T. D. Gardiner, G. Chourdakis, and P. T. Woods, "Injection seeding of an infrared optical parametric oscillator with a tunable diode laser", Optics Letters, 19, 281-283, (1994).

8. M. J. T. Milton, R. H. Bradsell, B. W. Jolliffe, N. R. W. Swarm, and P. T. Woods, "The design and development of a near-infrared DIAL system for the detection of hydrocarbons", Proceedings of the 14th International Laser Radar Conference, 370-373, (1988).

9. H. L. Walmsly and S. J. O'Connor, "The measurement of atmospheric emissions from process units using differential absorption lidar", in Lidar Atmospheric Monitoring, Jean-Pierre Wolf, Editor, Proceedings of SPIC Vol.3104, 60-72, (1997).

10. W. B. Grant, "Lidar for atmospheric and hydrospheric studies", Tunable Laser Applications edited by F. J. Duarte, New York: Marcel Dekker, (1995), pp 213-305.

11. J. C. Scoot, R. A. M. Maddever, and A. T. Paton, "Spectroscopy of methane using a Nd:YAG laser at 1.34 μm", Appl. Opt., 31, 815-821, (1992).

12. W. R. Bosenberg and D. R. Guyer, "Broadly tunable, single-frequency optical parametric frequency-conversion system", J. Opt. Sot. Am. B, 10, 1716-1722, (1993).

13. T. D. Raymond, P. Esherick, and A. V. Smith, "Widely tunable single-longitudinal-mode pulsed dye laser", Optics Letters, 14, 1116-1118, (1989).

14. W. R. Bosenberg, L. K. Cheng, and J. D. Bierlein, "Optical parametric frequency conversion properties of KTA," Appl. Phys. Lett., 65,2765 (1994).

15. W. Koechner, Solid-state laser engineering, New York: Springer-Verlag, (1992).

16. B. J. Orr, M. J. Johnson, and J. G. Haub, "Spectroscopic applications of pulsed tunable optical parametric oscillator", Tunable laser applications edited by F. J. Duarte, New York: Marcel Dekker, (1994), pp 11-82.

98 URC004

Fourier Transform Infrared (FT-IR) Spectroscopy of Atmospheric Trace Gases HCl, NO and SO_2

C. Haridass, A. Aw-Musse, E. Dowdye, C. Bandyopadhyay and P. Misra†
Center for the Study of Terrestrial and Extraterrestrial Atmospheres and
Department of Physics and Astronomy
Howard University, Washington, D.C. 20059, U.S.A.

and

H. Okabe
Center for the Study of Terrestrial and Extraterrestrial Atmospheres and
Department of Chemistry
Howard University, Washington, D.C. 20059, U.S.A.

ABSTRACT

Fourier Transform Infrared (FT-IR) spectral data have been recorded in the spectral region 400-4000 cm^{-1} of hydrogen chloride and sulfur dioxide with I cm^{-1} resolution and of nitric oxide with 0.25 cm-i resolution, under quasi-static conditions, when the sample gas was passed through tubings of aluminum, copper, stainless steel and teflon. The absorbance were measured for the rotational lines of the fundamental bands of $^{1}H^{35}Cl$ and $^{1}H^{37}Cl$ for pressures in the range 100-1000 Torr and for the $^{14}N^{16}O$ molecule in the range 100-300 Torr. The absorbance were also measured for individual rotational lines corresponding to the three modes of vibrations (υ_1-symmetric stretch, υ_2-symmetric bend, υ_3-anti-symmetric stretch) of the SO_2 molecule in the pressure range 25-150 Torr. A graph of absorbance versus pressure was plotted for the observed rotational transitions of the three atmospherically significant molecules, and it was found that the absorbance was linearly proportional to the pressure range chosen, thereby validating Beer's law. The absorption cross-sections were determined from the graphical slopes for each rotational transition recorded for the HCl, NO and SO_2 species. Qualitative and quantitative spectral changes in the FT-IR data will be discussed to identify and characterize various tubing materials with respect to their absorption features.

Key Words: FT-IR, atmospheric, spectroscopy, infrared radiation, absorbance, absorption cross-sections.

INTRODUCTION

At the present time, environmental issues like global climatic change, photochemical smog formation, acid rain, stratospheric ozone hole and deforestation are of great concern worldwide. These issues are related to increasing concentrations of atmospheric trace constituents, such as carbon dioxide (CO_2), nitric oxide and nitrogen dioxide (NO_x), nitrous oxide (N_2O), sulfur dioxide (SO_2), methane (CH_4), and hydrogen chloride (HCl), and their average residence times in the atmosphere. [1] As an illustration, the influence of CO_2, NO_x and N_2O on stratospheric ozone depletion depends on the altitude, and furthermore, CH_4 acts in general against ozone depletion, yet accelerates the depletion within the ozone hole. Measures for reduction of the emissions are either already in effect, or will be implemented, since the concentrations of most of the trace gases, except presumably NO, and SO_2, are expected to increase further, mainly because of their longevity in the atmosphere. The greenhouse effect" is related to the radiation budget of the earth and responsible for the fact that the mean earth temperature is 330 C above its radiation temperature as viewed from space. The contribution of a given gaseous compound to the greenhouse effect is determined by its atmospheric abundance and IR absorptivity. Gases like CO_2 and CH_4 are found to be responsible for more than 800/0 of the enhancement of the greenhouse effect.[4] Nitric oxide can hinder ozone destruction, which is manifested by low nitric oxide concentration in the ozone hole.

Small concentrations of atmospheric pollutants can also have dramatic effects leading to smog formation. The production of smog is generally restricted to densely populated and industrial areas, where increased concentrations of toxic compounds, such as sulfur dioxide (SO_2), occur near their sources. This type of smog, known as *winter smog* or *London smog,* has lost much of its previous importance owing to drastic reductions in emission of so-called primary pollutants, like sulfur compounds from heating systems. Recently, another type of smog known as *summer smog or Los Angeles smog,* has attracted a great deal of interest. This type of smog is produced photochemically from primary pollutants, essentially from NO_x and voltalic organic compounds (VOCs), under the influence of solar ultraviolet radiation. Hydrochloric acid (HCl), one of many key atmospheric species, is also considered a possible entity in stratospheric ozone depletion. The HCl molecule is formed in the stratosphere predominately by reactions between hydrogen containing species (such as HO_2 and H_2) with CH_4 and Cl radicals. The transport of HCl down into the troposphere, and the competition between various stratospheric photochemical reactions, govern the regeneration of free chlorine and hence its return into the O_3 destruction cycle. This latter process effectively removes Cl from the ClO_x cycle.

Different measurement techniques can be applied to measure the small absorption encountered in trace detection. Fourier Transform (FT) spectrometers are operated with conventional broad band sources, in conjunction with interferometers, to achieve the desired spectral resolution. FT systems possess higher optical and greater observation time efficiencies, since all spectral elements are simultaneously observed and measured. FT-IR spectrometers are very useful for chemical analyses of multi component mixtures,[5,6] because of the available broad infrared range and simultaneous observations in several spectral regions.

Matrix isolation, in combination with the FT-IR technique [7] has been considered very attractive for laboratory analyses of the composition of sampled air, because it avoids the problem of interferences due to reduction in spectral congestion compared togas phase spectra. This technique can be applied to a wide range of stable and moderately labile atmospheric trace gases with detection limits in the parts per trillion range. Long-path absorption spectroscopy with lead salt diode lasers has profited from the recent development of tunable IR lasers.[8,9] Very high spectral resolution, together with increased sensitivity and time response, are possible using the technique of tunable diode laser absorption spectroscopy (TDLAS). TDLAS has been performed both with open atmospheric paths and with confined air samples in long-path cells. The technique is not as universal as FT-IR spectroscopy, owing to the limited tuning range of a single diode laser.[10] Fried *et al.*[11] have reported results based on a laboratory study for detecting the important atmospheric molecule HCl using a tunable diode laser coupled with a multipass White cell. These authors also claim that in working with very low concentrations of the highly polar HCl molecule they have encountered problems arising from surface interactions of HCl with the White cell and the inlet plumbing. This has led us to undertake a systematic study of the interaction of atmospheric trace gases like HCl, NO, and SO_2, with the tubing materials used for transporting the gases from the main reservoir to the absorption cell. Thus, in the present work, we report absorption cross-sections of the individual rotational lines of the atmospherically significant molecules, namely HCl, NO, and SO_2, in the gas phase, when these gases were passed through tubes made of different materials, such as aluminum, copper, stainless-steel, and teflon. The measurements were done using a Nicolet Magna-IR 550 Fourier Transform infrared spectrometer. The objective of the present paper was to identify and characterize various tubing materials with respect to their absorption features, so that specific tube-gas combinations may be used to overcome the difficulty of measurement problems associated with a significant loss of sample concentration via adsorption on surfaces of tubings and associated instrumentation.

EXPERIMENTAL

The infrared spectra were recorded using a Nicolet Magna-IR 550 Fourier Transform spectrometer. A 10-m path length multipass absorption cell was used for recording the spectra.[12] A quasi-static system was used for recording

the spectra of the sample gases. Tubes of one meter length and made of different materials, namely aluminum, copper, stainless-steel and teflon, were used for recording signature spectra of specific gas samples and to examine the reactivity of the sample with the tubing material. Resolution was set at 1 cm-i for recording of the FT-IR spectra under quasi-static conditions for HCl and SO_2 gases, and at 0.25 cm-l for recording the FT-IR spectra of NO. Thirty six scans (with 1 cm-l and 0.25 cm-' resolution) were co-added after subtraction of the background, each time a sample was examined, and the results were superimposed and averaged for final spectral display. The time required to record each spectrum was less than five minutes. KBr windows were used in the sample cell for better transmission in the mid-infrared region.

Hydrogen chloride (HCl) at a pre-mixed concentration of 529.0 ppm in N_2; nitric oxide (NO) at a pre-mixed concentration of 503.0 ppm in N_2; and sulfur dioxide (SO_2) at a pre-mixed concentration of 484.0 ppm in air, obtained from Scott Speciality Gases, Inc., Plumsteadville, PA ,were used. The absorbance were measured for the rotational lines of the fundamental bands of HCl for pressures in the range 100-1000 Torr, and for the NO molecule in the range 100-300 Torr. Absorbance were also measured for the individual rotational lines corresponding to the three modes of vibration (vi-symmetric stretch, υ_2-symmetric bend, and υ_3-antisymmetric stretch) of the SO_2 molecule for pressures in the range 25-150 Torr. Characteristic features were obtained for HCl in the spectral range 3100-2600 cm^{-1}, for NO in the range 1950- 1780 cm-l, and for sulfur dioxide in the range 1425-400 cm-l. A graph of absorbance versus pressure was plotted for all of the observed rotational lines for the three atmospherically significant molecules, and it was found that the absorbance was linearly proportional to the pressure range chosen, thereby validating the Beer-Lambert law:

$$I = 10 \exp(-\sigma_{\upsilon} N \star l), \quad (1)$$

where σ_{υ} is the absorption cross section, 1 the length of the absorption cell and N the molecular density as a function of pressure given by $N = (N_A \times P_{Torr})/RT$, where N_A is the Avogadro Number, P_{Torr} the total pressure of the gas, R the molar gas constant and T the temperature. The absorbance of each line was then given by A = (l./l), where 1_0 is the incident intensity and I the absorbed intensity.

RESULTS AND DISCUSSION

*Hydrogen Chloride: **FT-IR** Spectra and Analysis*

The FT-IR spectra of the fundamental (l-O) band of the diatomic molecule $^1H^{35}Cl$ and its natural abundance isotopomer $^1H^{37}Cl$ were recorded in the spectral range 3100-2600 cm-l in dry N_2, at a resolution of 1 cm^{-1}, and with gas pressures of 100-1000 Torr under quasi-static conditions using 1 meter tubes made of aluminum, copper, stainless-steel and teflon, respectively. The main purpose of recording the fundamental band of HCl using the FT-IR spectrometer was to monitor the changes in the intensities of the individual rotational lines in the spectra when the experimental gas was subjected to flow through different tubing materials. The absorbance often rotational lines of the P and R branches for both $^1H^{35}Cl$ and $^1H^{37}Cl$ for six different pressures (100, 200, 300, 500,700 and 1000 Tom) were measured when the gas was flown in turn through aluminum, copper, stainless-steel and teflon tubes. Small changes in the spectra of the experimental gas can be quantitatively understood by calculating the absorption cross-sections for each observed rotational frequency. In the present work, we have calculated the absorption cross-sections by plotting the absorbance versus pressure and using the Beer-Lambert Eqn. (1). . It was found that the absorbance showed a linear relationship to the amount of HCl present in the gas cell, confining the validity of Beer's Law for the pressure range employed in the experiments (See Fig. 1). Based on the slope of the plot of absorbance versus P_{Torr}, the absorption cross-sections for the observed rotational lines for both $^1H^{35}Cl$ and $^1H^{37}Cl$ were determined. The calculated absorption cross-sections for $^1H^{37}Cl$ are given in Table 1. From the data given in this table, it is evident that there was greater detectable variation in the absorption cross-sections when the gas was passed through aluminum and copper tubings, in contrast to the values of absorption cross-sections obtained when the gas was passed through stainless steel and teflon tubings. The absorption cross-sections for both $^1H^{35}Cl$ and $^1H^{37}Cl$ are shown for comparison in Fig. 2 when the gas was passed through aluminum tubing.

*Nitric Oxide: **FT-IR** Spectra and Analysis*

Nitric oxide is produced in the earth's atmosphere mainly by the oxidation of nitrous oxide (N_2O, a by-product of microbial metabolism) by the excited atom of oxygen (in the ^1D state), and to some extent by high-altitude aircraft, nuclear blasts, volcanoes, and lightning. [13] Unlike HCl, the ground state of the NO molecule is a $^2\Pi_r$ state. The rotational structure of a band of the $^2\Pi$-$^2\Pi$ system gives rise to 12 branches. In the present work, because of low resolution (0.25 cm-l), it was possible to observe only the P and R branches of both the $^2\Pi_{1/2}$-$^2\Pi_{1/2}$ and $^2\Pi_{3/2}$-$^2\Pi_{3/2}$ systems as shown in Fig. 3. In the present work, our main goal was to record the fundamental (l-O) band of the nitric oxide molecule using the Fourier Transform spectrometer when the gas was subjected to flow through different tubes made of materials like aluminum, copper, stainless steel and teflon, in order to understand qualitatively and quantitatively the changes in the spectra. Our quantitative analysis was focused mainly on estimating the absorption cross-sections from the observed absorbance at five different pressures (100, 150, 200, 250 and 300 Torr). As mentioned in the previous section, the absorbance of 21 lines of the P and R branches of the two systems were measured for five different pressures. It was found that there was an increase in the absorbance as the pressure of the NO gas was increased in the

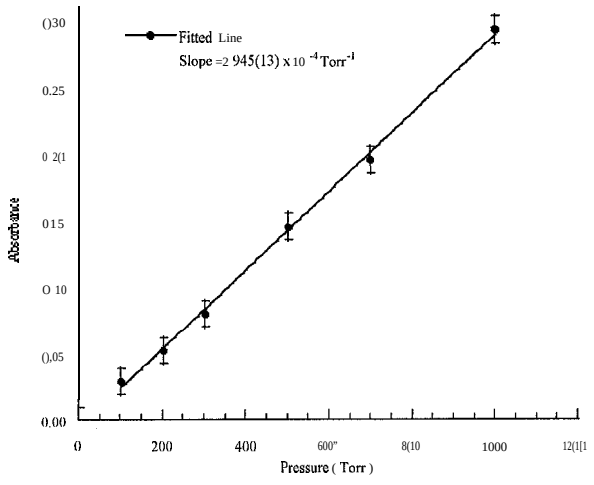


Fig. 1 A plot of absorbance versus pressure for the rotational line P(4) of $^{1}H^{37}Cl$ when the gas was passed through 1 meter teflon tubing.

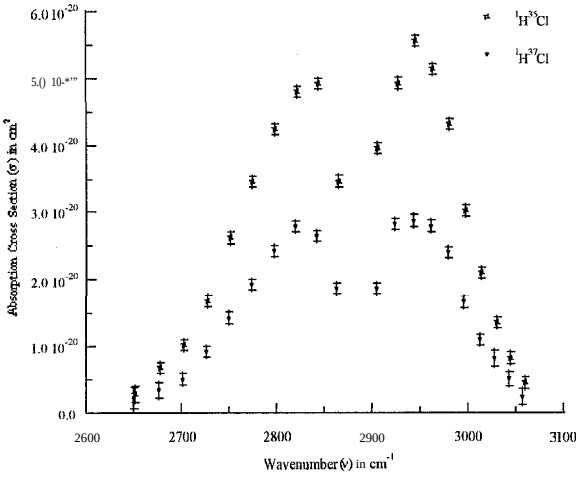


Fig.2. Absorption cross section (σ) versus wavenumber (ν) for $^{1}H^{35}Cl$ and $^{1}H^{37}Cl$ when the gas was passed through aluminum tubing,

Table 1. Absorption cross-sections' for $^{1}H^{37}Cl$ (under quasi-static conditions) when the gas was passed through different tubing materials.

	v (cm-l)	Absorption Cross-Section (σ_v)× 1020 cm^2				v (cm-])	Absorption Cross-Section (σ_v)× 10^{20} cm^2			
J	R(J)	Al	Cu	SS	Tef	P(J)	Al	Cu	SS	Tef
o	*2904.09*	1.87(8)	1.83(8)	1.69(8)	1 . 6 7 (8)					
1	*2923.69*	*2.83(8)*	*2.73(8)*	*2.55(8)*	*2.53(8)*	*2862.97*	1.87(8)	1.82(8)	1.69(8)	*1.68(8)*
2	*2942.71*	*2.87(8)*	*2.78(8)*	*2.60(8)*	*2.58(8)*	*2841.58*	*2.65(8)*	*2.55(8)*	*2.38(8)*	*2.65(8)*
3	*2961.10*	*2.79(8)*	*2.81(8)*	*2.55(8)*	*2.50(8)*	*2819.51*	*2.78(8)*	*2.68(8)*	*2.50(8)*	*2.78(8)*
4	*2978.66*	*2.40(8)*	*2.47(8)*	*2. 19(8)*	*2. 13(8)*	*2796.91*	*2.43(8)*	*2.34(8)*	*2.20(8)*	*2.43(8)*
5	*2995.78*	*I .66(8)*	1 .60(8)	1.49(8)	1.46(8)	*2773.77*	1.93(8)	*1.87(9)*	*1.74(8)*	*1.93(8)*
6	*3012.12*	*1.10(8)*	*1.05(8)*	*0.98(8)*	*0.96(8)*	*2750.09*	*1.42(8)*	1.35(8)	1.27(8)	*1.42(8)*
7	*3027.73*	*0.8* (1)	*0.72(8)*	*0.66(8)*	*0.63(8)*	*2725.88*	*0.92(8)*	*0.88(8)*	*0.84(8)*	*0.92(8)*
8	*3042.71*	*0.5* (1)	*0.42(8)*	*0.38(8)*	*0.37(8)*	*2701.13*	*0.5 1(8)*	*0.49(8)*	*0.45(8)*	*0.51(8)*
9	*3056.81*	*0.3* (1)	*0.24(9)*	*0.19(9)*	*0.17(8)*	*2675.96*	*0.3 (1)*	*0.29(9)*	*0.25(9)*	*0.26(9)*
10	*3072.92*		0.3 (9)	0.2 (8)	0.4 (8)	2650.19	0.2 (2)			0.12(9)

'The number in parentheses indicates the estimated error for the absorption cross-section based on the uncertainty in the slope. Abbreviations are: Al-Aluminum: Cu-Copper: SS- Stainless-Steel: Tef: Teflon.

10-m path length cell. Absorption cross-sections were calculated from the slope of the straight line obtained from the absorbance versus pressure plots. The straight line plot confirmed that Beer's law holds good for the pressure range of the NO gas used. Absorption cross-sections were calculated for all the observed rotational lines of both the $^{2}\Pi_{1/2}$-$^{2}\Pi_{1/2}$ and $^{2}\Pi_{3/2}$-$^{2}\Pi_{3/2}$ systems . Table 2 summarizes the absorption cross-sections for the R and P branches of the $^{2}\Pi_{1/2}$-$^{2}\Pi_{1/2}$ transition when the gas was passed through the four tubing materials. A detectable variation in the absorption cross-sections was observed from this table when the gas was passed through aluminum and teflon tubings, as compared to copper and stainless steel tubings.

***Sulfur** Dioxide: **FT-IR** Spectra and Analysis*

Sulfur dioxide is known as an important constituent for both planetology and astrophysics since it is present on Venus and in interstellar clouds. It is also significant for terrestrial atmospheres because large amounts of SO_2 can be expelled from strong volcanic eruptions. For example, the volcanic eruption of Mount Pinatubo in the Philippines released approximately 20 million tons of sulfur in the atmosphere in June 1991. Once the SO_2 molecules are present in the stratosphere, they are converted into sulfate aerosols, which affect both climate and stratospheric chemistry. Spectroscopic methods are used as standard techniques to track atmospheric concentrations of the SO_2 molecule. High levels of sulfur dioxide were observed from remote infrared measurements shortly after massive volcanic eruptions.[14,15] In the present work, our main goal was to quantitatively determine the reaction of SO_2 with different materials, so that one can emphasize the use of a particular material to transfer the experimental gas from the main reservoir to the absorption cell, and to overcome the loss of concentration from the gaseous adsorption on available surfaces. The fundamental bands of SO_2: ν_1= 1151.71352 cm-1, ν_2=5 17.8724 cm^{-1} and ν_3=1362.0603 cm-' were recorded using the FT-IR spectrometer at 1 cm^{-1} resolution for five different pressures (25, 50,75, 100, 125, and 150 Torr) when the gas was passed through aluminum, copper, stainless-steel and teflon tubes. The absorption cross sections were

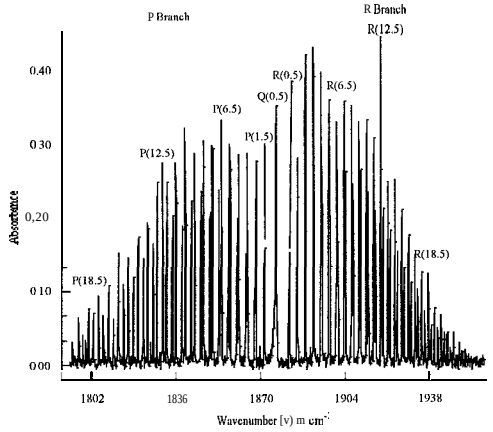


Fig. 3 FT-IR spectrum of nitric oxide (NO) under quasi-static conditions recorded at 0.25 wavenumber resolution with a gas pressure of 100 Torr using 1 meter teflon tubing.

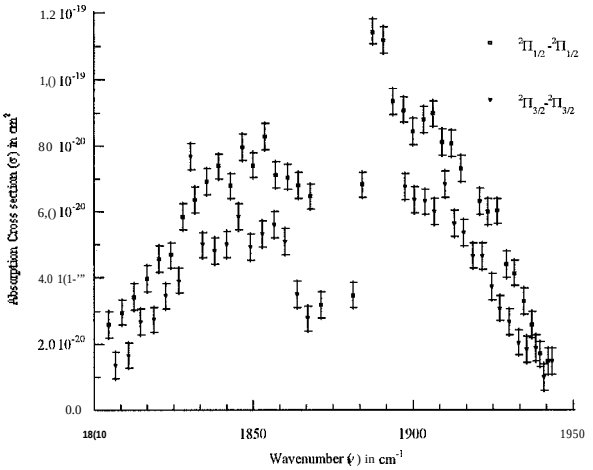


Fig. 4. Absorption Cross section (σ) versus Wavenumber (ν) for $^{14}N^{16}O$ when the gas was passed through Aluminum tubing.

Table 2. Absorption cross-sections' for the R and P branches of the $^2\Pi_{1/2}$ - $^2\Pi_{1/2}$ transition of $^{14}N^{16}O$ (under quasi-static conditions) when the gas was passed through different tubing materials

	v (cm^{-1})	Absorption Cross-Section (σ_v)× 10^{20} cm^2				v (cm^{-1})	Absorption Cross-Section (σ_v)× 10^{20} cm^2			
J	R(J)	Al	Cu	SS	Tef	P(J)	Al	Cu	SS	Tef
0.5	1881.07	3.47(39)	2.92(39)	2.67(39)	3.29(39)					
1.5	*884.34*	6.82(39)	5.18(39)	5.48(39)	5.40(39)	1871.08	3.71(39)	2.98(39)	2.77(39)	3.09(55)
2.5	*887.60*	11 .46(39)	*8.29(39)*	*8.44(39)*	*8.71(39)*	1867.73	6.46(39)	5.73(39)	4.89(39)	5.63(39)
3.5	*890.81*	11 .22(39)	*9.20(39)*	*8.86(39)*	*9.36(39)*	1864.25	6.79(39)	5.56(39)	4.96(39)	5.95(39)
4.5	*893.94*	*9.34(39)*	*7.71(39)*	*7.62(39)*	*8.61(39)*	1860.79	7.05(39)	5.88(39)	5.98(39)	6.17(39)
5.5	*897.04*	*9.08(39)*	*6.92(39)*	*7.00(39)*	*7.40(39)*	1857.29	7.13(39)	6.23(39)	5.43(39)	6.17(39)
6.5	900.10	*8.44(39)*	*6.93(39)*	*6.46(39)*	*7.18(39)*	1853.76	8.26(39)	6.72(39)	6.48(39)	6.60(39)
7.5	903.16	*8.78(39)*	*6.90(39)*	*6.67(39)*	*7.30(39)*	1850.19	7.41(39)	6.23(39)	5.90(39)	6.65(39)
8.5	906.18	*8.97(39)*	*6.89(39)*	*6.62(39)*	*7.21(39)*	1846.60	7.95(39)	6.23(39)	5.94(39)	6.36(39)
9.5	909.16	*8.10(39)*	*6.37(39)*	*6.08(39)*	*7.02(39)*	1842.96	6.78(39)	5.68(39)	5.62(39)	6.24(39)
10.5	912.10	*8.08(39)*	*6.61(39)*	*6.09(32)*	*6.84(39)*	1839.30	7.37(39)	6.46(39)	5.64(39)	6.58(39)
11.5	915.01	*7.32(62)*	*6.05(62)*	*5.77(39)*	*6.30(62)*	1835.59	6.93(62)	4.84(62)	5.35(39)	5.62(62)
12.5	917.95	*25.38(39)*	*8.30(39)*	*7.13(39)*	*9.51(39)*	1831.85	6.35(39)	5.06(39)	5.11(39)	5.15(39)
13.5	920.74	*6.33(39)*	*4.76(39)*	*4.75(39)*	*5.04(39)*	1828.08	5.83(39)	4.20(39)	4.61(39)	4.97(39)
14.5	1923.53	6.02(39)	5.09(39)	4.16(39)	5.1 1(39)	1824.27	4.68(39)	4.20(39)	4.13(39)	4.10(39)
15.5	1926.32	6.02(39)	4. 18(39)	3.79(39)	4.34(39)	1820.44	4.56(39)	3.42(39)	3.12(39)	3.54(39)
16.5	1929.06	4.40(39)	3.87(39)	3.55(39)	3.67(39)	1816.56	3.97(39)	3.32(39)	3.03(39)	3.17(39)
17.5	1931.75	4.15(39)	2.92(39)	2.93(39)	3.14(39)	1812.66	3.42(39)	3.10(39)		3.05(39)
18.5	1934.42	3.30(39)	2.48(39)	2.58(39)	2.45(39)	1808.72	2.96(39)	2.08(39)		2.28(39)
19.5	1937.05	2.57(39)	2.02(39)	1.99(39)	2.93(39)	1804.75	2.60(39)	1.88(39)		
20.5	1939.64	1 .70(39)	1 .70(39)			1800.74		1.77(39)		
21.5	1942.23	1.48(39)	1.48(39)							

'same Footnote as in Table 1.

calculated from the slopes of the plots of absorbance versus pressure and are given in Table 5 for the three modes of vibration. An evaluation of the data presented in Table 3 indicates that there was no appreciable variation in the absorption cross-section when the gas was passed through four different materials confirming that the SO_2 gas is fairly inert compared to HCl and NO, as far as reactivity with aluminum, copper, stainless-steel and teflon are concerned.

CONCLUSION

Hydrogen chloride, nitric oxide and sulfur dioxide are trace gases that are present in the atmosphere and are important for the understanding of stratospheric ozone depletion, global warming and smog formation. A sensitive Fourier Transform infrared spectrometer was used to measure the absorbance of HCl, NO and SO_2 for various pressures when the gases were passed separately through four different 1-m long tubings of aluminum, copper, stainless-steel and

Table 3. Absorption cross-sections' for the rotational transition associated with the ν_1, ν_2, and ν_3 vibrational modes of SO_2 (under quasi-static conditions) when the gas was passed through different tubing materials

	v (cm-l)	Absorption Cross-Section $(\sigma_v)\times 10^{20}$ cm^2				v (cm^{-1})	Absorption Cross-Section $(\sigma_v)\times 1020$ cm^2			
No	ν_1 mode	Al	Cu	SS	Tef	ode	Al	Cu	SS	Tef
1	1122.279	1.71(7)		1.673(3)	1.79(8)	:9?301	2.7(9)	2.56(1)	3.183(7)	3.021(8)
2	1139.683	2.66(7)	2.66(7)	2.62(7)	2.78(9)	497.148	3.15(7)	3.5 (1)	3.10(7)	3.94(7)
3	1163.307	2.97(7)	3.33(7)	3.35(7)	3.48(9)	503.209	4.01(7)	3.77(6)	3.46(7)	4.37(9)
4	1173.346	2.34(7)	2.25(7)	2.30(7)	2.46(9)	522.773	1.87(9)	1.9 (1)	1.54(9)	2.34(9)

No	ν_3 mode	Al	Cu	SS	Tef
1	1342.679	2.14(7)	2.07(7)	2.11(7)	2.18(9)
2	1349.671	2.35(7)	2.28(7)	2.28(7)	2.40(9)
3	1356.880	2.10(7)	2.05(7)	2.04(7)	2.14(9)
4	1371.690	3.04(7)	2.95(7)	2.96(7)	3.14(9)

'Same Footnote as in Table 1.

teflon. The absorption cross-sections were determined from the plots of absorbance versus pressure for the observed rotational lines of the fundamental bands of HCl, NO and SO_2. The plots confined the validity of Beer's law for the pressure range used. Detectable differences in absorbance and absorption cross-sections for the various tubing-gas combinations were observed. For the HCl molecule, based on the values of the absorption cross-sections, it was concluded that metallic materials (i.e Al and Cu) showed a greater change in absorption, whereas stainless steel and teflon tubings were comparatively inert. The absorption-cross sections calculated for the NO molecule showed that there was a greater change in absorption when the gas was passed through aluminum and teflon tubings, as compared to copper and stainless steel tubings. On the other hand, the absorption cross-sections calculated for the SO_2 molecule showed no appreciable variation in absorption when the gas was passed through aluminum, copper, stainless steel and teflon tubings. Thus, we can conclude that great care should be taken in reporting the absolute concentrations of trace molecules of relevance for atmospheric and stratospheric modeling considerations, since one must include the reaction of the gases with tubing materials employed in the inlet and outlet plumbing of the equipment used.

REFERENCES

1. Graedel, T. E. and Crutzen, P. J. 1989. The changing atmosphere. *Sci. Am. 261/9, 28-36.*
2. Bolin, B., Dőős, B. R., Jager, J., and Warrick, R. A., Eds. 1986. *The Greenhouse Effect, climatic Change, and Ecosystems.* Wiley, Chichester.
3. Schneider, S. H. 1989. The Changing Climate. *Sci. Am. 261/9, 38-47.*
4. Dickinson, R. E. and Cicerone, R. J. 1986. Future global warming from atmospheric trace gases. *Nature (London), 319, 109-115.*
5. Griffiths, P. R. 1975. Chemical Infrared Fourier Transform Spectroscopy, *Chem. Anal., 43,* Wiley, New York.
6. Theophanides, T., Ed. 1984. *Fourier Transform Infrared Spectroscopy,* Reidel, Dordrecht, The Netherlands.
7. Almond, M. J. and Downs, A. J. 1989. Spectroscopy of matrix isolated species. *Adv. Spectrosc. 17.*
8. Kuritsyn, Yu. A. 1985. Infrared absorption spectroscopy with tunable diode lasers. In *Laser Analytical Spectrochemistry* (V.S. Letokhov, Ed.), Chapter 4. Adam Hilger, Bristol.
9. Gristar, R., Bőttner, H., Tacke, M., and Restelli, G., Eds. 1992. *Monitoring of Gaseous Pollutants by Tunable Diode Lasers.* Kluwer Academic, Drodrecht, The Netherlands.
10. Schiff, H. 1992. Ground based measurements of atmospheric gases by spectroscopic methods. *Ber. Bunsenges. Phys. Chem. 96,296-306.*
11. Fried. A., Sams, R., and Berg, W. W. 1984. Application of tunable diode laser absorption for trace stratospheric measurements of HCl: Laboratory results. *Applied Optics, 23/11,* 1867-1880.
12. Aw-Musse, A. 1997. Fourier Transform Infrared (FT-IR) Spectroscopy of Trace Gases HCl And NO of Relevance to Atmospheric Phenomena. M.S. Thesis, Howard University, Washington, D.C.
13. McDermid, I.S. and Landenslager, J. B. 1982. *J. Quant. Spectrosc. Radiat. Transfer, 27,483.*
14. Mankin, W. G., Coffey, M. T., and Goldman, A. 1992. *Geophys. Res. Lett., 19, 179-182.*
15. Goodman, J., Snetsinger, K. G., Pueschel, R. F., Ferry, G. V., and Verma, S., 1994. *Geophys. Res. Lett., 21,* 1129-1132.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support received from the *Center for the Study of Terrestrial and Extraterrestrial Atmospheres* (CSTEA) (grant # NAGW-2950), NASA Lewis Research Center (grant # NAG3-1677) and U.S. Environmental Protection Agency's Office of Exploratory Research (grant # R819720-0 1-3).

Stimulated Raman amplification, oscillation, and linewidth in barium nitrate

Christopher L. McCray and Thomas H. Chyba
Research Center for Optical Physics
Department of Physics
Hampton University
Hampton, Va. 23668
(757) 727-5922 Phone
(757) 727-5955 Fax
cmccray@gprc.hamptonu.edu

Abstract

Measurements of Raman gain in a $Ba(N0_3)_2$ crystal are reported at 532 nm using a Raman oscillator/amplifier arrangement for differential absorption lidar measurements of ozone. The experimentally determined gain coefficient will be compared with theoretical results. The effect of single and multi-longitudinal mode pumping upon the amplification process will be discussed. Measurement of the Raman linewidth for 1 st, 2nd and 3rd stokes shifts arc presented.

Introduction

The Raman shifting method has been used to help widen the spectral coverage of lasers for various applications. Raman shifting is based on stimulated Raman scattering. This is an inelastic scattering process in which light from a laser at wavelength λ_1 is converted into light at another wavelength λ_2, which is the result of excitation or deexcitation of an internal resonance mode in the Raman medium. Radiation at the shifted wavelength λ_2 experiences gain at the expense of the incident laser radiation. Raman scattering occurs in liquids, gases, solids and plasmas. It can involve transitions between electronic states, rotational or molecular vibrations, lattice vibrations, spin states in semiconductors or electron waves in plasmas. Raman scattering is a two photon process and obeys the selection rules for two successive dipole transitions [1].

The first single pass Raman scattering in crystalline barium nitrate, $Ba(NO_3)_2$, pumped at 1064 nm by a Q-switched frequency doubled, Nd:glass laser was reported in 1980 [2]. Later work was published on Raman scattering in $Ba(NO_3)_2$ pumped at 532 nm by a Q-switched frequency-doubled, Nd:YAG laser **[3,4].** Preliminary experiments [5] at NASA LaRC suggested that a $Ba(NO_3)_2$ Raman laser could serve as a compact solid state replacement to other existing lasers for uv lidar measurements. A $Ba(NO_3)_2$ Raman crystal pumped at 532 nm by a frequency doubled Nd:YAG laser will create first Stokes (shifted wavelength) laser output at 563 nm and second Stokes output at 599 nm with very high efficiencies.

Preliminary results on a compact visible $Ba(NO_3)_2$ Raman laser have been obtained. From these initial experiments we obtained a total conversion efficiency greater than 65% at 30 Hz PRF for first and second Stokes [6]. Optimization of the $Ba(NO_3)_2$ laser system can be accomplished by modifying the laser design to better mode match the pump beam with the resonator mode, enabling the pump photons to make more passes in the resonator before leaving the resonator, selecting the mirror reflectivities to suppress unwanted Stokes lines, and by accounting for the thermal lensing and birefringence effects in $Ba(NO_3)_2$.

Laboratory measurements with the oscillator have shown that output beam divergence increases with pump power. Also, it becomes increasingly difficult to suppress unwanted Stokes lines. To address these difficulties, we have conducted preliminary measurements with an oscillator/amplifier arrangement.

Experiment

The experimental setup used to measure the Raman gain coefficient in $Ba(NO_3)_2$ is shown in Fig. 1. The pump beam is from a frequency-doubled, multimode, flashlamp- pumped Nd:YAG laser. The temporal pulsewidth of the laser is 8 ns. For a multimode Nd:YAG there is no significant spectral broadening due to the Raman medium. The beam is divided in half by a beam splitter and then passed through two attenuators in order to separately vary the pump energies to the oscillator and amplifier. The pump beam for the oscillator is collimated with a telescope. The $Ba(NO_3)_2$ Raman oscillator is the same configuration as used before in earlier experiments [6,7]. A spatial filter is used to make the beam more collimated before entering the Raman amplifier. Spatial filtering significantly limits the Raman seed energy. A delay line is created in the second path using two mirrors in order to better match the 563 nm and 532 nm beams temporally. They are then sent collinearly through a single pass $Ba(NO_3)_2$ amplifier and then to a detector. Measurements of beam size and spatial profile were made using a spatial profiler. The spatial profiles of the pump and seed beams were essentially gaussian, with beam diameters equal to 1.7 mm and 1.8 mm, respectively.

For two values of incident 1st Stokes seed energy into the amplifier, the amplified 1st Stokes energy was measured as a fiction of pump power. Preliminary data shown in Fig. 2 indicates a Raman gain of approximately 1600 for a seed energy of 2 μJ and a Raman gain of 100 for a seed energy of 35 μJ.

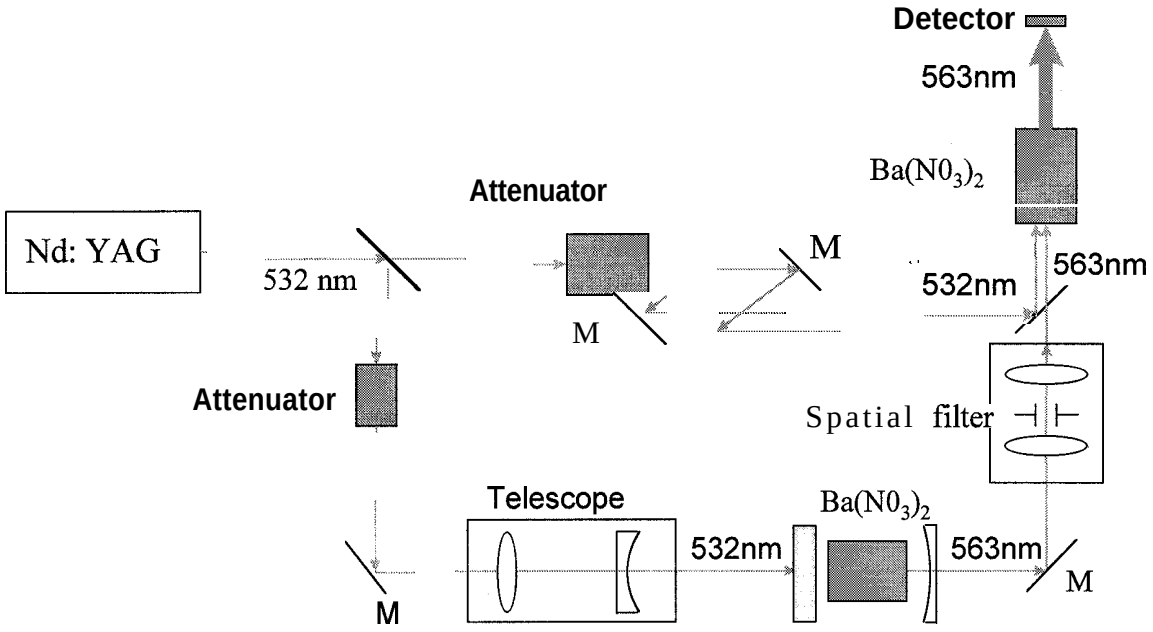


Fig 1. Schematic of the oscillator/amplifier.

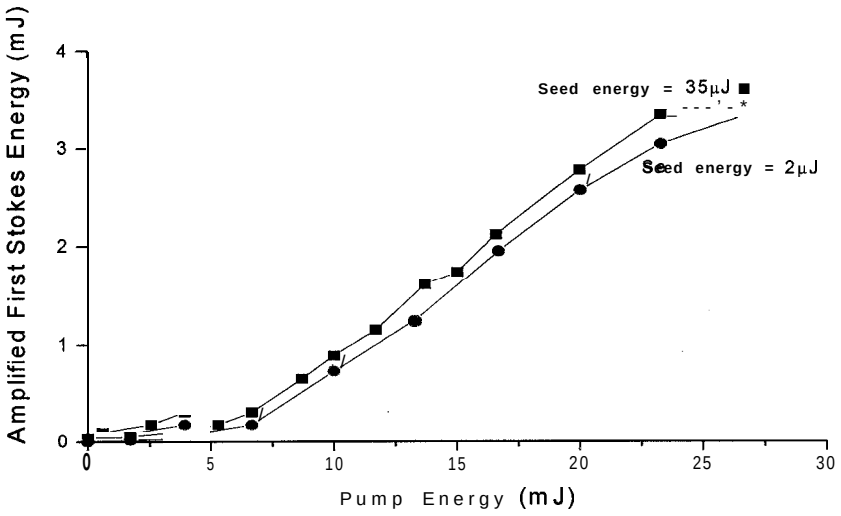


Fig. **2.** Results of oscillator/amplifier. A seed energy of **35** μJ yields a gain of 100 and a seed energy of 2 μJ yields a gain of 1600.

To properly determine the Raman gain coefficient, the spatial, temporal, and spectral properties of the pump and seed beams need to be taken into account [3].

Linewidth Broadening

Knowledge of linewidth broadening is important in the development of amplifiers and nonlinear converters. Crystalline barium nitrate has a strong Raman line at a frequency of about 1047 cm^{-1}. Recently, the room temperature linewidth has been measured to be 0.56 $\pm$ 0.1 cm^{-1} [8]. However, earlier values reported in the literature include 0.3 $\pm$ 0.12 cm^{-1} [9] and 1.5 $\pm$ 0.5 cm^{-1} [3]. We investigated linewidth broadening in barium nitrate in a laser oscillator at room temperature, for three Raman shifts (563, 599, and 639 nm) from a 532 nm Nd:YAG pump laser.

The experimental setup used to measure the Raman line width broadening is shown in Fig. 2. This setup is similar to the schematic used for Raman amplification. The pump beam is from a frequency doubled Nd:YAG laser (Continuum NY-80 C). This laser is injection seeded to operate in a single longitudinal mode. The temporal pulsewidth of the laser is 10 ns.

The Raman beam is collimated after the resonator by a 300 mm lens. Two brewster angle prisms are used to separate the Stokes and fundamental wavelengths. A Burleigh pulsed wave meter, WA -5500, is used to determine the wavelength and linewidth of each stokes shift.

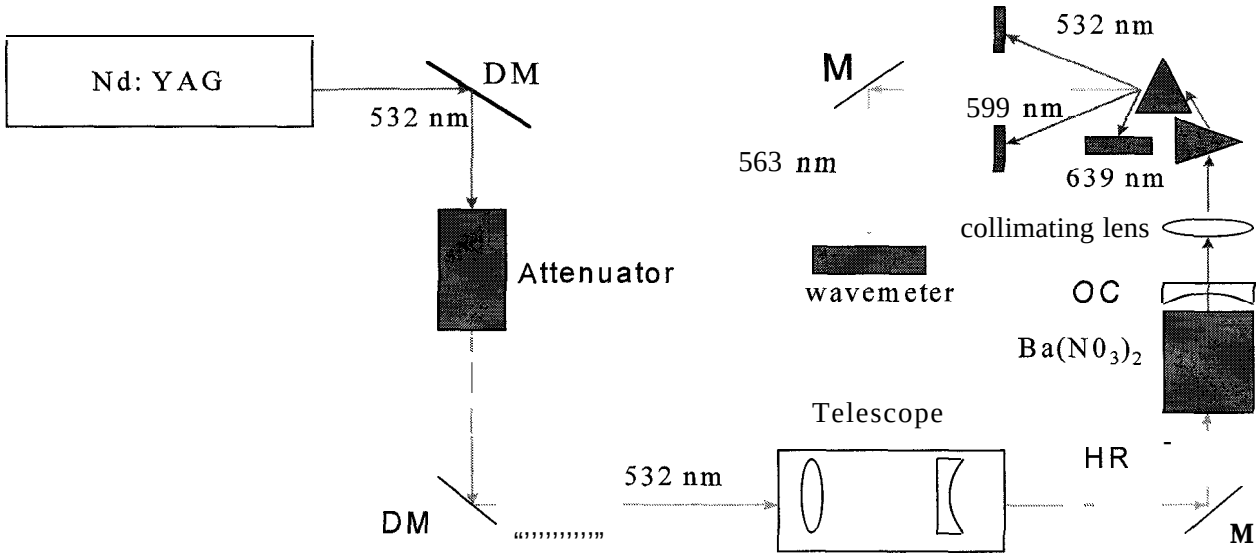


Fig 3. Schematic for measurement of linewidth broadening in $Ba(N0_3)_2$

The $Ba(NO_3)_2$ resonator was first pumped with a single longitudinal TEM_{00} mode with linewidth, $\Delta\upsilon_{LS}$ = 0.0045 cm^{-1} (manufacturer's specification). The wavemeter has a spectral resolution, $\Delta\upsilon_w$ of approximately 0.3 cm^{-1} (manufacturer's specification). Since the wavemeter's etalon coatings are optimized for 190-680 nm, we decided to measure its resolution directly using two lasers with linewidths much smaller than this value: the injection seeded Nd:YAG laser at 532 nm and a single longitudinal mode He:Ne laser at 632 nm. When measured with the wavemeter, each had a linewidth of 0.146 cm-l. We therefore adopted this value as the wavemeter resolution, $\Delta\upsilon_W$. Since all the spectral profiles are Lorentzian, the Raman contribution, $\Delta\upsilon_R$, to the total linewidth, $\Delta\upsilon_T$, can be calculated by subtraction of the wavemeter linewidth from the total linewidth: $\Delta\upsilon_R = \Delta\upsilon_T - \Delta\upsilon_W$. The results of this experiment are shown in Table 1. Each value represents the mean of 20-40 measurements. The stated uncertainty is the standard deviation of those measurements.

Stokes Wavelengths	Raman Broadened Linewidth ($\Delta\upsilon_R$)
563 nm	0.055 cm-" $\pm$ .01 cm^{-1}
599 nm	0.069 cm^{-1} $\pm$ .01 cm^{-1}
639 nm	0.081 cm^{-1} $\pm$ .03 cm^{-1}

Table 1. Stokes wavelengths and their Raman broadened linewidths

This experiment was repeated for the case of the multimode pump. While the values obtained are consistent with those in the Table, the large fluctuations in the linewidth of the multimode Nd:YAG laser give rise to uncertainties too large to allow a precise measurement of the Raman broadening. These values are significantly less than others reported recently. This effect may be due to the presence of the oscillator and will be explored further.

Acknowledgments

We acknowledge support from NASA Langley Research Center, NASA grants NCC-1-215 and NAGW 2929, the National Physical Science Consortium (NPSC), and the Virginia Space Grant Consortium for this project. We thank Sangwoo Lee for assistance with the measurements and Dr. B. Tabibi and C. Smith for use of their wavemeter.

References

1, A. Penzkofer, A. Laubereau, and W. Kaiser, "High Intensity Raman Interactions," *Principles of Quantum Electronics 6, 55-140 (1979)*

2. A. S. Eremenko, S. N. Kapukhin, and A. I. Spepanov, "Stimulated Raman scattering of the second harmonic of a neodymium laser in nitrate crystals," *Sov* J. *Quantum Electron.* **10,** 113-*114(1980)*

3. S. N. Karpukhin, and V. E. Yashin "Stimulated emission and amplification of radiation under conditions of stimulated Raman scattering in crystals," *Sov. J. Quantum Electron.* **14,** 1337-*1342 (1984)*

4. P. G. Zverev, J. T. Murray, R. C. Powell, R. J. Reeves and T. T. Basiev, "Stimulated Raman scattering of picosecond pulses in barium nitrate crystals," *Optics Comm. 97, 59-64(1993)*

5. D. Richter, W. Marsh and J. Barnes, private communication 1995

6. C. He and T. H. Chyba, "Solid -state barium nitrate Raman laser in the visible region," *Optics Communications 135,273-278 (1997)*

7. C. L. McCray and T. H. Chyba, "Development of an All-Solid State Raman Laser for Ozone DIAL Measurements," *NASA Universities Research Centers Technical Advances on Education, Aeronautics, Space, Autonomy, Earth and Environment.* **1497-500** *(1997)*

8. P G Zverev, and T T Basiev, "Investigation of the line Broadening of an SRS-active vibration in a barium nitrate crystal by two photon Raman amplification spectroscopy," *Quantum Electronics 25 (12) 1204-1207 (1995)*

9. P G Zverev, and T T Basiev, "Barium Nitrate Raman Laser for Near IR Spectral Region", *OSA Proceedings on Advanced Solid-State Lasers, 1995,* Vol. 24, B. Chai and S. Payne, (eds.), 288-295 (1995)

Structural Loading On The QCM/SAW Instrument Aboard The ER-2 Used for Atmospheric Testing+

Peter M. Bainum*, Phyllis D. Jones", Sandra M. Irish**, and Guangqian Xing*
*Department of Mechanical Engineering, Howard University, Washington, D.C. 20059
**Mechanical Systems Analysis Group, NASA Goddard Space Flight Center, Greenbelt, MD. 20771

Introduction

Several experiments have been proposed to capture and evaluate samples of the atmosphere where SST's travel. One means to achieve this is to utilize the quartz crystal microbalance (QCM)/ surface acoustical wave (SAW) instrument installed aboard the ER-2, formerly the U-2 reconnaissance aircraft. The QCM is a cascade impactor designed to perform in-situ, real-time measurements of aerosols and chemical vapors at an altitude of 60,000-70,000 feet[1]. The primary use of the ER-2 is by NASA for Earth resources to test new sensor systems before being placed aboard satellites. One of the" main reasons the ER-2 is used for this flight experiment is its capability to fly approximately twelve miles above the sea level (can reach an altitude of 78,000 feet)[2]. Because the ER-2 operates at such a high altitude, it is of special interest to scientists interested in space exploration or supersonic aircraft. The purpose of some of the experiments is to extinct data from the atmosphere around the ER-2. For the current CSTEA flight experiment, the housing of the QCM is in a frame that connects to an outer pod that attaches to the fuselage of the ER-2. Due to the location of the QCM within the housing frame and the location of the pod on the ER-2, the pod and its contents are subject to structural loads. In addition to structural loads, structural vibrations are also of importance because the QCM output data is based on the determination of beat frequencies between a pair of oscillators (one coated, the second uncoated, according to the chemical reaction being monitored). A structural analysis of this system can indicate whether potential resonances may exist between the (higher) structural modal frequencies and the beat frequencies. In addition undesirable deformations may result due to maximum expected static or dynamic loads during typical flight conditions. If the deformations are excessive they may adversely affect the accuracy the instrumentation output.

Structural Analysis Procedure

In order to perform a structural analysis of the QCM within its housing frame, a finite element model of the system (QCM and housing frame) was generated. Initially, separate finite element models of the QCM and housing frame were constructed, Figures I and H, respectively. The models were graphically generated by the finite element graphical software package MSC/PATRAN[3]. By inputting the structure's geometry, material properties, inertia, and boundary conditions, MSC/PATRAN has the capabilities to generate a finite element model. The geometry consists of

'Research supported by the NASA/Howard University Center for the Study of Terrestrial & Extraterrestrial Atmospheres (CSTEA).

the physical make-up of the structure and its dimensions are, for the QCM, length= 18.3 in., width = 4.5 in., and height= 4.8 in. and for the housing frame length= 66.5 in., width= 12.5 in., and height = 13.2 in. The material of both the QCM and housing frame is aluminum with material properties of 107 lb/in for Young's modulus, 0.333 for Poisson's ratio, and 0.098 lb/in^3 for the density. The QCM rests lengthwise on an aluminum plate within the frame and is rigidly bounded at both ends.

Finite Element Model of the QCM

The QCM was subdivided into three sections to simplify its input into MSC/PATRAN because of its anti-symmetry. The sections selected were the top base, beams and bottom base. The geometrical model of the structure was initially generated in MSC/PATRAN. The model consists of each section's length, width, and height. Therefore, from this model, the physical composition of the structure was created. The model's grid points are the minimum number of coordinates required to construct the geometric model with respect to the origin. The sections were input independently with respect to each other but not with respect to their respective origin. The difference between grid points and nodal points is that the grid points are the minimum number of coordinate points to define the structure's geometry whereas the nodal points are connectivity points between the finite elements. After the input of each section, the geometrical model of the QCM was complete with 86 grid points. The finite element model is created from the geometrical model. The finite element model consists of the structure's nodal points (nodes) as well as the finite elements. MSC/PATRAN automatically generates the finite element model based on the given information from the geometric model. The length between each node is selected as 0.5 inches. For this structure, there are 2,756 nodes and 1,461 finite elements. The types of finite elements chosen are CHEXA solid polygonal elements possessing three translational degrees of freedom.

Finite Element Model of the Housing Frame

Also, a finite element model of the housing frame was generated. Within this structure, there are two different types of beams and a support plate. One type of beam is a C-Channel used along the sides of the structure and the other is an L-Bracket used in the mid-section. As with the QCM, the housing frame and the support plate are composed of aluminum; therefore, they possess the same material properties as the QCM. For the support plate, the thickness of the plate is 0.25 inches. As with the QCM, the finite element model is created from the geometric model. For the housing frame, there are 35 geometrical points, 178 nodes, and 197 finite elements. The finite element used for the plate is the CQUAD4. This solid element was chosen because of its general purpose for plate elements capable of carrying in-plane forces, bending forces, and transverse shear forces. CQUAD4 elements possess five degrees of freedom. The sixth one, the rotational degree of freedom, remains completely constrained in all directions to prevent singularities. In MSC/NASTRAN, the assumptions for CQUAD4 finite elements are as follows:

- A thin plate is one in which the thickness is much less than the next larger dimension,
- The deflection of the plate's midsurface is small compared with its thickness,
- The midsurface remains neutral (longitudinal fibers do not change in length) during

bending, and

- The normal to the midsurface remains normal during bending.

For the remainder of the structure, the L-Brackets and C-Channels, the finite element used is the CBAR. CBARs are straight, prismatic elements with axial, bending, and torsional stiffness. The CBAR was chosen because it is a general purpose beam that supports tension and compression, torsion, and bending in two perpendicular planes and shear in two perpendicular planes. CBARs can exhibit six degrees of freedom at each node. The assumptions for the CBAR are as follows:

- Plane sections remain plane,
- It must be straight and prismatic (properties cannot vary along the length),
- The shear center and neutral axis must coincide, and
- The principal moment of inertia need not coincide with the element axis.

Numerical Results

Vibration Frequencies

Table I shows the distribution of the frequencies corresponding to the structural modes of vibration according to increasing mode number for the QCM/Housing Frame System. It can be seen that there is a three order of magnitude variation between the fundamental vibration frequency of 30.159 Hz and the highest vibrational modes. The quartz crystals within the QCM have a natural vibration frequency of 10 MHZ and the beat frequency between each pair of oscillators ranges from 3kHz to 9 kHz[1]. Thus, there is some concern about possible internal resonance between the higher modes of structural vibration and the beat frequent y.

Mode Number	Frequency (Hz)
1	30.159
2	67.713
3	97.214
4	108.108
*	*
107	1873.468
108	1889.484
109	1893.380
110	1942.543

Table 1: Structural Modes of Vibration

Response to Static Loads

According to the ER-2 flight profile[4] the maximum static load experienced is 8G (8 times the normal sea level gravitational loading). The maximum expected deformations in response to the

8G static loading are minute, of the order of thousandths of an inch, see Table II. Fig. II shows (an exaggerated) view of the deformed system in response to an 8G static load in the +Y- Direction - direction of flight.

However, for the special case where the system is assumed to be vibrating at only a single modal frequency, the maximum displacement expected from an 8G static load could reach as much as 2-6 inches (e.g. mode 16 or mode 10 at node 255 at the rear of the housing frame) and fractions of "an inch at node 343 (immediately in front of the QCM on the housing frame).

	X-DISP (in)	Y-DISP (in)	Z-DISP (in)
+X-DIRECTION	1.7612E-2	2.7612E-4	3.2778 E-3
+Y-DIRECTION	2.7509E-4	2.6276E-3	1.0284E-3
+Z-DIRECTION	1.8966E-3	1.1070E-3	1. 1080E-3
-X-DIRECTION	0.0	2.7539E-4	3.0651 E-3
-Y-DIRECTION	2.7604E-4	0.0	5.8118E-4
-Z-DIRECTION	7.0554E-4	9.0979E-4	0.0

Table II: Deflections from 8G Static Load

Response to Dynamic Load

An example of dynamic loading that maybe applied to the QCM/Housing Frame could result from wind gusts/wind shear. In the upper atmosphere (lower stratosphere) wind gusts can reach velocities of at least 50 ft/sec. Over the SST North Atlantic flight corridor[5]. Periodic wind gust can be represented in the form of an acceleration load by

$$A_g = v\,\omega \sin(\omega t)$$

where v = velocity amplitude and ω = frequency. For the special case where ω is tuned to the fundamental or harmonics of the natural vibration frequency a possibility of externally induced resonance may occur.

Fig. 111 shows the deformation at node 343 immediately in front of the QCM on the housing frame, as a function of time for one cycle of wind gust at the same frequency as the fundamental vibrational mode of the system. It is seen that for this special case deformations of the order of 2 inches could occur. It should be noted that, although this case depicts a truly "worst case" scenario, deformations of this magnitude could impede the air flow and amount of air particulate samples entering into the orifice of the QCM stack, whose diameter is 2.36 inches[1].

Conclusions and Recommendations

Results from the structural analysis indicate several orders of magnitude separation between the vibration frequency of the quartz crystals within the QCM and the vibrational modes of the system. However, there is a possibility of resonance between the beat frequency and the higher modes of vibration. The expected deformation range in response to the maximum expected static loading of 8G's is of the order of thousandths of an inch. However, dynamic loading induced by

wind guests tuned to the fundamental (and lower) frequencies could produce deformations of the order of inches - which might affect the accuracy of the QCM readings. To compensate for such a situation it is suggested that a damped vibration absorber could be designed and attached to specific nodal locations where the largest deformations could occur[6],[7].

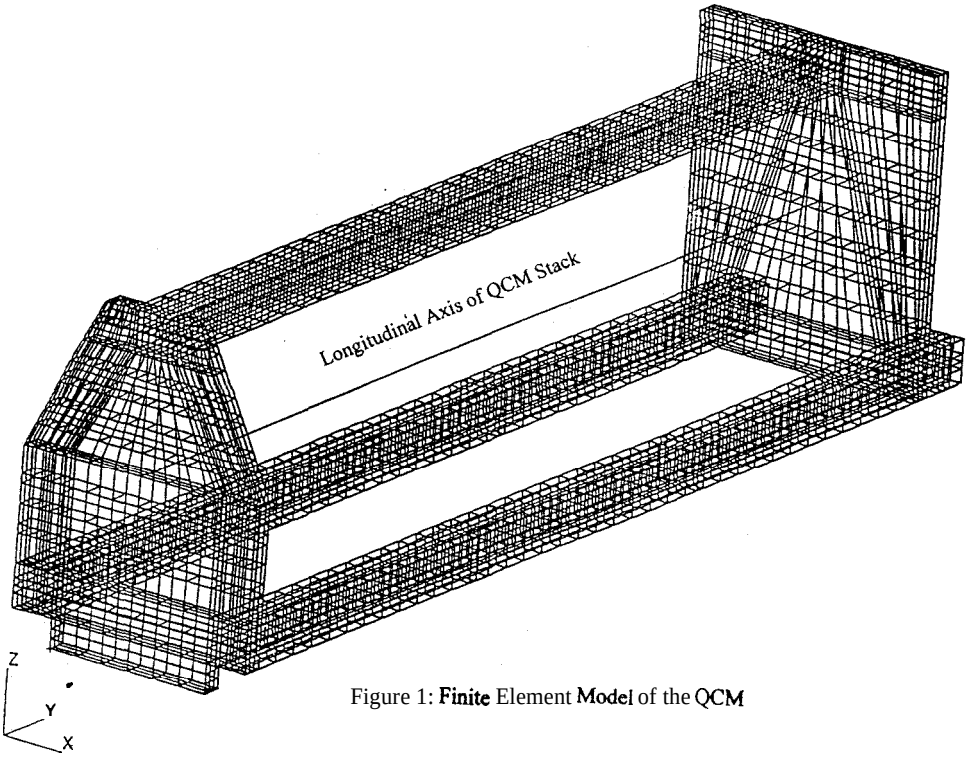


Figure 1: Finite Element Model of the QCM

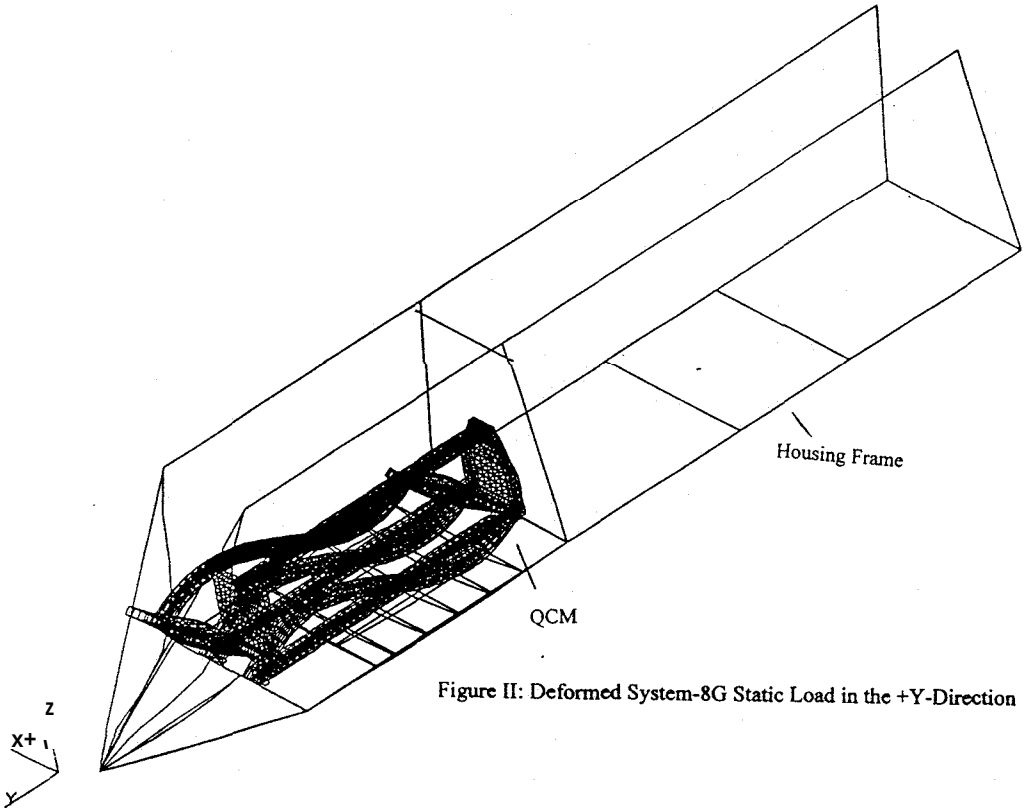


Figure II: Deformed System-8G Static Load in the +Y-Direction

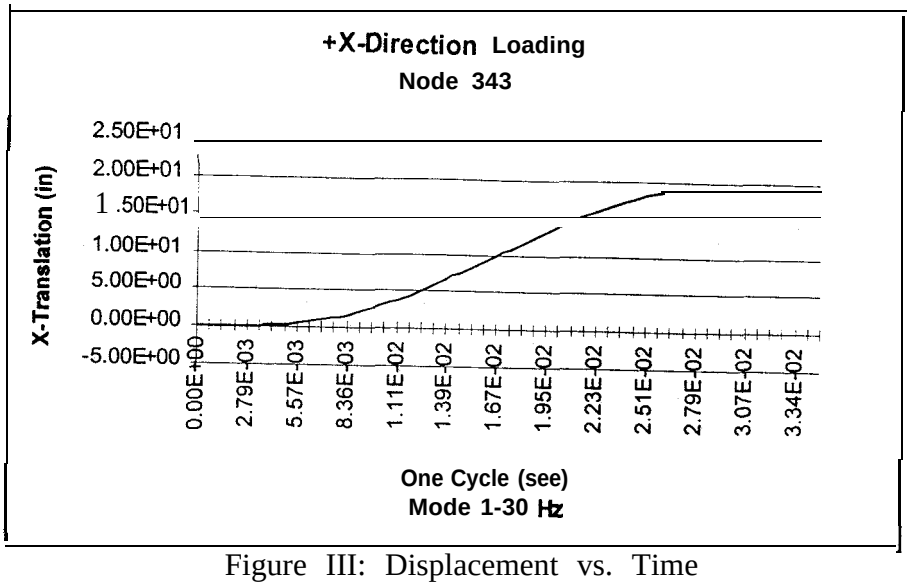


Figure III: Displacement vs. Time

References

[1] Jones, Phyllis, "Description of the QCM and its Use in the Concorde/ER-2 Flight Experiment," Technical Note, Dept. Of Aerospace Engineering Sciences, University of Colorado Boulder, CO., August 1994.

[2] Stewart, Doug, "Above the Sky," Air and Space, August/September 1993, pp. 11-13.

[3] Miller, Mark, User's Guide: Getting Started with MSC/NASTRAN, 1st Ed., MacNeal Schwendler Corp., Los Angeles, CA, December 1993,

[4] ER-2 Investigator's Handbook, Vol.I, High Altitude Missions Branch, Airborne Missions & Applications Division, NASA Ames Research Center, Moffett Field, CA, 94035, May 1988.

[5] Scott, Leslie, British Airways Concorde Captain, Briefing to Howard University AIAA Student Chapter, John F. Kennedy Airport, New York, N. Y., April 20, 1996.

[6] Marsis, Wisjnu and Bainum, Peter, "Hybrid Control System for Space Mast Structures," AAS/AIAA Astrodynamics Specialist Conference, Durango, CO, August 19-22, 1991, Paper No. AAS-91-374.

[7] Hartog, Den, J. P., Mechanical Vibrations, McGraw-Hill, 4th Ed, New York, NY, 1956.

Photosynthetic Response to Long- and Short-Term Changes in Carbon Dioxide in Sweetpotatoes Grown Hydroponically with Enhanced Mineral Nutrition

Casey Hamilton[1], Anita Terse[1], Douglas R. Hileman[1],
Desmond G. Mortley[2] and Jill Hill[2]

[1]Department of Biology and [2]Department of Agricultural Sciences
Tuskegee University, Tuskegee, AL 36088

ABSTRACT

Sweetpotato [*Ipomoea batatas* L.(Lam.)] has been selected by NASA as a potential food for long-term space missions. In previous experiments, sweetpotato plants grown hydroponically under elevated levels of CO_2 depleted the nitrogen in the nutrient solution between the hi-weekly solution replacements. In this experiment, the effect of enhanced nutrient replenishment on photosynthetic rates of sweetpotato was determined. CO_2 response curves were determined for 'TU-82-155' and 'Georgia-Jet' sweetpotatoes grown hydroponically in growth chambers at three different CO_2 concentrations (400, 750, and 1000 $\mu mol\ mol^{-1}\ CO_2$). Gas exchange measurements were made using infrared gas analysis, an open-flow gas exchange system, and a controlled-climate cuvette. Photosynthetic measurements were made at CO_2 concentrations from 50-1000 $\mu mol\ mol^{-1}\ CO_2$. Net photosynthetic rates showed an increase with increasing measurement CO_2 in all nutrient regimes, but the response of photosynthetic rates to the growth CO_2 conditions varied among the experiments and between the two varieties. Enhanced mineral nutrition led to increased net photosynthetic rates in 'Georgia Jet' plants, but not in 'TU-82- 155' plants. The results of this study will help to determine the CO_2 requirements for growth of sweetpotato on proposed space missions.

INTRODUCTION

NASA has been developing a controlled ecological life support system for use on extended space missions (Averner, 1989). During these missions crops will be grown for food as well as to regenerate oxygen. One of the crops that is being considered for space is sweetpotato [*Ipomea batatas* L. (Lam.)]. A system for growing sweetpotato hydroponically using a nutrient-film technique has been developed (Hill et al., 1989; Loretan et al., 1989; Mortley et al., 1991; Bonsi et al., 1992). One of the environmental conditions that must be considered in a controlled ecological life support system is the level of carbon dioxide. The concentration of CO_2 directly affects photosynthesis, which can in turn affect plant growth. In a closed life support system, the concentration of CO_2 may rise or fall depending on the ratio of plant material to inhabitants.

Numerous studies related to global climate change have been carried out on the effects of a doubling of atmospheric CO_2 concentration on photosynthesis in agricultural crops (Allen, 1990; Woodward et al., 1991). In general, photosynthetic rates increase as the CO_2 concentration surrounding the leaf, and thus the CO_2 concentration inside the leaf, increases, especially at low CO_2 concentrations (Farquhar and Sharkey, 1982; Woodward et al., 1991; Wullschleger, 1993). However, growth of plants at elevated CO_2 concentrations often leads to lower photosynthetic rates when these measured at a common CO_2 concentration, a phenomenon usually referred to as "acclimation" or "down-regulation" (Bowes, 1993). When measured at the same CO_2 concentration that the plants are grown under, photosynthetic rates usually increase with CO_2 enrichment, because the increase in photosynthetic rates due to higher leaf internal CO_2 is usually greater than the decrease in photosynthetic rates due to growth at elevated CO_2 (Cure and Acock, 1986).

Field studies have shown that both yield (Biswas et al.; 1996) and net photosynthetic rates (Strachan, 1993) of sweetpotato respond strongly to a doubling of the earth's current CO_2 concentration. However, inside a closed ecological life support system, CO_2 concentrations may reach levels that are several times the earth's current atmospheric concentration. Only a few studies have looked at photosynthetic rates at these higher CO_2 levels (Wheeler et al., 1993).

Photosynthetic rates of plants can be affected by mineral nutrient status. Specifically, light-saturated photosynthetic rates often correlate strongly with leaf nitrogen levels (Field and Mooney,

1986; Evans, 1989). Preliminary analyses of the hydroponic solutions used to grow sweetpotatoes under CO_2 enrichment indicated that some nutrients, especially nitrogen, were being depleted between the biweekly solution replacements. Thus, we wanted to investigate the effects of nutrient supply on photosynthetic rates.

We studied the photosynthetic responses to long-term and short-term changes in CO_2 concentration using sweetpotato grown hydroponically at two levels of mineral nutrition. This study was designed to answer three questions. First, do photosynthetic rates of hydroponically-grown sweetpotato show the same positive response to elevated CO_2 that was seen in field-grown plants? Second, what effects does CO_2 enrichment to levels greater than double the current atmospheric concentration have on photosynthetic rates of sweetpotato? Third, does enhancing the mineral nutrient supply affect photosynthetic rates in hydroponically-grown sweetpotato?

MATERIALS AND METHODS

Plant Material, Growth Conditions and Experimental Design

The study was conducted at the Center for Food and Environmental Systems for Human Exploration of Space (CFESH), George Washington Carver Agricultural Experiment Station, Tuskegee University, Tuskegee AL, in three environmental growth chambers. Four plants each of the sweetpotato cultivar 'Georgia Jet' and the breeding clone 'TU-82-155' were planted in 0.15 x 0.15 x 1.2 m growth channels. One channel of each variety was placed in each of the three growth chambers. The growth chambers were maintained at a diurnal temperature of 28/22 'C, a 14/10 light/dark photoperiod, and a relative humidity of 70%. The light source in the chambers consisted of a mixture of incandescent and fluorescent lamps. Photosynthetically active radiation (PAR) levels were approximately 600 $\mu mol\ m^{-2}\ s^{-1}$ at the canopy level.

The three growth chambers were maintained at 400,750, and 1000 $\mu mol\ mol^{-1}\ CO_2$. The nutrient solution consisted of a modified half-Hoagland solution with a 1:2.4 N:K ratio. Solution pH was maintained between 5.5 and 6.0 by the addition of either dilute NaOH or HCl biweekly. Electrical conductivity (1. 1-1.2 $mS\ cm^{-1}$) was monitored biweekly. The solution was pumped continuously at 1.0 L m^{-1} from each 30 L reservoir to the high end of each channel by a magnetic drive pump.

Nutrient Replacement Protocols

In experiments 1 and 2, the entire nutrient solution was replaced every two weeks. In experiments 3 and 4, the nutrient solution was supplemented three times per week, leading to about a 70% increase in the nutrient supply. In experiment 3, the nutrient solution initially was replenished using 10x Hoagland stock solution. By day 72, the plants exhibited nutrient deficiency symptoms, and analyses showed depletion of nitrate and phosphate in the nutrient solution. Therefore, the nutrient solution was replaced and the replenishment was resumed using 1/3 Hoagland solution. In experiment 4, the nutrient solution was replenished using 1/2 Hoagland solution.

Photosynthetic Measurements

An open-flow gas exchange system with a controlled-climate leaf cuvette was used for all gas exchange measurements. Leaf temperature and relative humidity in the cuvette were maintained at 28 'C and 50%, respectively. PAR from the growth chamber lamps and from four projector bulbs (type EYF) attached to the leaf cuvette, was maintained at levels of at least 700 $\mu mol\ m^{-2}\ s^{-1}$. CO_2 response curves (A/C_i curves) were generated by mixing CO_2-free air with air containing 3000 $\mu mol\ mol^{-1}\ CO_2$, using mass flow controllers. The CO_2 concentration of the air entering the cuvette was reduced from 1000 to 50 $\mu mol\ mol^{-1}$ in 12 steps. Measurements were made using the 3^{rd} - 5^{th} unfolded leaf on a growing shoot tip. In experiments 1, 2, and 4, three replicate leaves were measured for each variety, in each treatment. In experiment 3, two replicate leaves were measured before and two after the change in the nutrient protocol at day 72.

Data Analyses

Data were analyzed by analysis of variance (ANOVA), with plant variety and growth CO_2 concentration as the main effects. Measurement CO_2 concentration was treated as a split-plot effect, since each A/C_i curve was determined on a unique leaf.

In experiments 1-3 net photosynthetic rates of the two varieties did not differ significantly. In these experiments, the data were pooled together with respect to variety. In experiment 3, where the nutrient replenishment regime was changed at day 72, preliminary analyses of the data

indicated that photosynthetic responses to CO_2 before day 72 differed from those after day 72. Therefore, the data from the two periods of experiment 3 were analyzed separately.

RESULTS

There were no significant CO_2 effects on photosynthetic rates, when measured at the growth CO_2 concentration, in experiments 1,3a, and 4 in either variety. In experiments 2 and 3b, photosynthetic rates increased significantly with increasing CO_2 concentration (Fig. 1). Thus, CO_2 enrichment led to increased photosynthetic rates in one of the two replicate experiments at each nutrient regime.

To separate the effects of measurement and growth CO_2 concentrations, the points from the A/C_i curves, corresponding to 400,750, and 1000 $\mu mol\ mol^{-1}$ measurement CO_2 were analyzed by ANOVA. Photosynthetic rates increased with increasing measurement CO_2 in all experiments (Figs. 2,3). There were no significant effects of growth CO_2 on photosynthetic rates in both experiments when the nutrient solution was replaced every two weeks (Fig. 2).

When the nutrient solution was replenished three times per week, growth CO_2 concentrations had variable effects on photosynthetic rates. In the first half of experiment 3, photosynthetic rates were greater in plants grown at 400 $\mu mol\ mol^{-1}$ CO_2 than at higher CO_2 concentrations. In the second half of experiment 3, plants grown at 750 $\mu mol\ mol^{-1}$ CO_2 exhibited greater photosynthetic rates than did plants grown at 400 or 1000 $\mu mol\ mol^{-1}$ CO_2. In experiment 4 the two varieties responded differently. Photosynthetic rates of 'TU-82- 155' plants did not differ among growth CO_2 treatments, while 'Georgia Jet' plants exhibited greater photosynthetic rates in plants grown at 400 $\mu mol\ mol^{-1}$ CO_2 than at higher CO_2 concentrations (Fig. 3).

Combined analyses of all four experiments showed that enhancing the mineral nutrition supply did not have a significant effect on net photosynthetic rates of 'TU-82- 155' sweetpotatoes (15. 1 vs. 15.9 $\mu mol\ mol^{-1}$ CO_2 m^{-2} s^{-1} for nutrient replenishment of three times per week vs. every two weeks, respectively). However, 'Georgia Jet' sweetpotatoes grown in nutrient solution that was replenished three times per week had significantly higher net photosynthetic rates than did plants grown in nutrient solution that was replaced hi-weekly (16.5 vs. 13.8 $\mu mol\ mol^{-1}$ CO_2 m^{-2} s^{-1}, respectively; $p=0.02$).

DISCUSSION

Photosynthetic rates of hydroponically-grown sweetpotato did not respond as strongly to CO_2 enrichment as did field-grown plants (Strachan, 1993). Photosynthetic rates measured at the growth CO_2 concentration increased significantly with CO_2 enrichment in only two of the four experiments. This weak photosynthetic response to CO_2 enrichment in hydroponically-grown plants is likely due to a combination of factors. The lower light levels in the growth chambers may have limited the response of the plants to CO_2 enrichment. Leaves grown under high light usually have higher photosynthetic capacities (Larcher, 1980), and the photosynthetic rates measured in our experiment were about half of those reported for field-grown plants (Strachan, 1993).

The increase in photosynthetic rates of sweetpotatoes with increasing measurement CO_2 concentration was expected, since photosynthetic rates are usually limited by CO_2 supply at lower CO_2 levels (Farquhar and Sharkey, 1982). The lack of a significant increase in photosynthetic rates between 750 and 1000 $\mu mol\ mol^{-1}$ CO_2 indicates that photosynthesis becomes CO_2-saturated at around 750 $\mu mol\ mol^{-1}$ CO_2 in sweetpotato grown under these conditions. Therefore, it seems unlikely that enrichment to CO_2 concentrations above 750 $\mu mol\ mol^{-1}$ will lead to significant increases in yield.

Acclimation to growth at elevated CO_2 (i.e. decreases in photosynthetic rates measured at a common CO_2 concentration) did not occur in plants grown with hi-weekly changes of the nutrient solution. When the solution was changed three times per week, acclimation was observed in the first half of experiment 3 and in 'Georgia Jet' plants in experiment 4. Thus, mineral nutrient supply may have played a role in acclimation of sweetpotatoes to CO_2 enrichment. Acclimation to elevated CO_2 has been linked to feedback inhibition due to excessive carbohydrate accumulation and to a decrease in rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) activity, but the mechanism of acclimation is poorly understood and likely differs among species (Bunce, 1992; Bowes, 1993; Tissue et al., 1993). In cotton, acclimation to elevated CO_2 was more pronounced

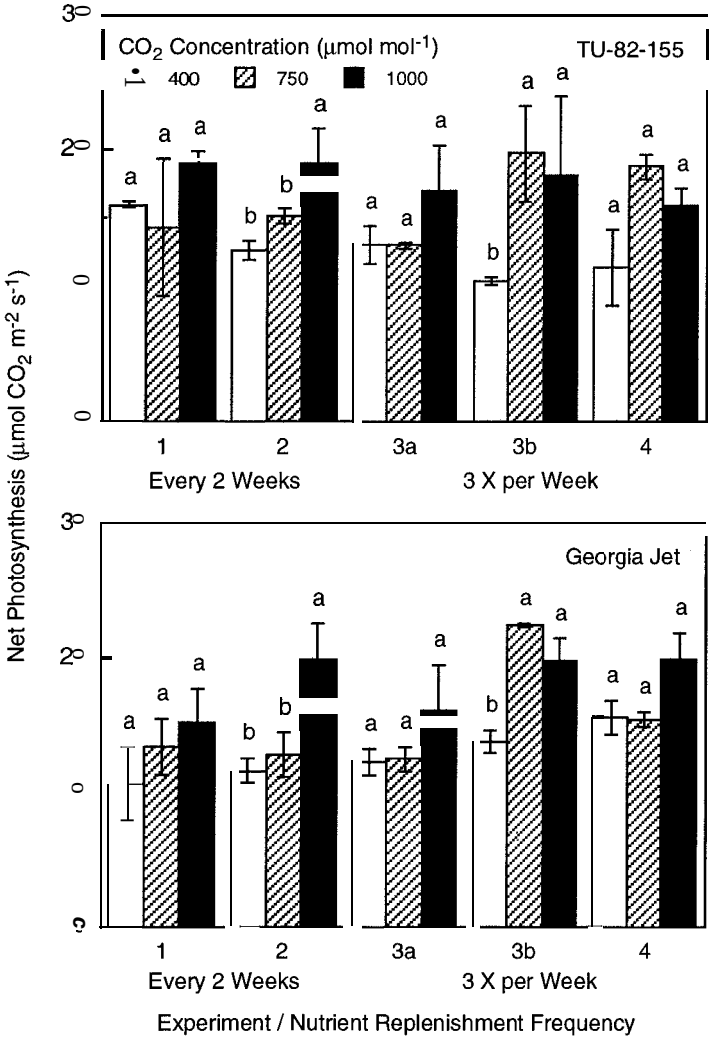


Fig. 1. Effect of CO_2 enrichment on net photosynthetic rates (measured at the growth CO_2 concentration) of two varieties sweetpotato grown hydroponically in two different nutrient regimes. Within each group of bars, different letters indicate significant differences between treatments.

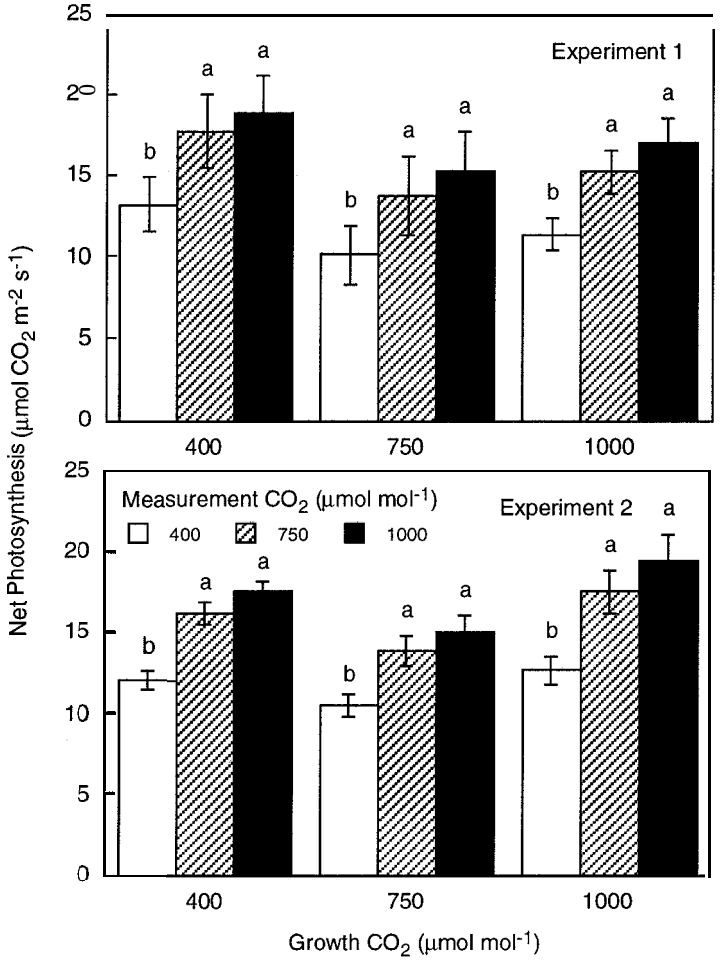


Fig. 2. Effects of measurement and growth CO_2 concentrations on net photosynthetic rates of sweetpotato grown hydroponically in two experiments with the nutrient solution replaced every two weeks. Within each group of bars, different letters indicate significant differences between measurement CO_2 concentrations. Photosynthetic rates did not differ between growth CO_2 concentrations.

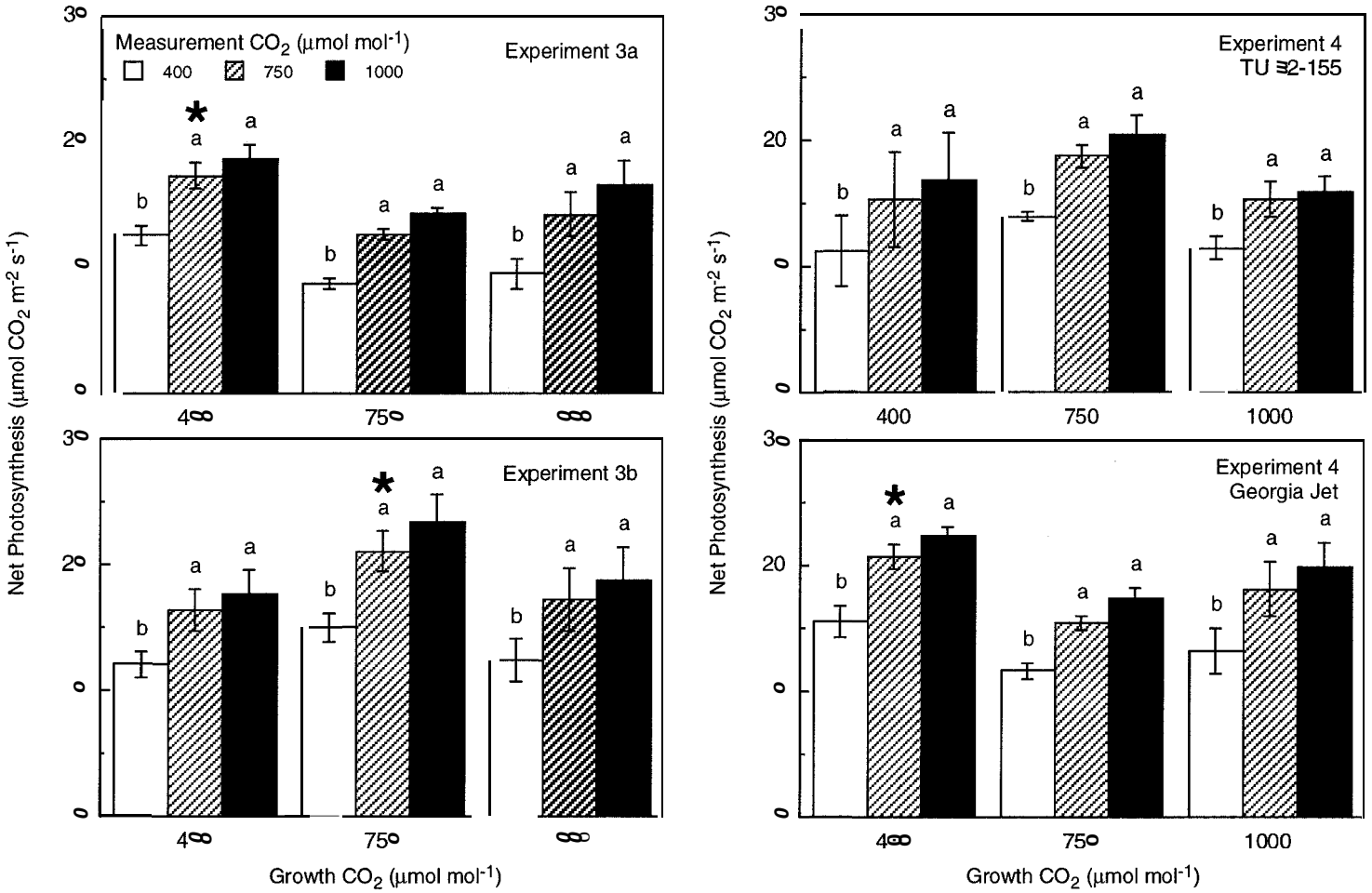


Fig. 3. Effects of measurement and growth CO_2 concentrations on net photosynthetic rates of sweetpotato grown hydroponically in two experiments with the nutrient solution replenished three times per week. Within each group of bars, different letters indicate significant differences between measurement CO_2 concentrations. Asterisks above a group of bars indicates that net photosynthetic rates were significantly higher at that growth CO_2 concentration.

in plants supplied with lower concentrations of nitrate, which is the reverse of the trend that we observed (Wong, 1979).

Photosynthetic response to CO_2 enrichment may also have been affected by nutrient limitations, as indicated by the rapid depletion of nitrogen, from the nutrient solution in experiments 1 and 2. The enhanced mineral nutrient supply used in experiments 3 and 4 led to an overall increase in photosynthetic rates in 'Georgia Jet' plants, but not in 'TU-823- 155' plants. This indicates that at least the 'Georgia Jet' plants may have been nutrient-limited when the solution was replaced be-weekly. However, 'TU-82- 155' plants may have engaged in luxury consumption of the additional nutrients supplied in experiments 3 and 4. In an earlier experiment at ambient CO_2, increasing the frequency of nutrient solution replacement led to increased nutrient concentrations in storage roots of hydroponically-grown sweetpotato (Grant et al., 1993). More experiments are needed to clarify the relationship between CO_2 enrichment and mineral nutrient supply in hydroponically-grown sweetpotato.

ACKNOWLEDGMENTS

We thank J. Seminara and D. Douglas for technical assistance. This work was supported by the National Aeronautics and Space Administration (Grant No. NAGW-2940 and Cooperative Agreement NCC 9-51) and by USDA/CSREES (Grant No. ALX-FS-2).

LITERATURE CITED

Allen, L. H., Jr. 1990. Plant responses to rising carbon dioxide and potential interactions with air pollutants. J. Environ. Qual.19:15-34.

Avener, M. 1989. Controlled ecological life support system. p. 145-153 *In* D.W. Ming, and D.L. Heninger (cd.) Lunar Base Agriculture: Soils for Plant Growth. ASA-CSSA-SSSA, Madison, WI.

Biswas, P. K., D.R. Hileman, P.P. Ghosh, N.C. Bhattacharya, and J.N. McCrimmon. 1996. Growth and yield responses of field-grown sweetpotato to elevated carbon dioxide. Crop Sci. 36:1234-1239.

Bonsi, C., W. Hill, D. Mortley, P. Loretan, C. Morris, and E. Carlisle. 1992. Growing sweetpotatoes for space missions using NFT. p. 110-119 *In* W.A. Hill, C.K. Bonsi, and P.A. Loretan (cd.) Sweetpotato Technology for the 21st Century. Tuskegee University, Tuskegee, AL.

Bowes, G. 1993. Facing the inevitable: plants and increasing atmospheric CO_2. Ann. Rev. Plant Physiol. Plant Mol. Biol. 44:309-332.

Bunce, J.A. 1992. Light, temperature and nutrients as factors in photosynthetic adjustment to an elevated concentration of carbon dioxide. Physiol. Plant. 86:173-179.

Cure, J. D., and B. Acock. 1986. Crop responses to carbon dioxide doubling: a literature survey. Agric. For. Meteorol. 38:127-145.

Evans, J.R. 1989. Photosynthesis and nitrogen relationships in leaves of C_3 plants. Oecologia 78:9-19.

Farquhar, G. D., and T.D. Sharkey. 1982. Stomatal conductance and photosynthesis. Ann. Rev. Plant Physiol. 33:317-345.

Field, C., and H.A. Mooney. 1986. The photosynthesis-nitrogen relationship in wild plants. p. 25-55. *In* T.J. Givnish (cd.) On the Economy of Plant Form and Function. Cambridge Univ. Press, Cambridge.

Grant, P., J.Y. Lu, D. Mortley, P.A. Loretan, C.K. Bonsi, and W.A. Hill. 1993. Nutritive composition of hydroponically-grown sweetpotato storage roots as affected by frequency of nutrient solution change. Hortscience 28:802-804.

Hill, W. A., P.A. Loretan, C.K. Bonsi, C.E. Morris, L. J. Y., and C. Ogbuehi. 1989. Utilization of sweet potatoes in controlled ecological life support systems (CELSS). Adv. Space Res. 9(8):29-41.

Larcher, W. 1980. Physiological Plant Ecology, 2nd ed. Springer-Verlag, Berlin.

Loretan, P. A., C.K. Bonsi, W.A. Hill, C.R. Ogbuehi, D.G. Mortley, J.Y. Lu, C.E. Morris, and R.D. Pace. 1989. Sweet potato growth parameters, yield components and nutritive value for CELSS applications. J. Aerospace 98:1090-1094.

Mortley, D. G., P.A. Loretan, C.K. Bonsi, W.A. Hill, and C.E. Morris. 1991. Plant spacing influences the yield and growth rate of sweetpotatoes grown hydroponically. Hortscience 26:1274-1275.

Strachan, R.C. 1993. The effects of CO_2 enrichment and differential nitrogen nutrition on growth and photosynthesis of sweetpotato. M.S. Thesis, Tuskegee University.

Tissue, D. T., R.B. Thomas, and B .R. Strain. 1993. Long-term effects of elevated CO_2 and nutrients on photosynthesis and rubisco in loblolly pine seedlings. Plant Cell Environ. 16:859-865.

Wheeler, R. M., C.L. Mackowiak, L.M. Siegriest, and J.C. Sager. 1993. Supraoptimal carbon dioxide effects on growth of soybean [*Glycine max* (L.) Merr.]. J. Plant Physiol.142:173-178.

Wong, S.C. 1979. Elevated atmospheric partial pressure of CO_2 and plant growth. I. Interaction of nitrogen nutrition and photosynthetic capacity in C_3 and C_4 plants. Oecologia 44:68-74.

Woodward, F. I., G.B. Thompson, and I.F. McKee. 1991. The effects of elevated concentrations of carbon dioxide on individual plants, populations, communities and ecosystems. Ann. Bet. 67 (suppl. 1):23-38.

Wullschleger, S.D. 1993. Biochemical limitations to carbon assimilation to C_3 plants -- a retrospective analysis of the A/C_i curves from 109 species. J. Exp. Bet. 44:907-920.

AUTOMATED pH CONTROL OF NUTRIENT SOLUTION IN A HYDROPONIC PLANT GROWTH SYSTEM

B. Smith, N. Dogan, H. Aglan, D. Mortley, and P. Loretan

Center for Food and Environmental Systems for Human Exploration of Space
Tuskegee University, Tuskegee, AL 36088

INTRODUCTION

Over, the years, NASA has played an important role in providing to and the development of automated nutrient delivery and monitoring, systems for growing crops hydroponically for long term space missions. One example are the systems used in the Biomass Production Chamber (BPC) at Kennedy Space Center (KSC). The current KSC monitoring system is based on an engineering workstation using standard analog/digital input/output hardware and custom written software. The monitoring system uses completely separate sensors to provide a check of control sensor accuracy and has the ability to graphically display and store data form past experiment so that they are available for data analysis [Fortson, 1992]

In many cases, growing systems have not been fitted with the kind of automated control systems as used at KSC. The Center for Food and Environmental Systems for Human Exploration of Space (CFESH) located on the campus of Tuskegee University, has effectively grown sweetpotatoes and peanuts hydroponically for the past five years. However they have adjusted the pH, electrical conductivity and volume of the hydroponic nutrient solution only manually at times when the solution was to be replenished or changed out according to its protocol (e.g. one-week, two-week, or two-day cycle). But the pH of the nutrient solution flowing through the channel is neither known nor controlled between the update, change out, or replenishment period. Thus, the pH of the nutrient solution is not held at an optimum level over the span of the plant's growth cycle. To solve this dilemma, an automated system for the control and data logging of pH data relative to sweetpotato production using the nutrient film technique (NFT) has been developed,

This paper discusses a microprocessor-based system, which was designed to monitor, control, and record the pH of a nutrient solution used for growing sweetpotatoes using (NFT).

DATA ACQUISITION AND COMPUTER HARDWARE

The Data acquisition hardware used in this project consists of an AT-MIO-16E-2 and a AT-MIO-16XE-10 analog/digital card. The AT-MIO-16E-2 (DAQ) card consists of 16

single-ended or 8 differential analog inputs with a resolution of 12 bits. It also provides two analog outputs capable of producing a voltage from O to 10 V, or plus/minus 10 V, as well as 8 digital input/output channels. The AT-MIO-16XE- 10 analog/digital card has the same features as the aforementioned card with the exception of having a resolution of 16 bits. The signal received by the DAQ cards is obtained from a Hanna pH controller. The controller acts as a sensor which provides a voltage corresponding to the pH of the nutrient solution to the computer for decision making. Actually the controller provides a O to 20 mA output signal which is then conditioned to a voltage ranging from O to 5 V.

The computer chosen for this project is an IBM compatible Personal Computer (PC) chosen because of its low cost, flexibility, and ability to perform the required functions. An NT operating platform, developed by the Microsoft Corporation is being used with this PC to provide a fully automated monitoring, control, and data logging system for pH of a hydroponic nutrient solution. The NT platform was selected for its ability to multi-task and provide stability to the operating system.

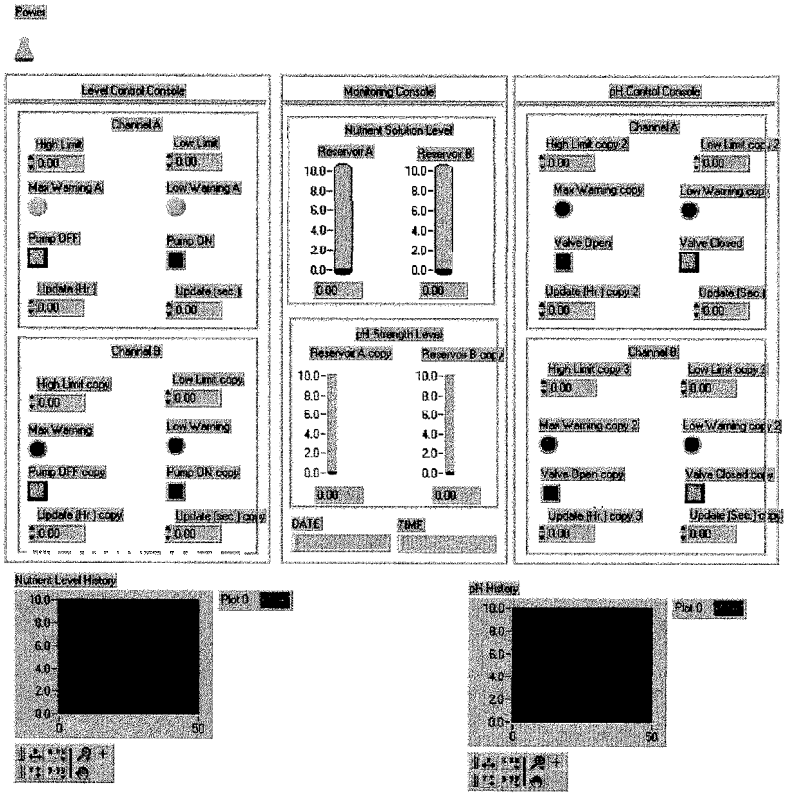


Figure 1 (Front Panel)

The data acquisition software utilized in the project is from National Instruments (Austin, Texas) entitled LabView Full Development System for Windows. In LabView, graphical panels, block diagrams, and icons are used to represent the algorithms and functions that the system utilizes to execute its programmed commands. The real time operating screen

or front panel *(Figure 1)* serves as the user interface section of the program. On this screen are warning displays, gauges, charts, etc. that can be used by the researcher to fine-tune the experiment being conducted. This front panel serves as a virtual representation of an actual control panel and operates in the same manner. The block diagram panel *(Figure 2)* contains the actual system programming codes which dictate how the system operates. On this panel the I/O operations, computational components, acquisition and analysis routine icons are connected together. This method of programming does not utilize the conventional text-based programming codes, but the icons still represent all the functions that would be available using text-based codes.

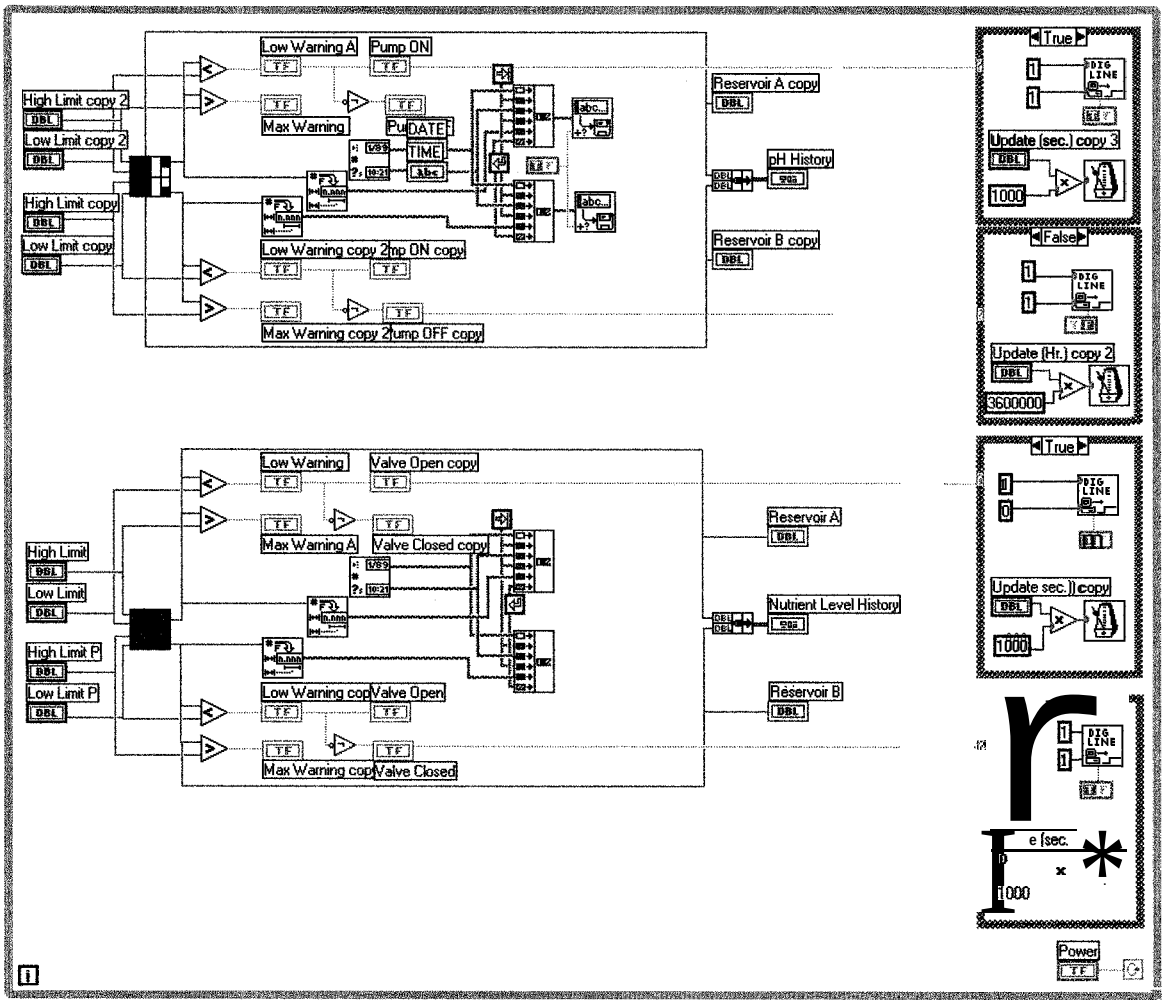


Figure 2 (Diagram Panel)

SYSTEM OPERATION

An explanation of the fictional diagram *(Figure 3)* showing the general operation of the system follows. The (DAQ) card is mounted within the PC, which receives a conditioned

signal from the controller proportional to the pH of the nutrient solution in the growth channel being monitored. In the diagram, the PC is the primary controller for the automated system in that it receives data from the controller and makes a decision concerning it based on the set point. To accomplish this the voltage input to the PC must first be converted to a pH value on a scale from O to 14. The voltage to pH conversion is achieved by using the calibration which has been previously input to the computer 'by charting the voltage against a pH value *(Figure 4)*. Once this step is completed, the pH value is subjected to a series of logic tests. The results of these tests then determine what action the control system should take to maintain the required system pH value. In determining whether an acid, base, or no solution at all needs to be added to the nutrient solution, the pH value is compared to preset maximum and minimum set points. If the pH is outside the maximum or minimum set point, one of two relays is energized which causes a normally-closed solenoid valve to open. If the pH is higher than the maximum set point, the solenoid valve controlling the acid (HCl) solution opens and acid is allowed to flow to reduce the pH. If the pH is lower than the minimum set point, the solenoid valve controlling the base (NaOH) opens and is allowed to flow to increase the pH. However, no matter which decision is made by the control system, all pH data will be chronologically logged into an Excel file for evaluation by the researcher.

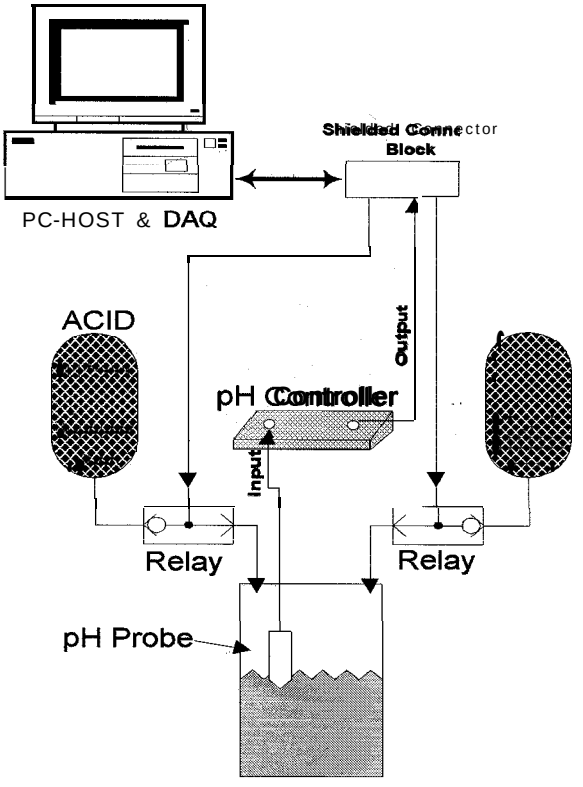


Figure 3 (Functional Diagram)

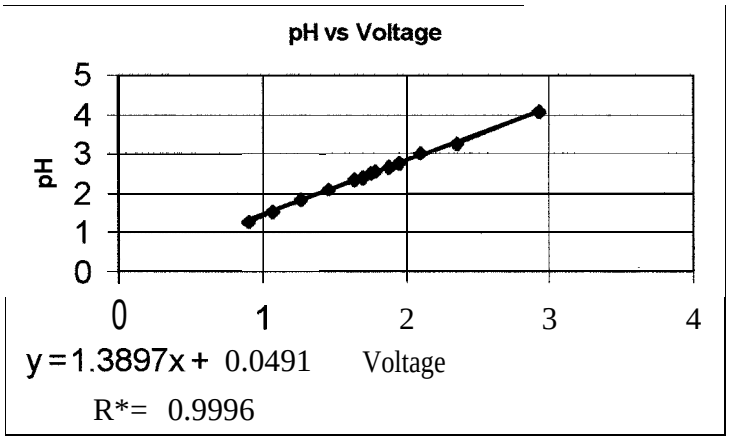


Figure 4 (pH vs Voltage)

FUTURE SYSTEM UPGRADES

Further development of the system will investigate the possibility of controlling the pH for several channels instead of one. Also, whether the controller can be programmed to change the protocol requirements at prescribed points in the plants' growth cycle will be examined. For accomplishing the former, consideration is being given to using a multiplexed interface, which will allow one controller to read and transmit data to the computer from multiple pH sensors. This method, however, has a down side because the system can only be calibrated by one of the probes attached to the multiplexer and each NFT channel will need its own pH sensor. The second method being considered incorporated the development of a mechanical system that will have the capability of transporting the sensor from one channel to the next. Such a system has the advantage of needing only one pH sensor, which avoids the aforementioned calibration problem and eliminates the need of having a pH sensor for each individual NFT channel.

ACKNOWLEDGEMENT

This research is supported by the National Aeronautics and Space Administration (Cooperative Agreement NCC 9-51) and USDA/CSREES (Grant ALX-FS-2).

REFERENCE

Fortson, R. E., J.C. Sager, J.O. Bledsoe, R.M. Wheeler, and W.M. Knott. 1992. Current performance of the NASA biomass production chamber. Paper No. 93-3510. ASAE, Chicago, IL.

Simulated Microgravity Induced Cytoskeletal Rearrangements Are Modulated by Protooncogenes.

C.D. Melhado, G.L. Sanford, F. Bosah and S.Harris-Hooker,
Space Medicine & Life Sciences Research Center,
Morehouse School of Medicine,
Atlanta, GA 30310-1495

INTRODUCTION: Microgravity is the environment living systems encounter during space flight and *gravitational unloading* is the effect of this environment on living systems. The cell, being a multiphasic chemical system, is a useful starting point to study the potential impact of gravity unloading on physiological function. In the absence of gravity, sedimentation of organelles including chromosomes, mitochondria, nuclei, the Golgi apparatus, vacuoles, and the endoplasmic reticulum may be affected. Most of these organelles, however, are somewhat held in place by cytoskeleton. Hansen and Igber suggest that intermediate filaments act to stabilize the nuleus against rotational movement, and integrate cell and nuclear structure (1). The tensegrity theory supports the idea that mechanical or physical forces alters the cytoskeletal structures of a cell resulting in the changes in cell: matrix interactions and receptor-signaling coupling (1). This type of stress to the cytoskeleton may be largely responsible regulating cell shape, growth, movement and metabolism.

Mouse MC3T3 El cells under microgravity exhibited significant cytoskeletal changes and alterations in cell growth (2). The alterations in cytoskeleton architecture may be due to changes in the expression of actin related proteins or integrins. Philopott and coworkers (3) reported on changes in the distribution of microtubule and cytoskeleton elements in the cells of heart tissue from space flight rats and those centrifuged at 1.7g. Other researchers have showed that microgravity reduced EGF-induced c-fos and c-jun expression compared to 1 g controls(4). Since c-fos and c-jun are known regulators of cell growth (5), it is likely that altered signal transduction involving protooncogenes may play a crucial role in the reduced growth and alterations in cytoskeletal arrangements found during space flight.

It is clear that a microgravity environment induces a number of changes in cell shape, cell surface molecules, gene expression and cytoskeletal reorganization. However the underlying mechanism for these cellular changes have not been clearly defined. We examined alterations in endothelial migration, and cytoskeleton architecture (microfilamentous f-actin and vimentin-rich- intermediate filaments) following wounding under simulated microgravity. We also examined the possibility that altered signal transduction pathways, involving protooncogenes, may play a crucial role in microgravity-induced retardation of cell migration and alterations in cytoskeletal organization. We hypothesize that, based on the tensegrity theory, cytoskeletal organization respond to gravitational unloading and through this response, cell behavior, function and gene expression are modified.

EXPERIMENTAL METHODS

Bovine aorta endothelial cells (BAEC) were obtained from NIGMS/Coriel Cell Repository, adapted and maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum and 1X antibiotic. The antisense oligonucleotides (AS-oligos) used were commercially prepared based on nucleotide published sequences for c-myc, c-fos and c-jun (Bio-synthesis Inc.) The AS-oligos were complimentary to the first twenty bases of each message; the c-myc sense strand was used as a control oligo.

OLIGONUCLEOTIDES

C-myc antisense:	5'-ACGATGCCCCTCAAGCTGAGC-3'
C-fos antisense:	5'-GGCGTTGAAACCCGAGAA-3'
C-jun antisense:	5'-CGTTTCCATCTTTGCAGT-3'
C-myc sense:	5'-TGCTACGGGGAGTTCGAC-3'

BAEC wound closure assay: The ability of BAEC to recover from a denudation injury (wound) when subjected to gravitational unloading was assessed as follows: Confluent cultures were grown in chamber slide plates, adapted to simulated microgravity by horizontal clinostat rotation at 30 rpm, for 12 hr; controls cultures were subjected to vertical clinostat rotation or were stationary cultures. There were differences in the response of vertically rotated or stationary control cells over the period studied. These cultures were subjected to denudation injury, rinsed 3X with HBSS, then replenished with DMEM. Cultures were microphotographed, under phase contrast, after the initial injury. Cultures were maintained under simulated microgravity or control conditions, then photographed after 6 hr.

Evaluation of AS-Oligos: Cultures were incubated for 2 hr with AS-oligos or control oligos (3 μg/ml) prior to preadapting to simulated microgravity for 12 hr. Following this period cultures were treated as outline under the *BAEC wound closure assay* above. However, following injury, the medium was replenished with DMEM containing 3 μg/ml oligos.

Fluorescence cytochemical labeling Fluorescence double labeling was used to assess changes f-actin and vimentin cytoskeletons. Cultures treated as given above were fixed at O and 6 hr with 3.7% formaldehyde in phosphate buffered saline (PBS), pH 7.4 for 30 min. Cultures were then dehydrated in absolute acetone for 4 min at -20 'C and permeabilized by a 2-min exposure to O. 1% Triton X-100 (1 1). Cells were labeled for vimentin by indirect immunostaining. Cells were incubated with 10% goat serum in PBS for 20 min to suppress non-specific binding of IgG, then with monoclinal anti-vimentin (1 O μg/ml) for 60 min. Cells were then incubated with biotinylated goat anti-mouse IgG (10 μg/ml) for 45 rein, rinsed 3X and incubated with FITC-conjugated -extravidin (1 :200 in PBS) for 15 min. These vimentin labeled cultures were then labeled for f-actin. Texas-Red conjugated phalloidin (50 μg/ml) was incubated with the cells for 40 min and washed with several changes of PBS, over 30 min. Double fluorescence-labeling of cytoskeletons were observed by confocal and/or epifluorescence microscopy.

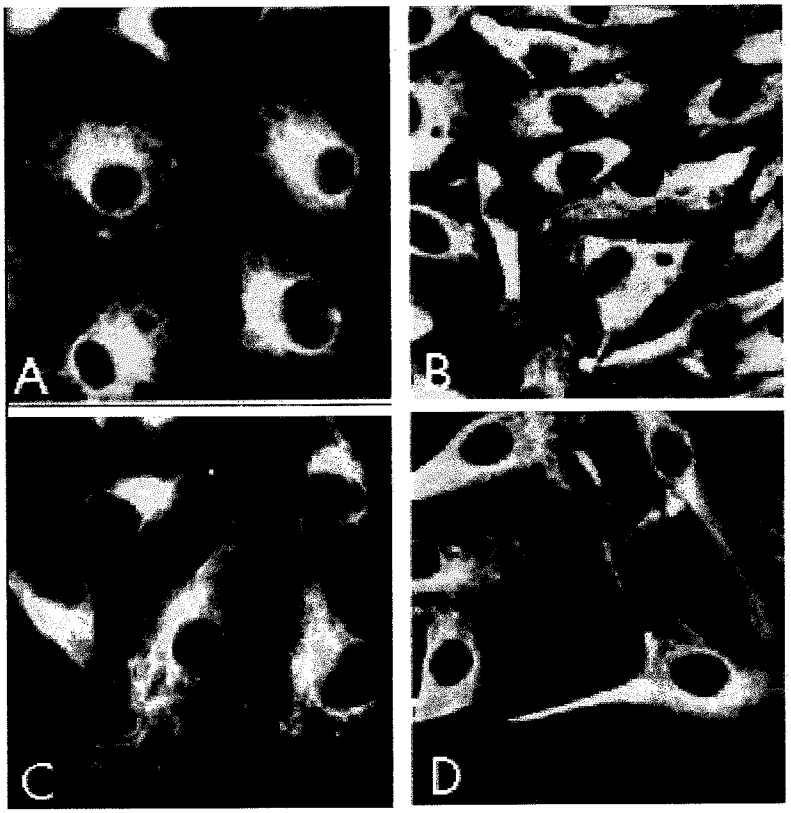


Figure 1. Double fluorescence labeled BAEC for **f-actin** (Texas-Red **phalloidin)** and **vimentin (FITC-antivimentin)**. Simulated **microgravity** treated cells in a non-injured (A) or injured area (B). **Stationary** controls in a non-injured (A) or injured area **(B)**.

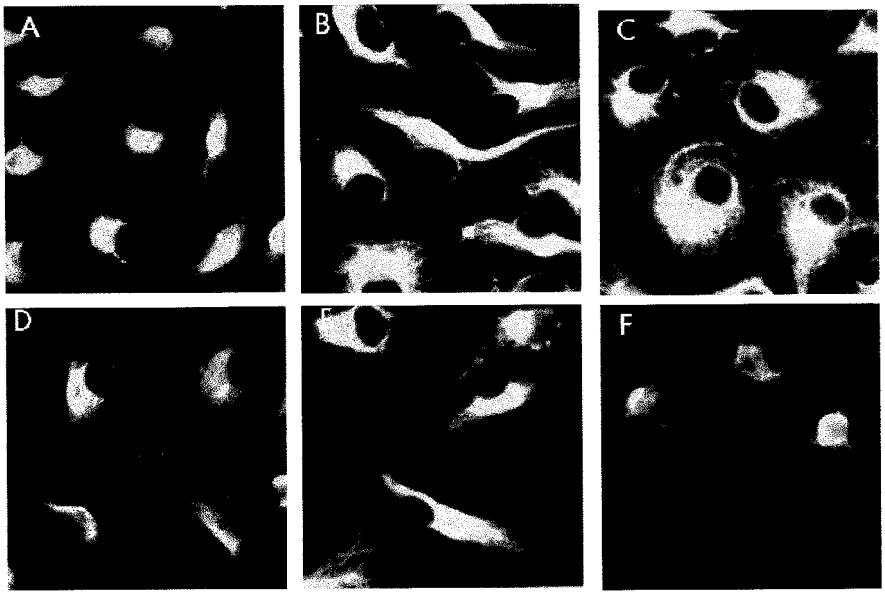


Figure 2. Double fluorescence labeled BAEC for **f-actin** (Texas-Red **phalloidin)** and **vimentin (FITC-antivimentin)**. Simulated **microgravity** treated cells in the presence of **c-myc AS-oligos** (A & B), **c-fos AS-oligos** (C & D) or **c-jun AS-oligo** (E & F) in a non-injured or injured area, respectively.

RESULTS AND DISCUSSION

Cytoskeletal Element Enabling Gravitational Force Unload Transmission Through the Cell: Double fluorescence-labelling of cytoskeletal components revealed some of the geometric parameters of the cell, the cytoskeketal arrangement and the physical relationship between the intermediate filament (vimentin) and microfilaments (actin). Simulated microgravity induced reorganization of actin microfilaments from a " dense banding pattern"(DBP) centrally and peripherally to a series of long, almost parallel stress fibers (Fig. 1). The f-actin filaments were randomly distributed and the fluorescence intensity enhanced. In addition, gravitational unloading induced cellular enlarging and flattening. There is strong evidence that the force transduction mechanisms in the endothelial cells are a combination of force transmission via cytoskeletal elements and transduction of physical forces to biochemical signals. F-actin appears to the principal transmission structures (1 1).

We also found vimentin immunolabeling in two prevalent patterns of distribution: (a) cytoplasmic filaments characteristic of dense, uniform VIF network (associated with stationary cultures); (b) a distinct, juxtanuclear focus with limited peripheral extensions (associated with gravitational unloading). The juxtanuclear focus of vimentin was seen in the centrosomal region. Additionally cells fail to form extended VIF networks, but were organized as perinuclear aggregates. Unlike f-actin, there is a lack of soluble pools of vimentin available for polymerization into filaments (7). Reorganization of endothelial vimentin during gravitational unloading may therefore occur as passive association of vimentin with f-actin, that is undergoing remodeling through cycles of polymerization/depolymerization. These findings are consistent with our observation that gravitational unloading induced vimentin juxtanuclear relocalization coupled with increased f-actin polymerization into parallel stress fibers. These studies suggest that there exist physical connections between these important structural protein families that determine the shape of anchorage-dependent cells, and it is probable that a coordinated rearrangement occurs in endothelial cell subjected to gravitational unloading. These observations are supportive of the tensegrity model.

Role of Protooncogenes in the Response of Cells to Gravitational Unloading: The molecular mechanisms responsible for transducing gravitational unloading into gene transcriptional changes have not been elucidated. It has been established that second messengers located at the plasma membrane and in the cytoplasm are coupled to "third messenger" transcription factors and serve as elements for information transfer to the nucleus where they induce the expression of immediate early response genes (IERG). Endothelial IERGs that are known to be responsive to changes of shear stress include c-myc, c-fos, and c-jun (1 O).

We used an antisense oligonucleotide (AS-Oligo) approach in conjunction with denudation injury to investigate whether protooncogene expression is involved in microgravity-induced cytoskeletal changes and ultimately their coupled involvement in wound healing. Cultures adapted to microgravity and incubated with AS-oligos then injured were allowed 6 hr for wound repair under both microgravty and normal conditions. As described earlier, the gravitational unloading induced vimentin-juxtanuclear relocalization with few peripheral extensions coupled with increased f-actin polymerization, for cells at injury and non-injury sites. Also, gravitational unloading retarded cell migration into the injury site (Fig. 3). These effects induced by gravitational unloading were synergistically enhanced by galactin AS-Oligo treatment. It was previously found that galactin AS-oligo treatment inhibits migration under stationary conditions (9).

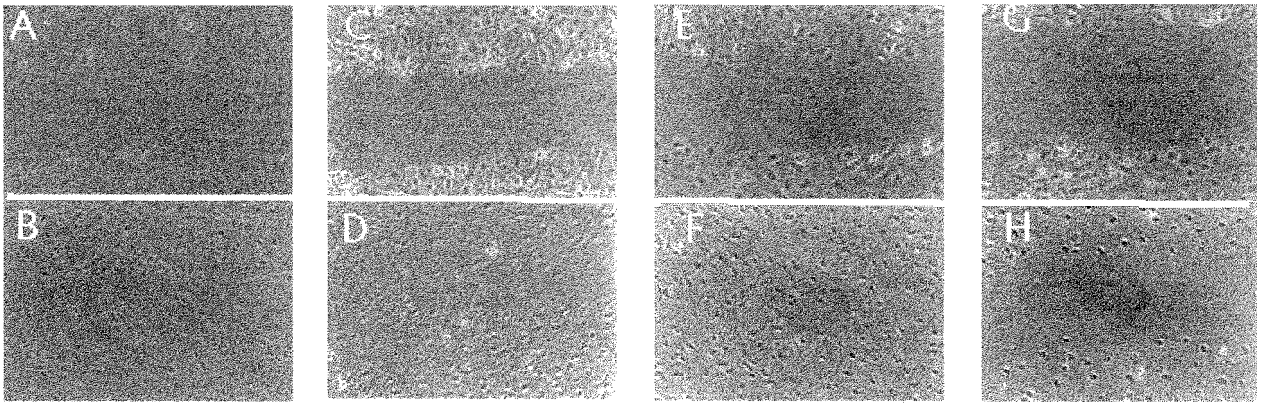


Figure 3. Phase contrast photomicrograph of BAEC injury site of microgravity treated cells incubated with 3µg/ml As-oligo to (C& D) c-myc, (E&F) c-fos, (G&H) cjun. (A &B) are untreated ontrols. The upper panel (A,C,E,G) were taken at immediately following injury. The lower pannel (B,D,F,H) photomicrographs were taken 6hr follwing injury.

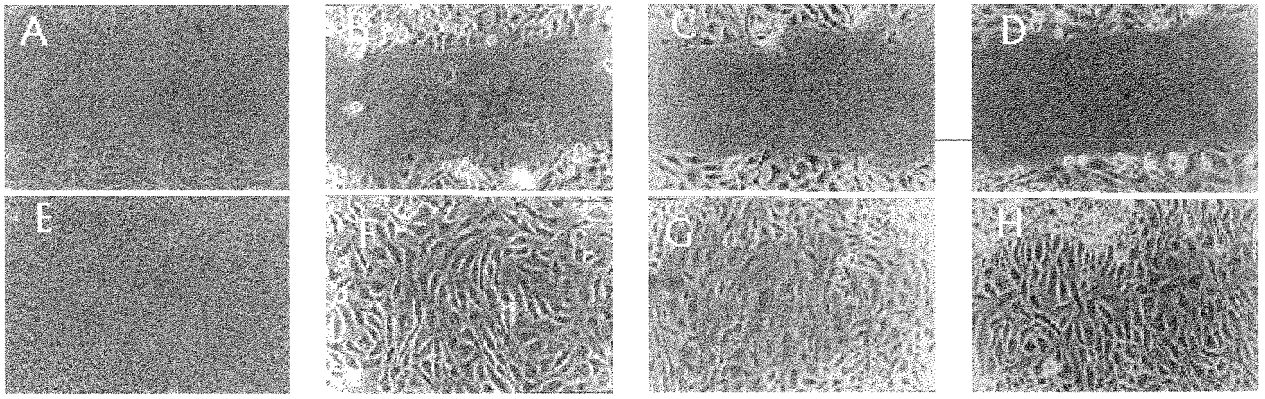


Figure 4. Phase contrast photomicrograph of BAEC injury site of stationary cultures incubated with 3µg/ml As-oligo to (B&F) c-myc, (C&G) c-fos, (D&H) c-jun. (A &E) are untreated ontrols. The upper panel (A,B,C,D) were taken at immediately following injury. The lower pannel (E,F,G,H) photomicrographs were taken 6hr follwing injury.

In contrast, gravitational unloading was ineffective in inducing the stress fiber assembly and migratory inhibition when cultures were treated AS-oligos to c-myc or c-fos (Figures 1 & 2). AS-Oligo treated (c-myc, c-fos) cell displayed a central and peripheral DBP of actin microfilaments and a vimentin juxtanuclear focus for migrating and non-migrating cells under microgravity, similar to stationary cultures (figures 1 &.4). There was a total loss of vimentin and peripheral extensions. Under microgravity, c-jun AS-Oligo treated cultures displayed vimentin filament and actin stress fiber disassembly analogous to c-myc and c-fos treated cultures, wound repair was also inhibited. These findings suggest that protooncogenes modulates the stress fiber assembly and delayed vimentin peripheral extensions associated with gravitational unloading , and that protooncogene suppression is necessary for stress fiber disassembly and delayed vimentin peripheral extensions redistribution necessary for cell locomotion and wound repair under microgravity . In addition, when cell locomotion is required, the centrosome redistributes to the front of the cell to establish cell polarity and acts as a modulator of the directionality of migration (8). Thus, the cytoskeleton is a dynamic intracellular system which regulates endothelial integrity and repair and is modulated by external stimuli.

ACKNOWLEDGEMENTS: These studies were supported by grants NASA NAG9-852 and NCCW0085

REFERENCES

1. Hansen L and Ingber DE, et al. (1994). Cellular Tensegrity: Exploring How Mechanical Changes in the Cytoskeletaon Regulates Cell Growth, Migratin, and Tissue Pattern during Morphogenesis. Int Rev Cyto 150:173-216.
2. Riijken et al. (1991). Aviat Space Environ Med 62:32-6.
3. Huges-Fuford M, Nelson K, Sblaug CG, et al. (1993). MC3T3-E1 Osteoblasts Grown in Microgravity on STS-56 Have Reduced Cell Growth, Glucose Utilization with Altered Actin Cytoskeleton and Increased Prostaglandin Synthesis. Am Soc Grav Space Biol Bull 7:31.
4. Philpott DE, Fine A, Mednieks M and Kato K (1985). Microgravity Changes in Heart Structure and c-AMP Metabolism. Physiologist 28(4):378, 1985
5. deGroot RP, Rijken PJ, Hertog J, et al. (1990). Microgravity Decreases c-fos Induction and Serum Response Element Activity. Cell Sci 97:33-38.
6. Bush SJ and Sassone-Corsi P (1990). Dimers, Leucine Zippers and DNA Binding Domains Trends Genet. (England) 6(2):36-40.
7. Watson PA (1991). Functions Following Form: Generation of Intracellular Signals by Cell Deformation. FASEB J 5:2013-2019.
8. Trevor KT, McGuire JG and Leonova EV (1995). Association of Vimentin Intermediate Filaments With the Centrosome. J Cell Sci 108(1):343-356.
9. Bosah FN and Sanford GL (1993). The Regulation of Galptin Gene Expression by Antisense Oligonucletides. Molec Biol Cell 4:1777.
10 Davies PF (1995). Flow-Mediated Endothelial Mechanotransduction. Physiol Rev 75:519-554.
11. Davies PF and Tripathi SC (1992). Mechanical Stress Mechanisms and the Cell. Circ Res 72:239-245.

A SYSTEM FOR MANAGING REPLENISHMENT OF A NUTRIENT SOLUTION USING AN ELECTRICAL CONDUCTIVITY CONTROLLER

D. Davis, N. Dogan, H. Aglan, D. Mortley, and P. Loretan

Center For food and Environmental Systems for Human Exploration of Space
Tuskegee University, Tuskegee, AL 36088

Introduction

Control of nutrient solution parameters is very important for the growth and development of plants grown hydroponically. Protocols involving different nutrient solution replenishment times (e.g. one-week, two-week, or two-day replenishment) provide manual periodic control of the nutrient solution's electrical conductivity (EC). Since plants take-up nutrients as they grow, manual control has a drawback in that EC is not held constant between replenishments. In an effort to correct this problem, the Center for Food and Environmental Systems for Human Exploration of Space at Tuskegee University has developed a system for managing and controlling levels of EC over a plant's entire growing cycle. A prototype system is being tested on sweetpotato production using the nutrient film technique (NFT), and it is being compared to a system in which sweetpotatoes are grown using NFT with manual control.

NASA has played an important role in the development of environmental control systems. They have become a forerunner in growing plants hydroponically with some control systems through the use of networked data acquisition and control using environmental growth chambers. Data acquisition systems which involve the use of real-time, calibration, set points, user panel, and graphical representation programming provide a good method of controlling nutrient solution parameters such as EC and pH [Bledsoe, 1993].

In NASA's Biomass Production Chamber (BPC) at Kennedy Space Center, control is provided by a programmable logic controller (PLC). This is an industrial controller which combines ladder computer logic which has the ability to handle various levels of electrical power. The controller controls temperature, light and other parameters that affect the plant's environment, In the BPC, the Nutrient Delivery System (NDS), a sub-system of the PLC, controls nutrient solution parameters such as EC, pH, and solution levels. When the nutrient EC measurement goes outside a preset range (120-130 mS/m) a set amount of a stock solution of nutrients is automatically added by a metering pump to bring the EC back into operating range [Fortson, 1992].

This paper describes a system developed at Tuskegee University for controlling the EC of a nutrient solution used for growing sweetpotatoes with an EC controller and a computer with LabView data acquisition and instrumentation software. It also describes the preliminary data obtained from the growth of sweetpotatoes using this prototype control system.

Materials and Methods

EC Control System Components

An EC control system, used in this application to control the EC of a nutrient solution for growing sweetpotatoes includes an EC controller manufactured by Signet Inc. The controller

consists of two channels, a 10-10,000 $\mu S\ cm^{-1}$ sensor, four output relays, and two analog outputs. Temperature measurements are collected with a Type J Mini-Size thermocouple meter manufactured by Newport Inc. Two 12 V, 1/16 in, normally-closed solenoid valves allow water or a concentrated stock solution to flow via a gravity feed, to a nutrient solution reservoir. EC and temperature measurements from the controller and the temperature meter are recorded on a PC using LabView data acquisition and instrumentation software. The controller and the temperature meter are connected to the computer through a fifty pin MIO connection block and a AT-MIO-16-E-2 data acquisition card (DAQ) *(Figure 1).*

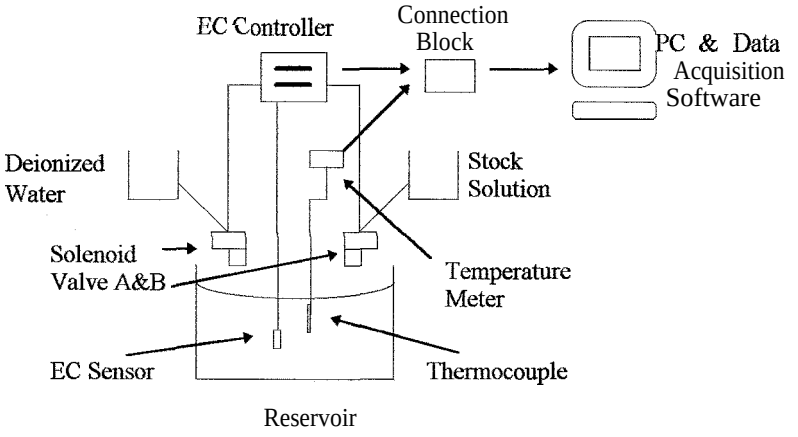


Figure 1: EC Control System

Standard NFT Growing Protocol

The above automated control system served as a prototype to control EC in a reservoir from which nutrient solution was fed to the roots of sweetpotato plants in a growth channel. Growth in this channel was compared to growth of sweetpotatoes in another growth channel in which EC was manually adjusted (a control). The standard protocol for growing sweetpotatoes in either channel is described in detail by Bonsi et al (1992), but the general characteristics of the system are described below.

The growth channel is rectangular (O. 15 x 0.15 x 1.2 m) and 15 cm sweetpotato cuttings (plants) are placed at a 25 cm spacing within the channel *(Figure 2).* A 45 L reservoir containing 30.4 L (8 gal) of modified half-Hoagland nutrient solution supplies the plants with the proper nutrients. For the automated EC system, the nutrient solution is replenished to the reservoir as the EC setting calls for. For the control channel, a modified half-Hoagland nutrient solution is changed every two weeks. When the volume of the nutrient solution in the reservoir drops below 8 L prior to the two week replacement interval, deionized water is added. Submersible pumps within the reservoir for each system pumps nutrient solution to the upper end of the growing channel. Each channel has 1% downward slope toward the reservoir. The nutrient solution is pumped to the growth channel at a 1 L min^{-1} flow rate which is set by a control valve and a bypass line back to the reservoir. The solution then spreads across the channel in a thin film and flows back to the reservoir creating a closed loop.

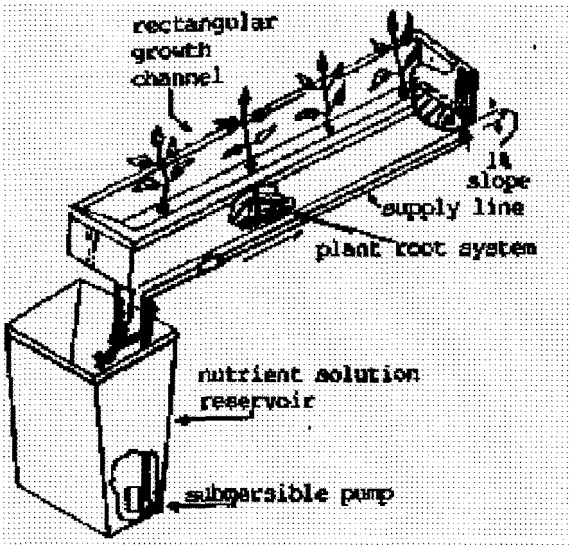


Figure 2: Standard NFT growing system

The standard harvest date for the plants is 120 days after planting (DAP) and yields are determined on a g/plant basis. All foliage is cut at the base, weighed flesh, dried for 72 h at 70^0C and reweighed dry. Storage roots are measured and weighed fresh.

Results and Discussion

Automated Control System Operation

The way the automated control system functions is relatively simple. The EC controller constantly monitors the nutrient solution. It automatically compensates for temperature for every EC reading. EC values are internally compensated 2% for every degree above or below 25^0C. A thermocouple monitors the temperature of the nutrient solution inside the reservoir but it is not responsible for temperature compensated EC readings. Maximum and minimum operating set points are programmed into the controller. The EC range of operation is within the range 1180-1220 $\mu S\,cm^{-1}$ with 1200 $\mu S\,cm^{-1}$ being the set point *(Figure 3).* If the EC drifts outside the set point, an internal relay is energized, closing a normally open switch, and a signal is sent to one of two normally closed solenoid valves. If the EC goes too high, the solenoid valve to a deionized water tank opens to allow water to flow and dilute the concentration of the nutrient solution in the reservoir. On the other hand, if the EC is too low, the valve to the concentrated stock solution tank opens to allow stock solution to flow into the nutrient solution reservoir. The solenoid valve remains open until the 1200 $\mu S\,cm^{-1}$ set point is reached at which time the relay de-energizes and closes the solenoid valve. The nutrient solution is constantly agitated due to a pump bypass line. However, some of the nutrient solution is pumped into the growing channel from the reservoir and circulates through the NFT channel back to the reservoir forming a closed loop. The gravity fed solutions, deionized water and concentrated stock solution enter the reservoir at a rate of approximately 1 mL per second. The slow feed rate allows for proper mixing in the nutrient solution reservoir and reduces the chance of over shooting the 1200 $\mu S\,cm^{-1}$ set point.

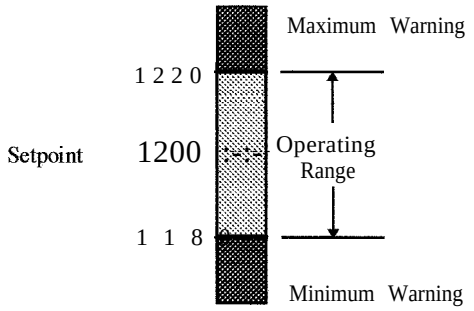


Figure 3: EC Controller Operating Range

Data Acquisition

Through the use of the LabView software, a virtual instrument (VI) panel is created for system monitoring. The panel displays the EC and temperature readings, charts, and set points (EC only). A graphical program is then created to suit the needs of the system. Every half-hour the computer takes a reading from the controller and the temperature meter. The readings from the controller and the temperature meter are voltages that are converted to EC and temperature values via linear equations. The same maximum and minimum limits on the controller are also set in the program. If an EC reading is outside the operating range, warning lights on the panel are activated warning the operator that there may be a problem. Since the EC controller is a stand-alone unit, the computer has no control over its operation. It is possible that the computer could take a reading while the controller is correcting the problem and, when the next reading is taken a half hour later, the nutrient solution may already be within the operating range. On the other hand, a warning light maybe an indication that any of several other problems have occurred, e.g., either the deionized water or concentrated stock solution tanks maybe empty, the set points on the controller and the computer do not match, or the controller may need to be re-calibrated. The EC message panel gives the operator a step-by-step method for correcting any problems that may occur. All EC and temperature readings taken are recorded and kept in an Excel file as a permanent record.

Plant Growth System Harvest Results

For the first trial run of the automated control system vs. manual control, the sweetpotatoes were harvested at 90 DAP, instead of the standard 120 DAP since the control plants developed a root disease. The control plants did not produce any storage roots, but the plants in the automated system were successful in producing four storage roots during the reduced growth period. The plants within the automated system produced nearly three times the amount of foliage as the manual control plants. The mean foliage dry weight for the automated and manual system was 26.5 and 9.0 g respectively. As expected, the automated control system was successful in holding the EC constant throughout the growth period of the plants. More experiments will be performed in order to further test and develop the automated system.

System Upgrades

Further development of the system will include studying the effect on plant growth of varying the EC over the plant growing period by programming all EC levels. The computer will control all activity. The EC controller will no longer be a stand-alone unit. AU decisions will be made at the computer. The controller will only be used as a sensor.

Devising methods of expansion and automation to control multiple channels will be examined. Multiplexing one channel of the controller to provide multiple inputs for sensors is one potential method *(Figure4)*. Secondly, the development of a mechanical system to move a probe from one reservoir to another is also being considered. Finally, ion specific electrodes will be introduced into the system to target key ions such as potassium and nitrate. Replenishment will not be solely based on EC readings alone, but specific ion readings together with EC readings will determine the proper replenishment.

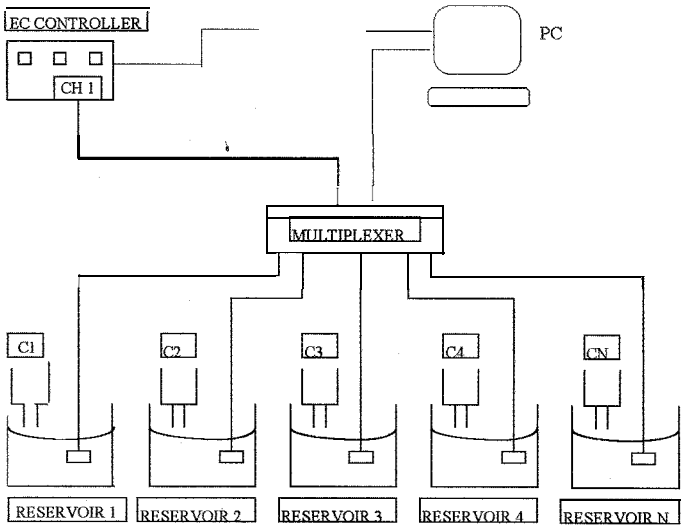


Figure 4: *Multiple Channel System*

Acknowledgment

This research is supported by the National Aeronautics and Space Administration through NASA Cooperative Agreement NCC 9-51.

References

Bledsoe, J. O., J.C. Sager, and R.E. Fortson. 1993. Networked data acquisition and control for environmental chambers. Paper No. 93-3510. ASAE, Chicago, IL.

Bonsi, C.K., W.A. Hill, D.G. Mortley, P.A. Loretan, C.E. Morris, and E.R. Carlisle. 1992. Growing sweetpotatoes for space missions using NFT. SWEETPOTATO TECHNOLOGY for the 21st CENTURY. 110-119 pp. IN Hill W. A., C.K Bonsi and P.A. Loretan (Ed). Tuskegee University, Tuskegee AL.

Fortson, R. E., J.C. Sager, J.O. Bledsoe, R.M. Wheeler, and W.M. Knott. 1992. Current performance of the NASA biomass production chamber. Paper No. 92-4001. ASAE, Charlotte, NC.

Observations of a Major Outburst of BL Lacertae in the Active Galactic Nuclei Monitoring Program in the Center for Automated Space Science

M.T. Carini, K.R. Hackney, S.D. Clements, R.C. Culler, R.L. Hackney, J.C. Noble, R.G. Gelderman, R.L. Scott, and C.H. McGruder III
Center for Automated Space Science (CASS)
Department of Physics and Astronomy
Western Kentucky University

Introduction

BL Lacertae (BL Lac) is the prototype for a class of Active Galactic Nuclei (AGN) known as BL Lac objects. These objects are characterized by a featureless optical continuum that is both highly polarized and extremely variable. Historically, this object is known to have an optical variability range of 5 magnitudes (Webb, *et al.*, 1987, Carini, *et al.* 1992) and exhibits episodic periods in which the source is observed to be bright (V = 12.0) and other periods when the source is observed to be quite faint (V = 16.5). In May, 1997, we found, in the course of our routine monitoring of these objects, that BL Lac was in an elevated state, with a V magnitude peaking near V = 12.5. The purpose of this paper is to report on our observations of the behavior of the optical continuum flux of BL Lac during this major outburst.

Observations

The observations presented here are a combination of our own observations, obtained at several different observatories, combined with the observations of other observers which have been made publicly available on the World Wide Web. This outburst marks the first time the WWW has been used to help identify and coordinate observations of an AGN during an outburst. Our observations were obtained at either the 1.05-meter Hall telescope at Lowell Observatory or the 0.75 meter telescope at the Rosemary Hill Observatory. The observations were made with CCD cameras equipped with a standard BVRI filter set. Repeated exposures of 90-120 seconds were obtained for the star field containing BL Lac and several comparison stars. These comparison stars (Smith, *et al.* 1985) were previously calibrated and located on the same frame as BL Lac, making them ideally suited as comparison stars. The observations were reduced following the method outlined by Howell and Jacoby (1986). Each exposure was processed through an aperture photometry routine that reduced the data as if it were produced by a multi-star photometer. Differential magnitudes were computed for BL Lac and several of the calibrated standard stars. Thus, simultaneous observations of BL Lac, several comparison stars, and the sky background allow one to remove variations that may be due to fluctuations in

atmospheric transparent y or extinction. The aperture photometry routine used to reduce these observations is the *apphot* routine in IRAF[1].

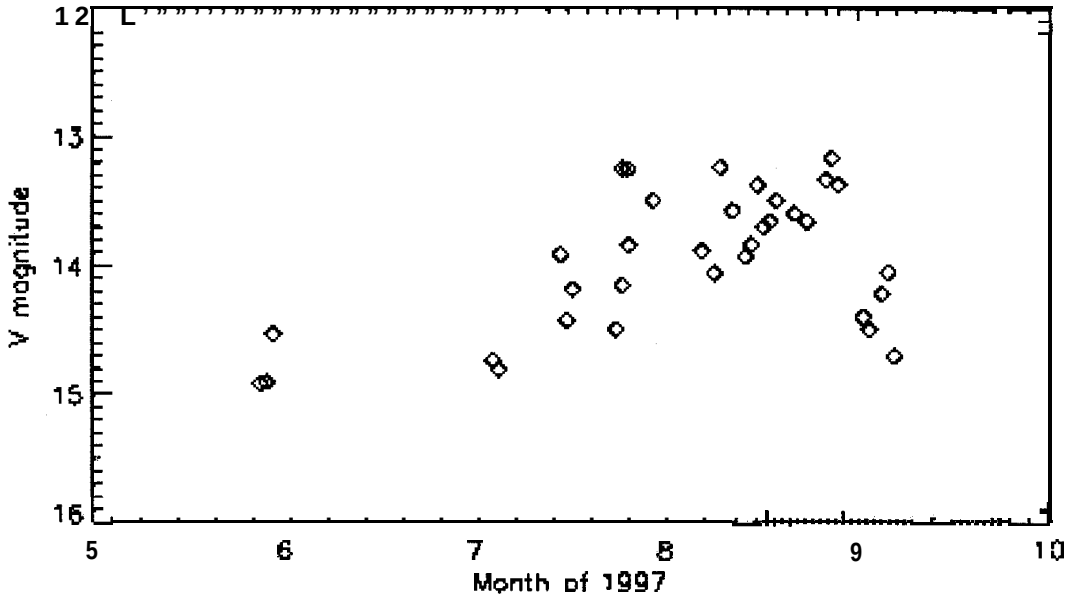


Figure 1-- Observations of BL Lac from May through September, 1997.

Results

Our observations, combined with the other publicly available data from the Web, are displayed in Figure 1. Typical errors are +/- 0.04 magnitude. The average V magnitude for BL Lac is V = 15.0. Over the last 5-7 years, the source has been observed at or fainter than this average magnitude.

However, we see that from late May through late August, the source brightened by 2 magnitudes, from V = 15.0 to V = 13.0. The apparent fading at the end of the data set is temporary; more recent observations indicate the source is still in a very bright state (Miller 1997, private communication). Closer inspection of the data during July and August shows variations of up to 1 magnitude in brightness on time scales of 1 day (Figure 2). More detailed observations presented by Bloom *et al.* (1997) (Figure 3) confirm this behavior. These authors detected a flare in the gamma ray flux occurring at the same time as the flare in the optical flux. Observations by Miller, *et al.* (1997) and Clements *et al. (1997)* show that rapid, large amplitude variations were occurring in BL Lac on time scales down to several hours with amplitudes of up to 0.5 magnitude.

[1]IRAF is distributed by the National Optical Astronomy Observatories, which is operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

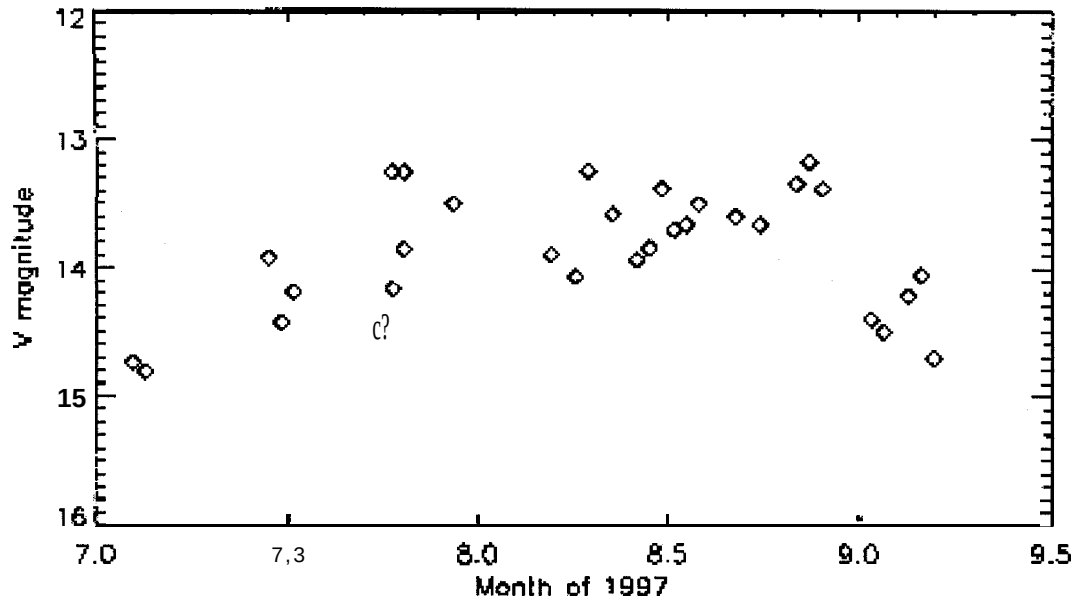


Figure 2-- Expanded plot of variations of BL Lac from July through September, 1997.

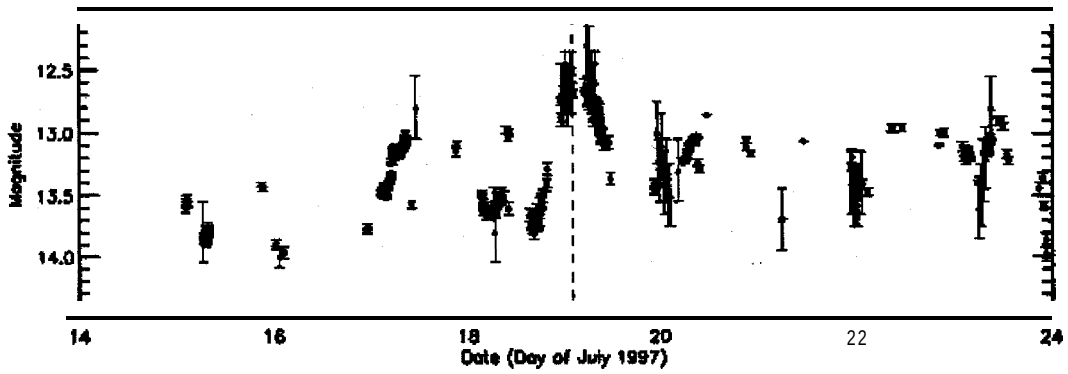


Figure 3-- Rapid variations in BL Lac with intra-day time scales, detected by Bloom *et al.* (1997), indicate a need for optical monitoring of this source with more intensive sampling.

In the standard model for AGN of the BL Lac class, the optical variations are attributed to disturbances propagating outward at relativistic velocities along a jet. If the jet is pointed nearly along the line of sight, then this orientation has the effect of enhancing the radiation and the associated variations. The overall increase in the source brightness observed since May is then the result of a shock traveling down the jet. The more rapid variations that are seen superimposed on this brightness increase arise as the result of disturbances or turbulence within the shock itself.

It is clear from a comparison of Figures 2 and 3 that coarse data sampling during an outburst of one of these objects is not sufficient to truly understand the variations the source is undergoing. The coverage in Figure 3 is far superior to that in Figure 2, yet even that does not adequately describe the source's behavior. Figures 1 and 2 are very typical representations of the standard way in which an outburst in one of these objects is monitored. Generally, a small number of observers will obtain a data point every night they can, working around weather and instrument schedules. This coverage provides a general description of the overall source behavior, but it misses shorter term variations. Figure 3 comes much closer to describing the actual variations the source is undergoing, but at a high price. A light curve such as that displayed in Figure 3 typically requires the dedication of multiple observers and instruments, all looking exclusively at one object for an extended period of time. However, a dedicated automated observatory located at a site with a high percentage of clear nights would produce nearly the same light curve. Even better, a world wide network of such telescopes could, in theory, provide a nearly complete (no data gaps) light curve of the outburst of one of these objects. Our goal is to achieve frequency of sampling beyond that of Figure 3, using the CASS 24-inch imaging telescope, which is nearing completion in Washington Camp, Arizona.

Conclusions

BL Lacertae, the proto-typical BL Lac object, has undergone a major optical outburst, after an extended quiescent period. The source brightened by over 2 magnitudes between late May and early July, 1997. It has remained in this bright state for several months. Extensive optical monitoring by other groups has shown a complicated variability pattern, requiring detailed observations to understand. Such observations could be better achieved using dedicated, automated observatories. We plan to begin a program, of monitoring these objects with such a facility in the next few months. We gratefully acknowledge support of NASA cooperative agreement NCC5-228.

References

Bloom, S. D., *et al.*, 1997, *Astrophys. J. Letters, 490, 145.*

Carini, M. T., Miller, H.R., Noble, J. C., Goodrich, B.D., 1992, *Astron. J*, 104, 15.

Clements, S.D., Carini, M. T., Noble, J. C., 1997, Paper to be presented to the American Astronomical Society.

Howell, S.B. and Jacoby, G. J., 1986, *Pub. Astron. Sot. Pac., 98, 802.*

Miller, H. R., 1997, private communication.

Miller, H.R., *et al.* 1997, Proceedings of the Perugia Conference on Variability in BL Lac Objects, (in press).

Smith, P. S., Balonek, T. J., Heckert, P.A., Elston, R., and Schmodt, G.D., 1985, *Astron. J., 90,* 1185.

Reverse Energy Cascade in Turbulent Weakly Ionized Plasmas

Kyron Williams, R. Appartaim, K. Belay, and J. A. Johnson III
Laboratory for Modern Fluid Physics
Center for Nonlinear and Nonequilibrium Aeroscience
Florida A&M University, Tallahassee, FL

1. Introduction. For systems far from equilibrium, the neglect of a role for viscous effects in turbulence may be generally inappropriate when the relaxation time for the molecular process approaches the local flow time (see Orou et al. (1996)). Furthermore, for stationary collisional plasmas, the conventional Reynolds number is irrelevant under circumstances where the standard features of turbulence in ordinary gases are observed in the plasma (see Johnson et al. (1987)). The current theoretical understanding of these turbulent phenomenon is particularly inadequate for turbulence associated with ionizing shock waves; generally speaking, thermodynamic, acoustic and pressure fluctuations are all seen as amplified across the shock wave followed by a dramatic decay (relaminarization) usually attributed to a lack of importance of viscosity in the turbulent regions. This decay would be accelerated when the flow speed is also reduced due to the importance usually given to the conventional Reynolds number (which is directly proportional to velocity) as a quality of turbulence index. However, evidence supporting this consensus is lacking.

By contrast, recent evidence of vanishing triple correlations form De Silva et al. (1996) provides strong support for early theoretical speculation of inherently molecular effects in macroscopic turbulence in Tsuge (1974). This specifically suggests that the role of compressive effects ordinarily associated with the shock wave could be significantly muted by the existence of a strongly turbulent local environment. There is also more recent theoretical speculation (in Frisch et al. (1984)) of an inherently and previously unsuspected non-dissipative nature to turbulence, with energy conservation being nurtured by reverse energy cascades in the turbulent fluctuation spectra. Furthermore, the role which might be played by fluctuations on quantum mechanical phenomena and variations in molecular parameters is completely unknown, - especially of the sort which might be found in optical spectra. A determination of the role which changing turbulent parameters might play on optical signatures behind a reflected shock wave should therefore provide a useful tool for the evolution of theoretical models for turbulence.

2. Experimental Set Up. In our experiments, a hypersonic shock wave is produced by discharging a 14.5 microFarad capacitor that has been charged to 18 KV into an argon gas filled cylindrical tube 5 cm in diameter and about 200 cm long. In this arc driven shock tube, the speed of the shock wave depends on the pressure of the gas inside the tube and the charging voltage. The tube is divided into (diaphragmless) driver and driven sections with a test section at the end of the driven section including the end plate where the end-wall materials can be changed (see Johnson et al. (1980) for details). The test section has been recently modified by the inclusion of a neutral and ion beam generator for studies of the diagnostic and physical implications for collisional processes in a turbulent plasma. An x-t diagram showing a typical shock wave evolution is plotted in Fig. 1(a).

Specifically, the shock wave gets reflected back into the tube (toward negative x) and is thereby able to rapidly decelerate the incoming turbulent plasma flow. The end plate has surfaces of either teflon, stainless steel, or aluminum with thermal conductivities of 1.43, 15, and 190 W/m-K respectively. The progress of the shock wave is monitored by three quartz pressure transducers that are located on the wall in the test section. There is also one set of three perpendicularly arranged optical windows in the test section. The optical signals are observed by a high gain and high signal to noise ratio Hamamatsu photo multiplier tubes (PMT) and a Cordin streak camera. (See Fig. l(b).) Our interest is focused on the effects of shock wave-plasma interactions undergone by the reflected shock wave as it reprocesses the turbulent flow produced by the primary shock wave.

For the neutral emissions into the streak camera, a band pass filter centered at 420±5 nm was placed at the entrance of fiber optics to block all other signals from entering the PMT. This was also done to enable us to observe only the resonance radiation from photons of wavelength of 422.8 nm which are emitted from singly ionized argon atoms as there are no other emissions in the neighborhood of this line. The monochromator of the streak camera was setup to pick up signals of wave length of 415.8±5 nm so that emissions excited from neutral argon atoms could be observed and not be confused with the fluorescence signals.

We also used laser induced fluorescence (LIF) as a principal diagnostic technique. In the case of ionized argon gas, an excited ion pumped at 488nm emits fluorescence at 488 nm and 422.8 nm wavelengths whenever it decays from

$4p^2D_{5/2}$ state to $4s^2$ or $4s^4P_{3/2}$ states respectively (see, e.g., Stern et al. (1975)). The results reported here are restricted to fluorescence of 422.8 nm wavelength in order to avoid confusion with the 488 nm pump signal. Specifically, a CW 488 nm 5 Watt laser beam is focused at a point in the test section through one of the test section's optical windows. The fluorescence signal is collected at a 50 MHz sampling rate from a window that is perpendicular to the incoming laser and then focused at the entrance of an adjustable monochromator set at 422.8 nm; the subsequent filtered optical signal is picked up by a photomultiplier tube providing input to the digitizing oscilloscope with a 50K per channel data point storage capacity. Data extraction and computations were done for both emissions and resonance radiation using LabView and Mathematical on a Power Macintosh.

As the shock wave propagates down the tube it compresses and heats the gas. We choose operating conditions such that M_S= 19.62 (M_S is the Mach number of the primary shock wave), based on direct measurements of the primary shock wave's speed (W_S) and the room temperature. We also make a direct measurement of the speed of the reflected shock wave (W_R). Using standard calculational procedures for this kind of flow environment (see Pain (1983)), we confirm that the approximate specific heat ratio gamma = 1.333 is obtained as appropriate for a fully ionized plasma. Our measurements also support the conclusion that the gas behind the' reflected shock wave is brought to rest since we see the correct asymptotic trend in the ratio of reflected to primary shock wave speeds. Using the value of gamma obtained from the measurements, we calculate:
T_S=1.89x10^4 K; n_S = 2.24x10^{22}/m^3; T_R = 1.02x10^5K; n_R=1.3x10^{23}/m^3; and the degree of ionization α is $\alpha_S \geq 0.6$, where in all of the above, the subscripts S and R refer respectively to the primary and the reflected shock waves, M is the Mach number, and T and n are temperature and particle number density respectively. Hence the gas behind the shock wave is a high density plasma ($\geq 10^{17}$/cc).

3. **Results and Discussions.** The fractal dimensions were calculated using the standard Grassberger-Procaccia correlation function to estimate the degree of complexity of our data as illustrated in Johnson et al. (1991). The method computes the dimension of N time series data by embedding it in an M-dimensional phase space; the fractal dimensions were calculated as the slope, in a stable regime, of the Grassberger-Procaccia correlation function from the fluorescence data. The slope of the power spectrum for the turbulent density fluctuations was computed by fitting the power spectrum to P- ω^{-n} where n is the power spectral index.

Fig. 2a shows a typical pressure signal (with the background removed) as monitored by one of the pressure transducers located at the shock tube's wall; the decay in the pressure reading is an artifact of the external circuitry. The first change in pressure corresponds to the primary shock wave and the next change occurs when the reflected shock wave comes back to the same pressure transducer. In Fig. 2b, we present a Fourier fluctuation power spectrum of a subset of the laser induced fluorescence data from the time period after the reflected shock wave from which the background has been extracted. We use the (uncalibrated) voltage fluctuations as a measure of relative local density fluctuations (see Chubb (1968)). As discussed in Johnson et al. (1987), the detailed mechanisms providing the source of turbulence are not yet known; however, reaction-diffusion processes arising from the nonequilibrium recombination Argon ions with (the onset of which increases the dimensionality of the system) have a relaxation time scale appropriate (when compared with the local flow time) for increased chaotic complexity in the system and correspondingly for the development of turbulence.

More than twenty shock tube firings were analyzed for each one of the three end wall materials. Figures 3(a) and 3(b) show samples of fluorescence fluctuations for two of the materials. One notices that the relative fluctuation levels for aluminum are significantly lower than those for teflon. There is a reduction for aluminum in the turbulent intensity as well as a corresponding reduction in the overall strength of the reflected shock wave. This can be understood as a consequence of the reduction in current loses into the external circuitry at the shock tube's end plate by the use of teflon.

Figs 3(c) and 3(d) show the evolution of the average values of theFourier spectral index n for these data. These indicate that the evolution of density power law index n shows remarkably little shock-to-shock variation for either one of the two materials shown. In both cases, the first point in the graph corresponds to data taken when the primary shock wave reached the optical window; each turbulence parameter determination was calculated on 256 data points separated by 25 microseconds. Thus, for turbulence characterizing purposes, averages over time for a specific reasli-zation of our system are seen to be the equivalent of averages over different realizations of our system at a fixed time.

In Fig. 4(a) are shown the average values for the chaotic dimension for the complete set of data for the time interval before and after the arrival of the reflected shock wave. No trends in the evolution in chaotic correlation dimension (D_2) or in the density fluctuation power law spectral index were found in one substance which were not also found in all three cases. This is shown in Fig. 3(c), 3(d) and 4(a). No value, at a fixed time either just before or just after the reflected shock wave, was found in the turbulence parameters just cited for one material which was statistically inconsistent with the values found in all three cases. This means that we are free to average, wherever it may be useful, without regard for the end-wall conductivity, all values measured for the turbulence parameters.

Figure 4(b) presents the short time evolution of the chaotic dimension and the spectral index for teflon for a short time interval immediately following the reflected shock wave. Teflon is the end wall material and the number of samples in the Fourier analysis has been reduced dramatically in comparison with Fig. 4(a); there are 30 data samples in each chaotic dimension determination (corresponding to a data rate of 0.6 microseconds per point in the figure) and there are 10 data samples in each spectral index determination (corresponding to a data rate of 0.2 microseconds per point in the figure). The first point in the graph corresponds to data immediately after the reflected shock wave passes the optical window.

From Figs. 3(c), 3(d) and 4(a), the evolution of density power law index n and the chaotic correlation dimension for all the experiments makes it easy to see that the reflected shock wave causes a transient increase in the two turbulence parameters followed by a period of relaxation toward pre-reflected shock wave values. The higher value of spectral index suggests that interaction of the shock wave with the plasma tends to increase the rate of transfer of energy through the turbulent scales. The higher value of the fractal dimension suggests that the interaction of the shock wave with the plasma tends to increase the degree of complexity of its turbulence. Although there is no theoretical explanation available at present for the shapes of the curves in Figs. 3(b) and 4(a), generally stated, these numbers all fall within the expected values for turbulent flows.

However, a distinction can be made in the behaviors shown. Specifically, the rate of change before and after the reflected shock wave is the same for the parameter n; the rate of change before and after the reflected shock wave is quite different for the parameter D_2. As previously noted, the temperatures, number densities, and velocities before and after the reflected shock wave are quite different. Therefore one might infer that the chaotic dimension as a turbulence indicator is sensitive to the full range of ordinary macroscopic parameters. By contrast, one might infer that the spectral index n is a fundamental indicator of the turbulent energy transfer processes, dependent only on the microscopic (molecular) constituents.

From Fig. 4(b), one sees no systematic change in the chaotic dimension; we interpret this as an indication of a steady turbulence state. However, by contrast, there is a clear systematic increase in the spectral index just of the sort predicted by Frisch et al. (1984); we interpret this as evidence of a very short term inverse energy cascade with energy moving from the small scales into the large scales of the flow. Thus there is a clear suggestion of a coalescence of eddies (following Frisch) in these data driven by small scale phenomena which, in the case of ions, could lead to locally concentrated electric fields.

In Fig. 5(a), samples of the neutral line emissions at 415.8nm just before and just after the reflected shock wave are given. Spectra are collected at a rate of one spectrum every 300 nanoseconds. The evolution of the neutral line emission parameters is shown in Fig. 5(b). From the neutral emission, a new behavior is clearly identified. The strength of the signal in Fig. 5(b) varies as is expected in connection with the changing neutral densities associated with the ionizing shock wave. The relative line shift, however, shows an anomalous shift toward the blue over a time scale approximately the same as that indicated in the inverse energy cascade in Fig. 4(b). Before and after this period, the line is approximately as expected given the thermodynamic conditions in the plasma.

It is possible to interpret the shift as a consequence of the quasiperiodic fluctuation fields arising from the clustering of ions as is also suggested by a molecular based interpretation of Fig.4(b) as discussed above. Reminiscent of the quadratic Stark effect, the turbulence based periodic fields are apparently inducing otherwise forbidden transitions which broaden and shifts the line on average toward Stark allowed satellites. Specifically, they correspond to transitions between the outermost sub-levels of the 5P level, e.g., 5P[3/2]->4S[3/2 and 5P[1/2]-> 4S[3/2]. Similar observations have been reported by Wiegart (1983) in a turbulent high frequency He discharge, in which the evolution of the satellites corresponding to forbidden transitions in He I were monitored in time. Notice that, when the flow becomes fully turbulent (after the period associated with the inverse energy cascade) the line width decreases as the line shift diminishes.

4. Conclusions. These analyses show that point fluorescence emissions and averaged spectral line evolutions in turbulent plasmas produced in both the primary (high velocity) and the reflected (approximately zero velocity) shock wave flows provide no evidence of relaminarization and no evidence of conventional turbulent dissipation. There is furthermore an insensitivity in the turbulence parameters to the plasma-surface interactions at the end-plate of the shock tube. The data show first evidence of a reverse energy cascade in the collisional turbulent plasma. This suggests that the fully turbulent environment can by described using a stationary energy state formulation. There is also compelling evidence for a turbulent Stark effect on neutral emission spectra lines in these data which is associated with evidence (simultaneous) of structures and which are reminiscent of previous observations in the glow discharge collisional plasmas (in emissions). In light of the recent observations of reduced molecular chaos, additional justification for a molecular based approach to turbulence physics seems to now exist.

Acknowledgment. The research was supported in part by grants to Florida A. and M. University from the Fusion Energy Division of the Department of Energy and from NASA Headquarters for the NASA-FAMU Center for Nonlinear and Nonequilibrium Aeroscience.

Chubb D (1968) Phys. Fluids 11:2363
De Silva U, Zhang J, Orou JC, Johnson JA III (1996) J. Phys. D: App. Phys 29:263 (
Frisch U, Sulem PL (1984) Phys. Fluids 27:1921
Johnson JA III, Santiago J, I Lin (1980) Rev. Sci. Instr. 51: 1430)
Johnson JA III, Ramaiah R (1987) Phys. Rev. A 36:774
Johnson JA HI, Johnson LE, Hong Y (1991) Phys. Letts. 158:144
Orou J C, Johnson JA III (1994) Phys. Fluids A, 6:415
Pain J (1983), The Physics of Vibrations and Waves, John Wiley and Sons 372-374
Stern RA, Johnson JA III (1975) Phys. Rev. Letts. 34:1548
Tsuge S (1974) Phys. Fluids 17:22
Wiegart NJ (1983) Phys. Rev. A 27:2114

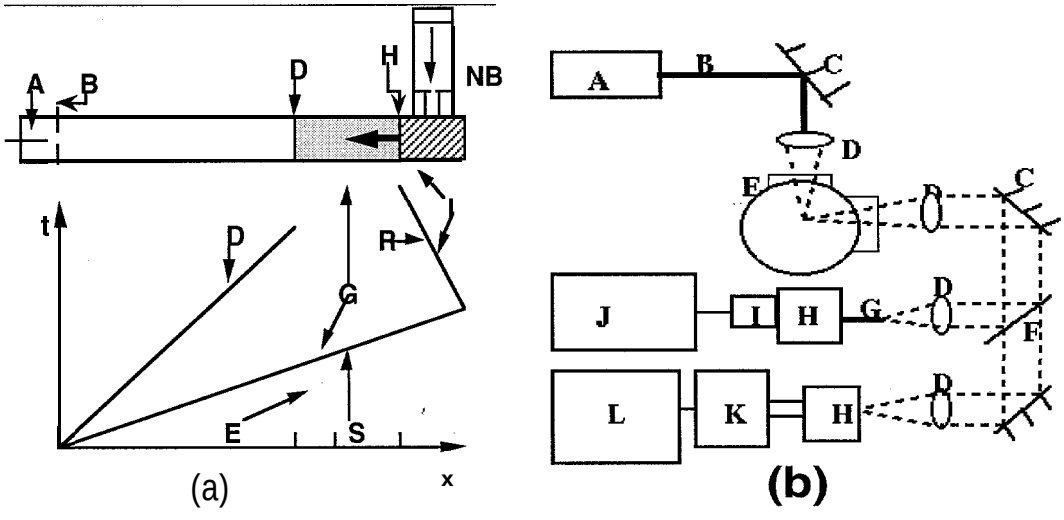


Figure 1. (a) An x-t Diagram Showing the Arc Driven Shock Wave Phenomena. A = High Voltage Electrode at the Set-Up Charging Voltage, B = Ground Electrode, D = Blast Wave, E = Ambient Arc Driven Shock Tube Gas, Prior to the Primary Shock Wave at the Set-Up Pressure, G = Post Primary Shock Wave Plasma, S = Primary Shock Wave, R=H= Reflected Shock Wave, I = Post Reflected Shock Wave Plasma, NB=Neutral Beam. The fluorescence measurements are made at a point 5.5 cm from the end-wall and 159 cm from the ground electrode. (b) Data Acquisition System: A = Laser, B = Laser beam, C = Mirror, D = Lens, E = Window, F = Beam splitter G = Fiber optic, H = Monochromator, I = PMT, J = TDS, K= Streak camera, L = Computer

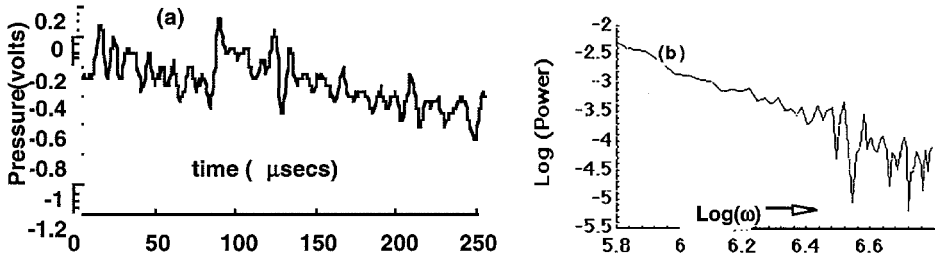


Figure 2. Sample Data. (a) Typical pressure signal of shock wave: The changes in pressure correspond to the primary and reflected shock waves respectively. (b) Power spectrum for the data from fluctuations in fluorescence in the post reflected shock wave region.

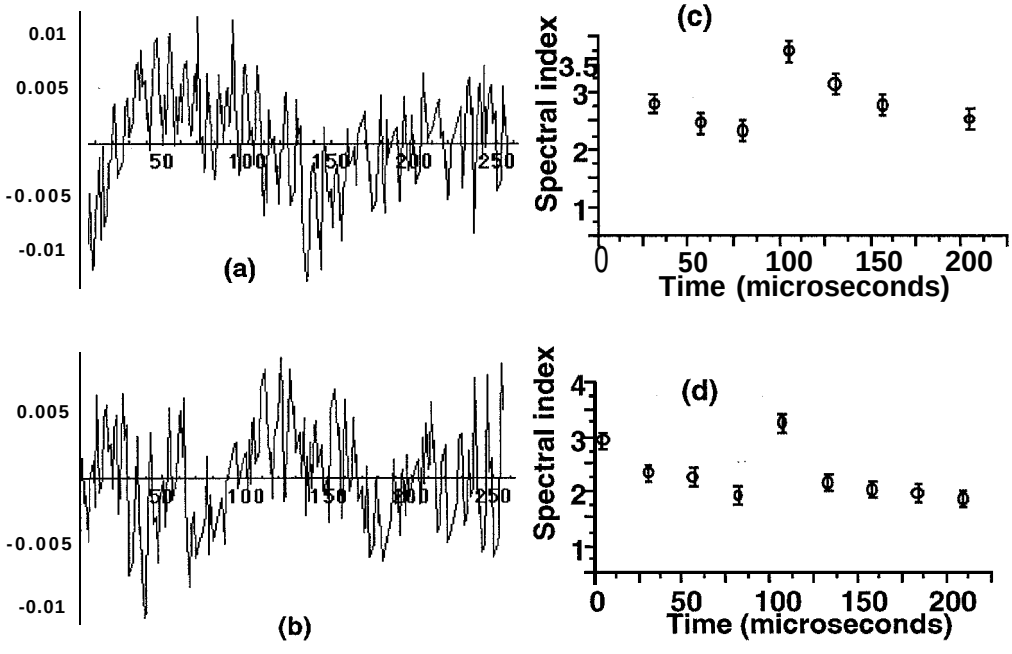


Figure 3. (a) Sample Fluctuations in Fluorescence with Teflon as the End Plate Material. (b) Sample Fluctuations in Fluorescence with Aluminum as the End Plate Material. (c)and (d) Evolution of the Fourier Spectral Index for the Fluctuations in(a) and(b) Respectively Averaged over 10 Shock Tube Firings. The time is measured from the arrival of the primary shock wave through the arrival of the reflected shock wave at the measurement station roughly 100 microseconds later.

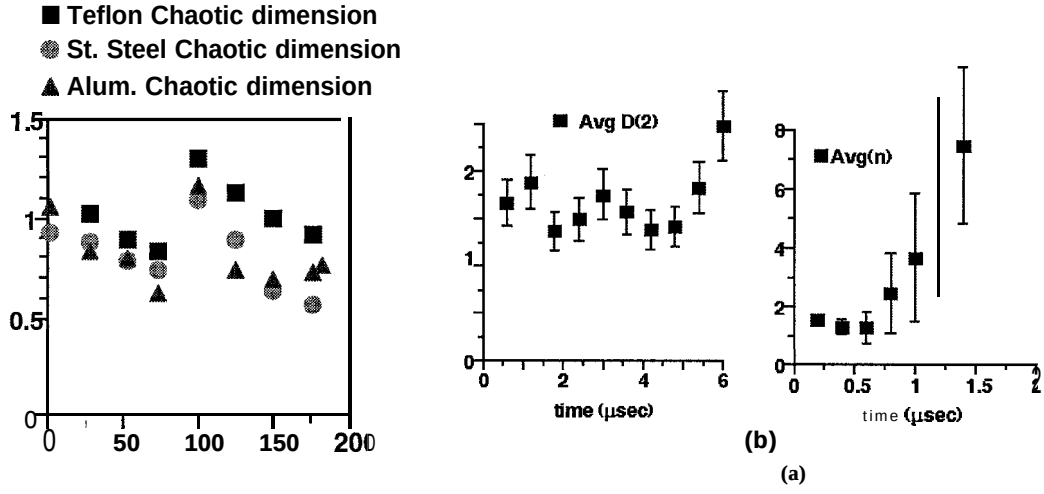


Figure 4. (a) Variation of Chaotic Dimension After the Primary Shock Wave and After the Reflected Shock Wave. (b) Variation of Spectral Index and Fractal Dimensions with Time: Immediately After the Reflected Shock Wave for a Teflon Coated Surface.

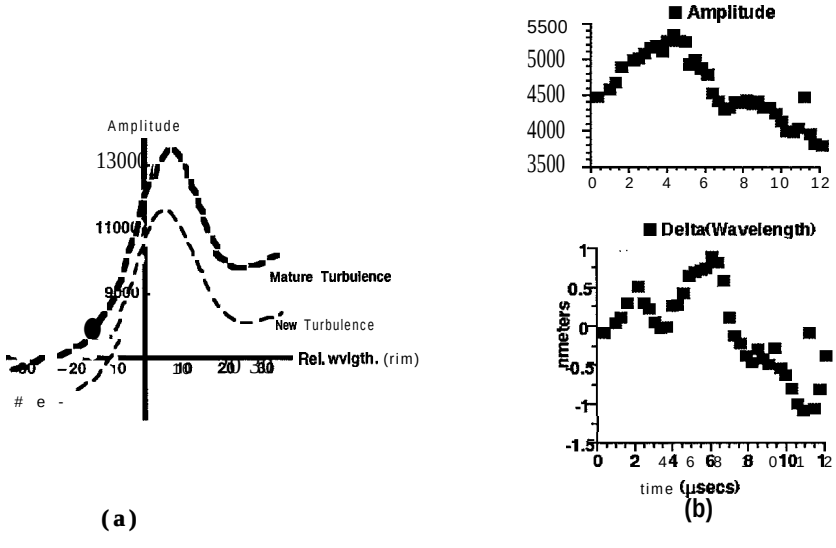


Figure **5.** (a) Sample Data Neutral Line Emissions with Blue Shift in Early Turbulence. (b) The Evolution of Neutral Line Emission Parameters.

Water Injection for Mixing Noise Reduction of a Heated Supersonic Rectangular Jet

D. Washington and A. Krothapalli
Center for Nonlinear and Nonequilibrium Aeroscience
Florida A&M University and Florida State University

Introduction/Motivation

Over the past several decades researchers have been interested in reducing the noise of supersonic jets. The starting point for most of these noise suppression techniques comes from theories concerning aerodynamic noise from Lighthill's[1] analogy. In Lighthill's theory, the equations of motion are broken down into the inhomogeneous wave equation where a distribution of acoustic sources (in the flow field) are responsible for the inhomogeneous terms. Unfortunately, this source distribution is of unknown origin and magnitude. Consequently researchers have had difficulties devising effective noise suppression techniques without incurring significant thrust penalties on existing aircraft. One currently available noise suppression method is to use non-circular nozzle geometries in order to produce streamwise vortices in the jet flow field. This idea comes from Pannu and Johannesen[2] who suggested that these types of structures could redirect the noise in jets. However, subsequent investigations[3] have shown that streamwise vortices create at best only modest magnitudes of noise reduction in the jet Overall Sound Pressure Level (OASPL). Despite the many active and passive noise suppression schemes devised by researchers until now, no method has been developed which has lead to OASPL reductions significant enough for use in the High Speed Civil Transport (HSCT) program. This is evident in figure 1, where a compilation of the noise suppression results for the past several decades is shown. In this figure one can see that in the 1970's, a 15 percent thrust penalty was sustained in order to produce a 12 dB reduction in jet noise. Similarly, in the 80's and 90's this improved to a 7 and 5 percent thrust penalty respectively, to reach this 12 dB reduction. However, the goal of the HSCT program currently has not been met, since it requires that a 12 dB reduction in jet noise can have at most a 2.5 percent thrust penalty.

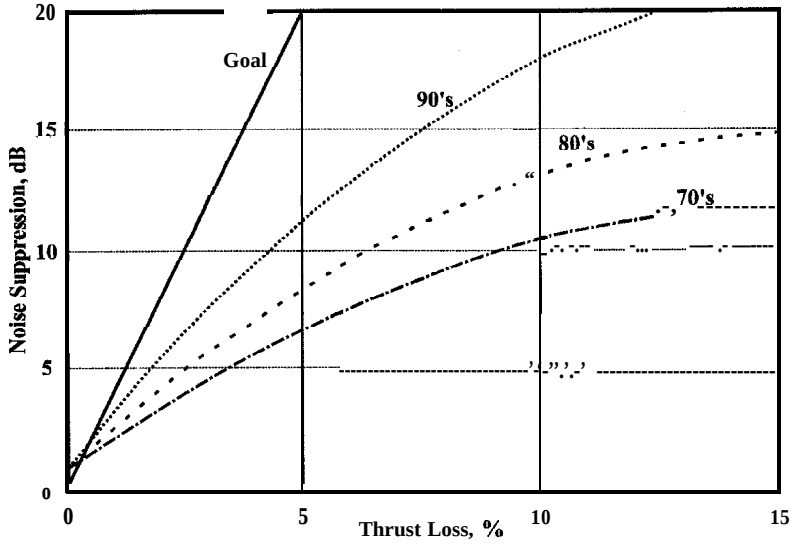


Figure 1 Noise Suppression Technology

It is known that supersonic jet noise consists of three components, which are broadband shock associated noise, screech tones, and turbulent mixing noise. The former two components are

created near the nozzle exit are blocked. Conversely, if the low frequency waves remain after the high frequency waves are blocked, then the strongest turbulent mixing noise sources exist downstream of the potential core. Hence, in order to resolve this issue an experiment is devised to visually look at the acoustic waves produced in a jet with and without a control present around the jet perimeter. For the purpose of this experiment, the control is fixed at a specific streamwise location. In addition, the control has a rectangular hole in the middle of the plate. This allows the jet to pass through its center without being obstructed. A schematic of this experiment can be seen in figure 3.

In figure 4, instantaneous shadowgraph images of an ideally expanded Mach 2.0 jet with a stagnation temperature of 671.4 K is shown. The two images in the figure represents a jet with and without a control present in the flow field. Both images show high frequency waves near the nozzle exit, while low frequency waves persist in the downstream direction. Moreover, when the control is inserted around the jet periphery, it is clear that the high frequency waves are reflected back upstream. Remarkably, the low frequency waves are still created by the jet when the control is in place. Thus, strong evidence is provided to support the notion that the bulk of turbulent mixing noise sources are produced downstream of the potential core.

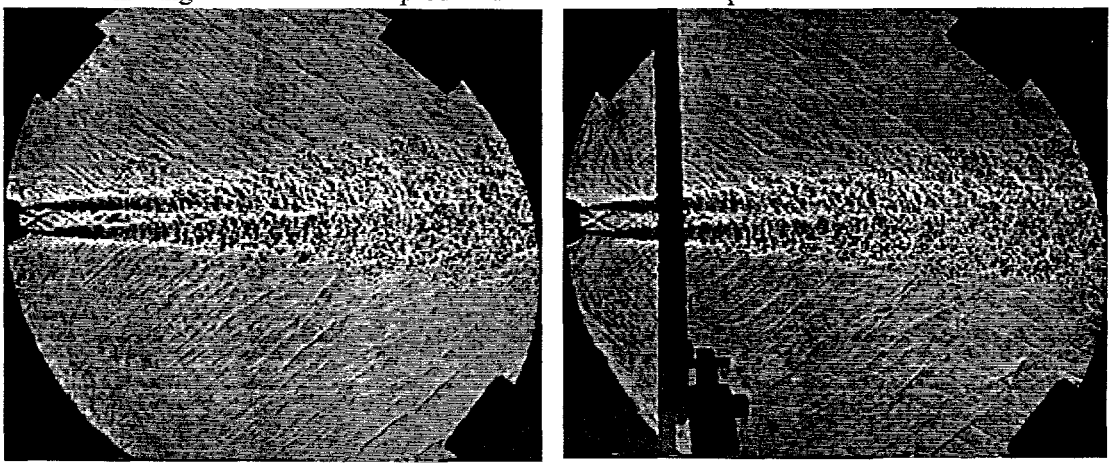
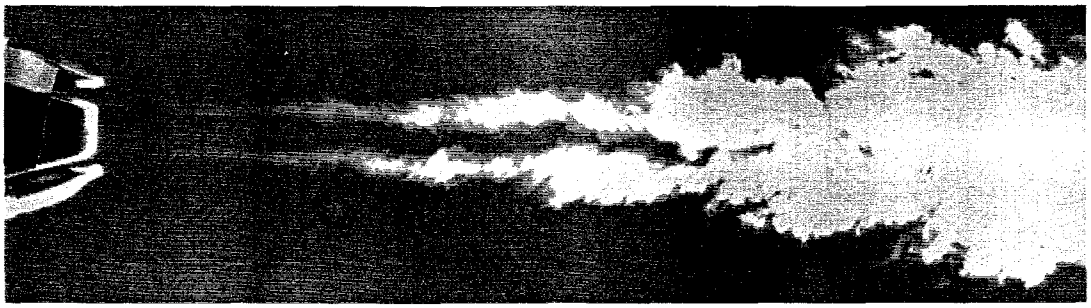
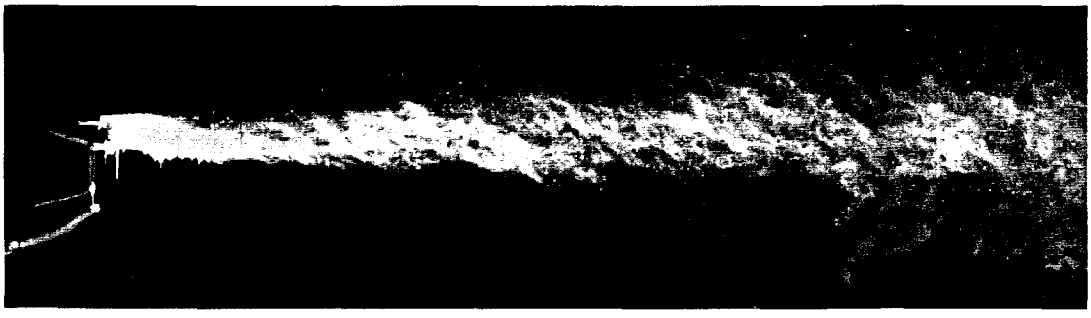


Figure 4 (a) Free Jet Case (left image); (b) Control Case (right image)

The results of the previous section suggests that in order to reduce the noise of supersonic jets, one must suppress the source of the turbulent mixing noise downstream of the potential core. One possible solution to this problem is to inject water into this region of the jet. Therefore, in Figures 5 instantaneous PLS images of a cold Mach 2 jet with (figure 5b) and without (figure 5a) water injected are shown. The bright streak in figure 5b is caused by the laser light



(a)



(b)

Figure 5 Instantaneous PLS Images (a) Free Jet Case; (b) Top Shear-Layer Water Injection

retracting off a water droplet on the injecting nozzle. The results of the PLS images qualitatively indicates that the addition of water significantly alters the shear-layer dynamics of the jet. This evidence comes from the fact that the large scale corrugated structures in the top shear-layer of figure 5b are reduced in size as atomized water freely entrains into the jet. Furthermore, it is expected that the pressure fluctuations due to the shear-layer instabilities are altered as a result. Based upon the evidence presented thus far, it is anticipated that the far-field OASPL should be modified due to low mass flows of injected water. However, until the acoustic measurements of this investigation are completed the magnitude of the change in the OASPL will not be known.

Conclusion

Evidence has been presented in the current paper which supports the idea that the source of the turbulent mixing noise is located downstream of the potential core. In an effort to suppress the turbulent mixing noise, atomized water is injected into a heated Mach 2.0 rectangular jet. The injection of water into the jet causes a change in the overall structure of the shear-layer instabilities. This is clearly evident in the PLS images of Figures 5. It is believed that the shear-layer modification attributed to water injection will cause a reduction of the OASPL in the jet even for low mass flow ratios of water to air (approximately 0.1). This assertion comes from the fact that a portion of the jet energy will be required to complete the evaporation process of the atomized water. Furthermore, the momentum exchange between the water and air in the jet shear-layer will add to this OASPL suppression. Thus, preliminary experimental results have shown positive benefits toward mixing noise reduction. In addition to the results already presented, aerodynamic experiments will be conducted in the future. These experiments will consist of pressure surveys, PLS images, PIV, and IR images. The combination of this experimental data will be used to determine how the shear-layer dynamics of the jet are modified as water is injected into the jet. This will allow us to interpret the results of the far-field acoustic measurement that will also be obtained in a subsequent investigation.

References

[1]Lighthill, M. J., "On Sound Generated Aerodynamically: L General Theory," *Proc. Roy. Sot. London, Ser. A*, 211, Mar. 1952, pp. 564-587

[2]Pannu, S.S. and Johannesen, N.H., "The structure of jets from notched nozzles," *Journal of Fluid Mechanics,* vol. *7,* part *3, 1976,* pp. 515-528

[3]Krothapalli, A., Alvi, F. S., Washington, D. and King, C.J., "Aeroacoustic Characteristics of a Supersonic Diamond-Shaped Jet," First Joint CEAS/AIAA Conference, Munich, 1995.

[4]Harper-Bourne, M. and Fisher, M.J., "The Noise from Shock Waves in Supersonic Jets," *Proceedings of the AGARD Conference on Noise Mechanisms,* AGARD CP-131, 1973, pp. 11-1 to 11-13.

[5]Tam, C. K. W., "Supersonic Jet Noise," *Annual Review of Fluid Mechanics, 27, 1995,* pp. 17-43.

[6]Tam, C. K.W., Golebiowski, M., and Seiner, J.M., "On the Two Components of Turbulent Mixing Noise from Supersonic Jets," AIAA paper 96-1716, May 1996, AIAA/CEAS 2"' Aeroacoustics Conference, State College, PA

[7]Zoppellari E., and, Juvé, D., "Reduction of Jet Noise By Water Injection," AIAA paper 97-1622-CP, May 1997, AIAA/CEAS 3" Aeroacoustics Conference, Atlanta, GA

[8]Lighthill, M.J., "Some Aspects of the Aeroacoustics of High-Speed Jets," *ICASE Report, 93-20,* May 1993, pp. 1-43

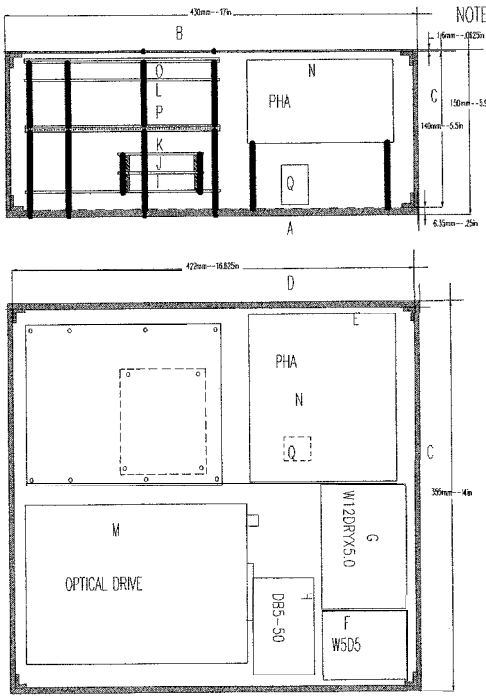
VERIFICATION AND QUANTIFICATION OF SINGLE EVENT EFFECTS ON HIGH SPEED SRAM IN TERRESTRIAL ENVIRONMENTS

H. Huff, Z. You, D. Williams, T. Nichols, J. Attia, T. N. Fogarty, K. Kirby, R Wilkins
PVAMU-CARR

R Lawton, Space Environments Consulting

As integrated circuits become more sensitive to charged particles and neutrons, anomalous performance due to single event effects (SEE) is a concern and requires experimental verification and quantification. The Center for Applied Radiation Research (CARR) at Prairie View A&M University has developed experiments as a participant in the NASA ER-2 Flight Program, the APEX balloon flight program and the Student Launch Program. Other high altitude and ground level experiments of interest to DoD and commercial applications are being developed. The experiment characterizes the SEE behavior of high speed and high density SRAM's. The system includes a PC-104 computer unit, an optical drive for storage, a test board with the components under test, and a latchup detection and reset unit. The test program will continuously monitor the stored checkerboard data pattern in the SRAMs and record errors. A version of the testbed is shown in Figure 1. Since both the computer and the optical drive contain integrated circuits, they are also vulnerable to radiation effects. A latchup detection unit with discrete components will monitor the test program and reset the system when necessary. The first results will be obtained from the NASA ER-2 flights, which are now planned to take place in early 1998 from Dryden Research Center in California. The series of flights, at altitudes up to 70,000 feet, and a variety of flight profiles should yield a distribution of conditions for correlating SEES. SEE measurements will be performed from the time of aircraft power-up on the ground throughout the flight regime until systems power-off after landing.

FIGURE 1: SEE TESTBED CONFIGURED FOR 400 HZ POWER SUPPLIES



NOTE: TOP AND FOUR SIDES ARE TO BE A UNIT REMOVABLE FROM THE BASE

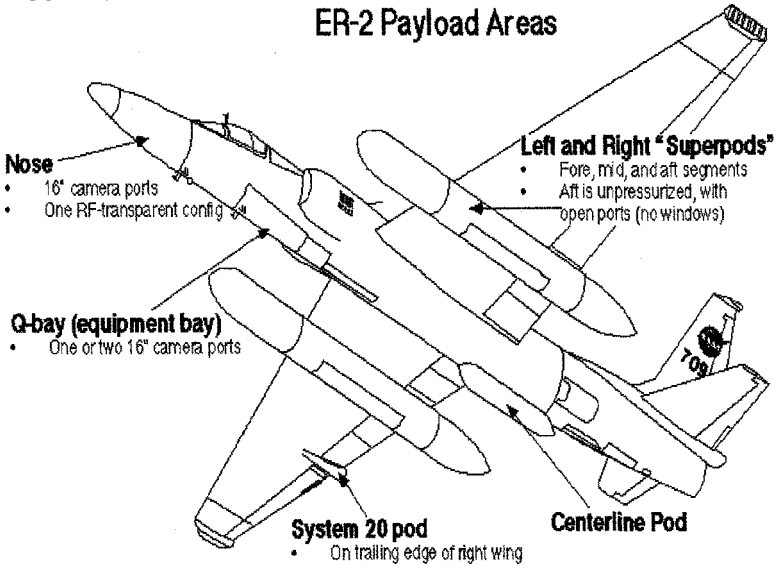
SYSTEM COMPONENTS

	THE BOX		
A	BASE	14"X17"X 25′	ALUMINUM
B	TOP	14"X17"X.0625"	ALUMINUM
C	2 SHORT SIDES	14"X5.5"X.1875"	ALUMINUM
D	2 LONG SIDES	16.625"X5.5"X.1 875"	ALUMINUM
E	ANGLE BRACES	.5"X.5"X.125"	ALUMINUM
	POWER SUPPLYS		
	ABBOTT		
F	W5D5	2 5"X3.5"X4.0625"	2.7lb
G	W12DRYX5 o	3.5"X4.5"X4.25"	2.7lb
	ACOPIAN		
H	DB5-50	3.5"X2.5"X1 85"	3lb
	CIRCUIT BOARDS		
I	SINGLE BOARD COMPUTER--ADVANTEC 4860	8"X5"X1"	
J	10 BOARD--DIAMOND SYSTEMS F07149	3.55"X3.775"X 625"	
K	SCSI-104--ADASTRA SYSTEMS mod800-400	3.55"X3.775"X.625"	
L	CUSTOM SURFACE MOUNT TEST BOARD	8"X5"X1"	
	NOTE STACK WEIGHT	<11b	
	ADDITIONAL DEVICES		
M	OPTICAL DRIVE--MOUNTAIN OPTECH P/N 2436-00 9	75"X5.75"X3.5" 2lb 13.4oz	
N	CUSTOM PULSE HIEGHT ANALYZER	6"X6"X3"	
o	2 CR39 PLASTIC SHEETS	8"X5"X 0625"	
P	POLYETHYLENE SHEET	8"X5.75"X 25"	
Q	LATCH-UP PROTECTION RELAY I 4?	"X1.1"X.85"	

FLIGHT PLATFORMS AND PROFILES

The ER-2(Figure 2) is the primary flight platform identified for use with SEE testbed (Figure 1). Its high altitude (60 to 70 thousand ft) normal cruise profile will provide data about SEES with the aircraft operating above 90% of the earth's shielding atmosphere. The first location identified on the ER2 is in the right Superpod along with other experiments from the Atmospheric Ionizing Radiation (AIR) effort. Alternately discussions with Lockheed Martin have identified the "System 20 Pod" as an ideal location for the testbed. The location is seldom used by other experiments because of its small size (9 in Dia. X 20 in. L). This maximizes the flight opportunities for the testbed which also makes a wider geographic range of data possible. Given enough flights a multi-dimensional SEE distribution profile maybe developed.

FIGURE 2:



The SR-71 Blackbird which is also stationed at the Dryden Flight Research Center, would allow the even higher altitude (up to 85 thousand ft. normal flight profiles and 95 thousand extreme case) data to be gathered. Because of the testbed's small size it could piggyback with other flight experiments, expanding the Blackbirds data to cost ratio. Use of this platform is in the preliminary feasibility study phase.

Project APEX reaches very high altitudes (100 to 120 thousand ft.) for extended periods of up to several days. The project's central experiment is aimed at determining high altitude Reynolds number data for a balloon launched remotely flown glider. The SEE testbed's small size and it's ability to operate on battery power makes it an ideal piggyback payload on currently planed flights. Discussions of possibly testing other ICs using the testbed and this platform are being considered. The effort will be funded by NASA's Faculty Award Research Program Dr. Richard Wilkins is the Prairie View principal investigator with this program.

The Student Launch Program has high altitude balloon flight capabilities (90 thousand ft) and provides an excellent platform for this testbed. It is also a great opportunity for students to get involved in solving real engineering problems. They will be responsible for developing power management plans for energy usage from the battery pack. Operating environment analysis and compensation will be crucial for the students to address in developing the testbed's physical and electrical support systems. The students will also work on data handling and analysis.

Prairie View's students have a exciting opportunity in the form of the Getaway Special program. The reservation for this payload was donated by the employees of McDonnell Douglas, Houston (now Boeing). It is expected that the engineering student group will configure the SEE testbed and support systems into a self contained form for flight on the Space Shuttle within the next two years It also provides another set of SEE mapping data crucial to developing safer more reliable electronics for low earth orbit applications.

Space Industries International which has built custom circuitry to the Center for Applied Radiation Research's specifications for the SEE testbed has identified an exciting possible platform on the back side of the University of Houston's Space Vacuum Epitaxy Center's Wake Shield Facility. The location would be ideal for the testbed's purposes because the back side of the Wake Shield is always pointed towards the sun to block the solar wind from the epitaxy experiments.

The Mars Robotic Precursor flights are the most exciting platforms yet contemplated for this testbed. Study of the recent Mars Exploration Technology Survey shows an urgent necessity for the data that can be gathered by the SEE testbed. The SEE testbed can provide information important to most categories listed in the survey. Specifically the data collected by the testbed will be crucial in efforts to qualify control electronics for virtually all systems exposed to the flight environment. The testbed's ability to provide real error rate data for specific environments and components makes it a necessary instrument for the early exploratory flights. The testbed's light weight, small power requirements and the usefulness of the data collected make it a natural piggyback candidate.

TEST SETUP AND TEST PLAN

Figure 3 is the system configuration for the test setup. The PC/104 CPU module is an embedded-PC controller that offers full hardware and software compatibility with the PC bus built in ultra-compact (3 .6"x 3.8") stackable modules. The compact form factor and the low power consumption (1-2 watts per module) make it really convenient for our system.

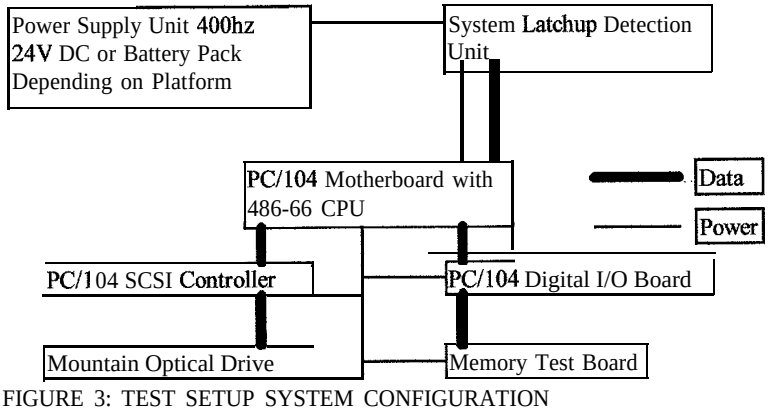
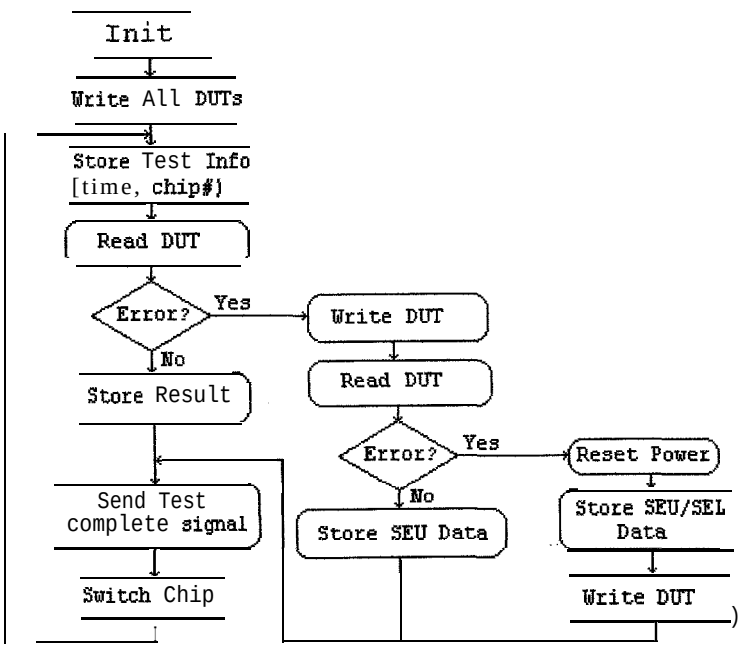


FIGURE 3: TEST SETUP SYSTEM CONFIGURATION

The CPU loads and runs the program from the optical drive which is controlled through the SCSI controller. The data gathered is stored to the drive for downloading later. The reason that we chose optical drive as our storage media is that from previous test results we found that normal magnetic hard disk drives do not function well under low pressure conditions. The Mountain Optical Drive we are using is specially designed for high-G impact protection and has been tested by NASA Langley Research Center to be flight qualified. The test vector is generated through the digital I/O board which has 48 I/O channels. There is a 100KHz test speed limit on these digital I/O boards, but since the testis designed as a fictional test, the test speed is not critical. However it will affect the amount of time used to test each chip and therefore the number of chips that we can put on the test board.

The memory devices under test (DUT) are state-of-art Motorola high speed (20ns access time) high density (4Mb) Static RAM MCM6246. The chips are equipped with chip enable (E) and output enable (G) allowing for greater system flexibility and eliminating bus contention problems. Either input, when high, will force the outputs into high impedance. This device has been tested at Texas A&M University Cyclotron Institute, and we have a complete data set for various particles. The DUT are mounted on the test board.

Since the test system itself contains integrated circuits which are as vulnerable to SEU/SEL (Single Event Upset/Single Event Latchup) as the DUT, we first considered putting plastic shielding above the test system. But considering the very high energy cosmic ray particles which will penetrate basically everything, a special unit(System SEU/SEL Detection Unit) was designed using discrete components. This Unit will receive a signal from the CPU module at certain intervals indicating that the test program is functioning correctly. When the CPU module fails to send the signal this indicates an SEU/SEL has occurred in the test system, the detection unit will cycle the power of the test system. The test program will store the time that the system power been cycled for post test analysis.



ER-2 Memory Test Flow

Figure 4

Figure 4 shows the test flow of the test program developed using C language. During the initialization the power up time is recorded to the optical drive. After that, a checker board data pattern is written to all the DUT chips on the test board. Then the test program checks each individual chip for errors. It should be noted that from past tests we observed that some errors caused by heavy ions can only be corrected by cycling the power of the memory chip. However these errors did not lower the power supply voltage by withdrawing significant amount of current which is an unique sign of Latchup. We classify these errors as memory cell SEL. From the test program flow chart we can see that when an error is detected, we will rewrite the data pattern and cheek again to identify if it is a SEU or SEL. Each chip has a chip number associated with their test data in the file stored in the optical drive. The test program will continue to run until power has been turned off.

CORRELATING INSTRUMENTS AND DATA CONFIGURATIONS

Each of the platforms the testbed will fly on log a flight profile for each flight. That profile provides time, altitude, longitude, latitude and in most cases temperature and pressure data. These data combined with the testbed's log of SEES and SELs allows a multi-dimensional map of SEES to be created. Given data from a number of flights relationships between these parameters and others maybe established. Data about the radiation environment is crucial to using the testbed's results and relating them to the radiation hardness of the devices under test. Comparing the field measured radiation environment and SEE data to Laboratory developed data may extend the usefulness of laboratory experiments as well. Two Instruments are currently being considered for providing correlating radiation environment data

The Pulse Height Analyzer developed by the Radiation Physics Office (RPO) of Goddard's Earth Science Directorate in cooperation with Clemson University. The analyzer uses arrays of p-n junctions formed in silicon. This analyzer utilizes structures similar to those sensitive to SEES, It's range of sensitivity and response to energetic particles correlates well with the testbed's most interesting data regions. It has been developed with space flight and high altitude measurement in mind and is correspondingly small and light weight which makes it easy for us to incorporate it into our flight package.

The Liulin-3 dosimeter-radiometer is a similar instrument developed under a cooperation agreement between Bulgarian, Russian, American, and German scientists for experiments on board the Bion-11 satellite. It was developed specifically to assess the radiation environment of low earth orbit flights. Its primary ranges were developed for biological hazard assessment but may still provide data useful in evaluating the conditions of concern to us.

ACKNOWLEDGMENTS

This work has been supported by NASA—CARR and a grant from NASA Dryden Flight Research Center. The authors wish to express their particular appreciation to their coworkers at the Center for Applied Radiation Research and at Prairie View A&M University.

The authors would like to thank the following for many useful discussions and support.

E.G. Stassinopoulos	Goddard Space Flight Center
Kumar Krishen	Johnson Space Center
Don Maiden	Langley Research Center
John Wilson	Langley Research Center
Kurt Kloesel	Dryden Flight Research Center
Irma Cox	Dryden Flight Research Center
Steve Paterson	Ames Research Center
Harvey Kent	Lockheed Skunkworks

REFERENCES

E.G. Stassinopoulos & P.J. McNulty "Anew Approach To Space Radiation Effects Measurements. Direct measurement of Linear Energy Transfer (LET) Values. NASA's Space Environment and Effects Program Fall 1996 issue.

NASA/GSFC-RPO Technical Description of Liulin-3

"PC/104 Resource Guide", April 1996. Ninth Edition

C. Kouba, "Single Event Upset Characteristics of the 486-DX4 Microprocessor" Radiation Effects Data Workshop, 34th Annual International NSREC 1997.

Southern Great Plains 1997 hydrology experiment: The spatial and temporal distribution of soil moisture within a quarter section pasture field

T. Tsegaye*[1], T. Coleman[1], W. Tadesse[1], N. Rajbhandari[1], Z. Senwo[1], W. Crosson[2] and J. Currency[1]

[1]Alabama A&M University, Center for Hydrology, Soil Climatology, and Remote Sensing (HSCaRS)
P. O. Box 1208, Normal, AL 35762

[2]Global Hydrology and Climate Center, 977 Explorer Blvd., Huntsville, AL 35806

ABSTRACT

Understanding the spatial and temporal distribution of soil moisture near the soil surface is important to relate ground truth data to remotely sensed data using an electronically scanned thinned array radiometer (ESTAR). The research was conducted at the A-ARM EF site in the Little Washita Watershed in Chickasha, Oklahoma. Soil moisture was measured on a 100 x 100- m grid on a quarter section (0.8 km by 0.8 km) size field where the DOE A-ARM SWATS is located. This site has several drainage channels and small ponds. The site is under four different landuse practices, namely active pastureland, non-grazed pastureland covered with thick grass, forest area covered with trees, and a single residential area. Soil moisture was measured with a Time Domain Reflectometry (TDR) Delta-T 6-cm theta-probe and gravimetric soil moisture (GSM) technique for the top 6 cm of the soil depth. A fourth order polynomial equation was fitted to each probe calibration curve. The correlation between TDR and GSM measurement technique ranges from 0.81 to 0.91. Comparison of the spatial and temporal distribution of soil moisture measured by the TDR and GSM techniques showed very strong similarities. Such TDR probes can be used successfully to replace the GSM techniques to measure soil moisture content rapidly and accurately with site specific calibration.

INTRODUCTION

Accurate and reliable data of soil moisture is required in several remote sensing, hydrological and agricultural related works. Soil moisture estimates provide information on the quantity of water in the soil profile and soil properties such as texture, organic matter, pore size distribution and bulk density that influence the availability of this quantity. The spatial and temporal distribution of soil moisture and its estimates can be used to derive hydrological parameters including surface and subsurface drainage, infiltration, runoff, evaporation, and transpiration (Crosson et al, 1997)

Remote sensing scientists have expressed a strong need for an independent and well-documented source of soil moisture data that can be used for regional scale model initialization. Large databases obtained by active and passive microwave imagery are often considered as a source of this information. One of the difficulties of obtaining such data using remote sensing techniques is that the sensitivity to detect soil moisture varies with vegetation type and soil properties.

Soil sampling techniques used as ground truth for remote sensing are most often not adequate due to the time and cost associated with them. The need for fast, less expensive, accurate, and less destructive measurement techniques in the remote sensing foot print area is increasingly growing. The time domain Reflectometry (TDR) may offer an alternative way to quickly measure soil moisture in the upper few centimeters of the soil depth (Tsegaye et al., 1997). The probe may also be au alternative for acquiring soil moisture observations within the experimental areas where the logistics of traditional gravimetric soil moisture (GSM) sampling techniques can restrict the number of samples collected. The objectives of this study were (1) to develop and examine TDR calibration curve for two types of soil texture within the Little Washita Watershed and (2) to compare the VWC values measured by a TDR device with the GSM technique, and to examine the similarity of the spatial distribution of soil moisture within a pasture field.

MATERIALS AND METHODS

The research was conducted within an 800 by 800-m pasture field in Little Washita River Watershed in Southwest Oklahoma. The Watershed covers 235.6 square miles and is tributary of the Washita River. The climate at this location is classified as moist and sub-humid and the average annual rainfall is approximately 885 mm. In this Watershed, the

summer is typically long, hot, and relatively dry. The average daily high temperature for July is 94°F, and the average accumulative rainfall for July is 56 mm. The topography of the land is gently to moderately rolling and the average slope within the entire sampling location ranges from 1 to 5 percent. This study focused mainly on a continuous pasture field. Such fields are particularly suitable for repetitive sampling within a watershed because the land-use management of such fields rarely changes and the vegetation cover and roughness characteristics vary little throughout the study period.

The traditional gravimetric soil moisture (GSM) and the thetaprobe which follow the Time Domain Reflectometry (TDR) principle were used to measure volumetric water content (VWC) within pasture and wheat fields. The GSM technique is the ratio of the mass of water present in the soil sample to the mass of the sample after it has been dried at 1050C. The bulk density of the soil was used to convert mass soil water to volumetric soil water content. The VWC is the ratio between the volume of water present and the total volume of the sample. It is a dimensionless parameter, expressed either as a percentage by volume or as a ratio in m^3/m^3. It was measured with a four-rod theta-probe manufactured by Delta-T Devices (DELTA-T DEVICES LTD, 1996). The theta probe measures VWC by the well-established method of responding to changes in the apparent dielectric constant. These changes were converted into direct current voltage, virtually proportional to the soil moisture content over a wide working range.

Soil moisture measurements and GSM sampling were done on a 100 X 100-m grid on a pasture field and on a 10 X 5-m grid on a wheat field. Forty-nine sampling points were established throughout the pasture field to determine the spatial and temporal distribution of soil moisture within a single pixel of remote sensing foot prints area. On each sampling point, both the TDR and the GSM measurements were taken on the same location. The GSM sampling was collected using a specially designed 6-cm^3 sampling scoop.

RESULTS AND DISCUSSION

Calibration curves developed for the probes used in SGP'97 field experiment are shown in Fig. 1-6. Each thetaprobe is adjusted by the manufacturer to provide a consistent output for a known measuring media i.e., for organic or mineral soil making it readily available for immediate use. The response from each thetaprobe was quite similar for homogeneous media. However, when the soil is highly variable and the drainage pattern and soil texture change it is necessary to have a site-specific calibration to minimize measurement errors among instruments. For both the BERG and LW03 DOE site, a fourth order polynomial equation was fitted to each calibration curve. These probes were giving similar results for saturated and totally dry soil conditions. However, when the soil is between field capacity and wilting point, the probes show some variations in voltage output for a sandy clay loam soil at the berg site versus the sandy loam textured soil at the LW03 DOE site. Such variations in voltage output should be calibrated and the data can then be used to relate to remotely sensed or hydrological measurements.

The calibrated volumetric water content data from the thetaprobe measurements and the converted data from GSM to VWC were compared for three date, June 21, 24, and 28, 1997, and the result are shown in Table 1. The mean and variance associated with the two measurement techniques were very comparable. The mean VWC were higher for TDR measurements with lower variance values associated with the measured data sets. Lower mean values of GSM technique may have been the result of drying effect due to transportation of samples between field and laboratory. The correlation, on the other hand, was ranging from 0.86 to 0.91 indicating that the TDR technique can replace the GSM technique without loosing too much information.

Fig. 1. Calibration curve of theta probe DT202.10 for LW03 site.

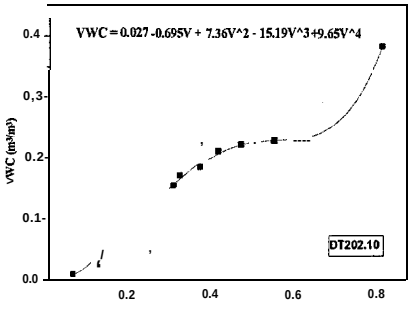


Fig. 2. Calibration curve of theta probe DT204.2 for LW03site.

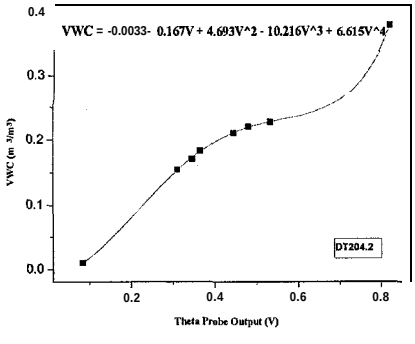


Fig. 3. Calibration curve of theta probe DT203.9 for LW03 site.

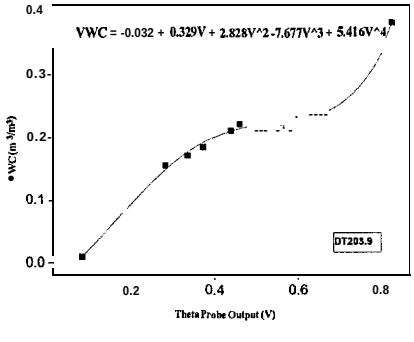


Fig. 4. Calibration curve of theta probe DT202.10 for Bergsite

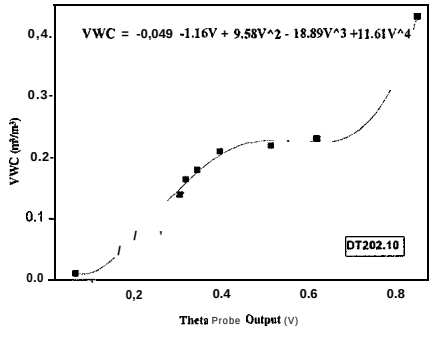


Fig. 5. Calibration curve of theta probe DT204.2 for Bergsite.

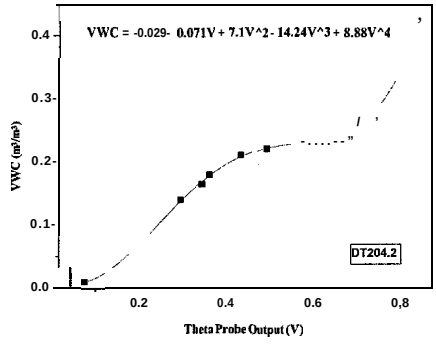
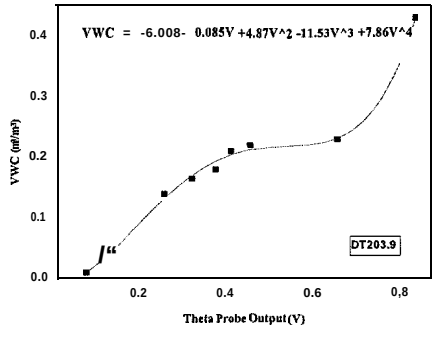


Fig. 6. Calibration curve of theta probe DT203.9 for Bergsite.



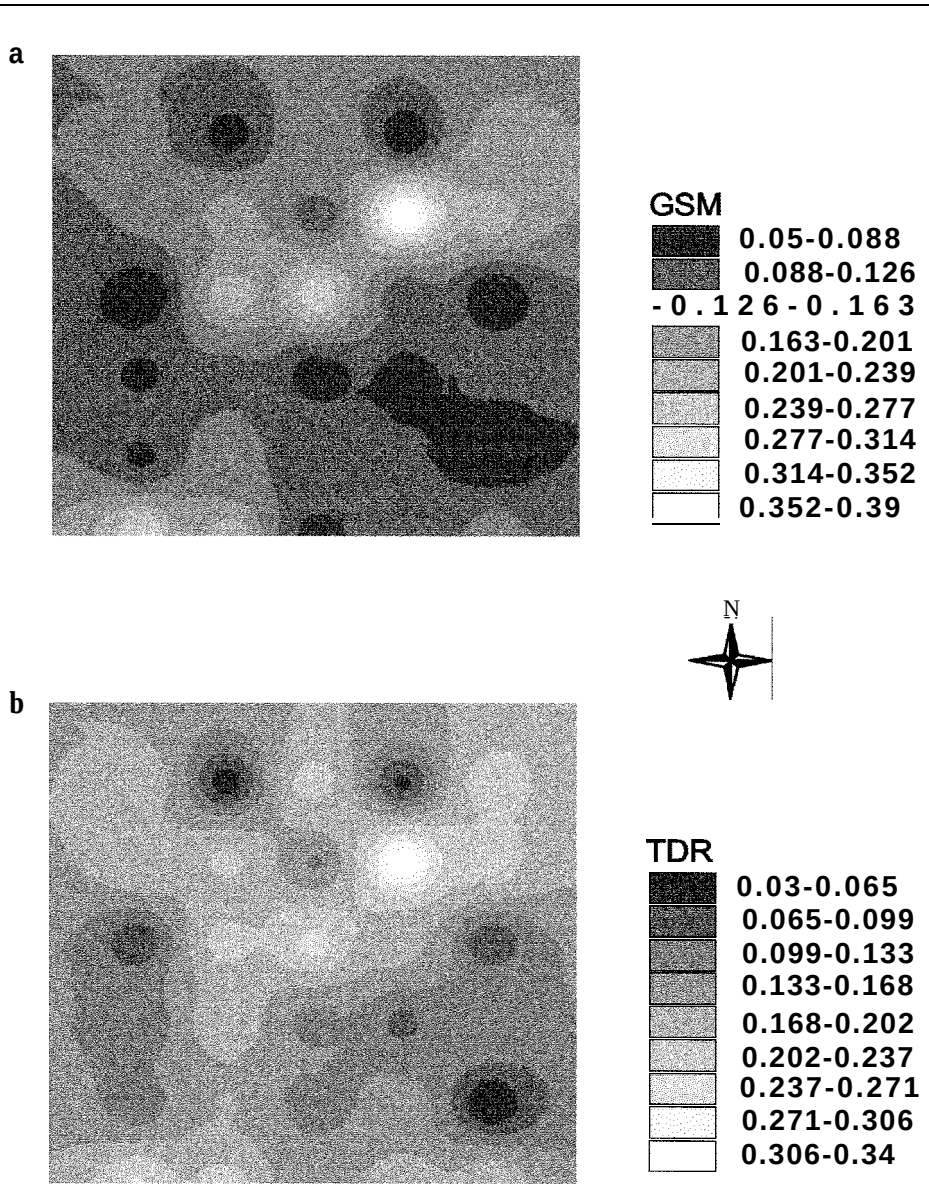


Figure 7. Comparison of the spatial distribution of soil moisture for 0 -6 cm depth by (a) GSM and (b) TDR measurement techniques for pasture field (sampled June 21, 1997).

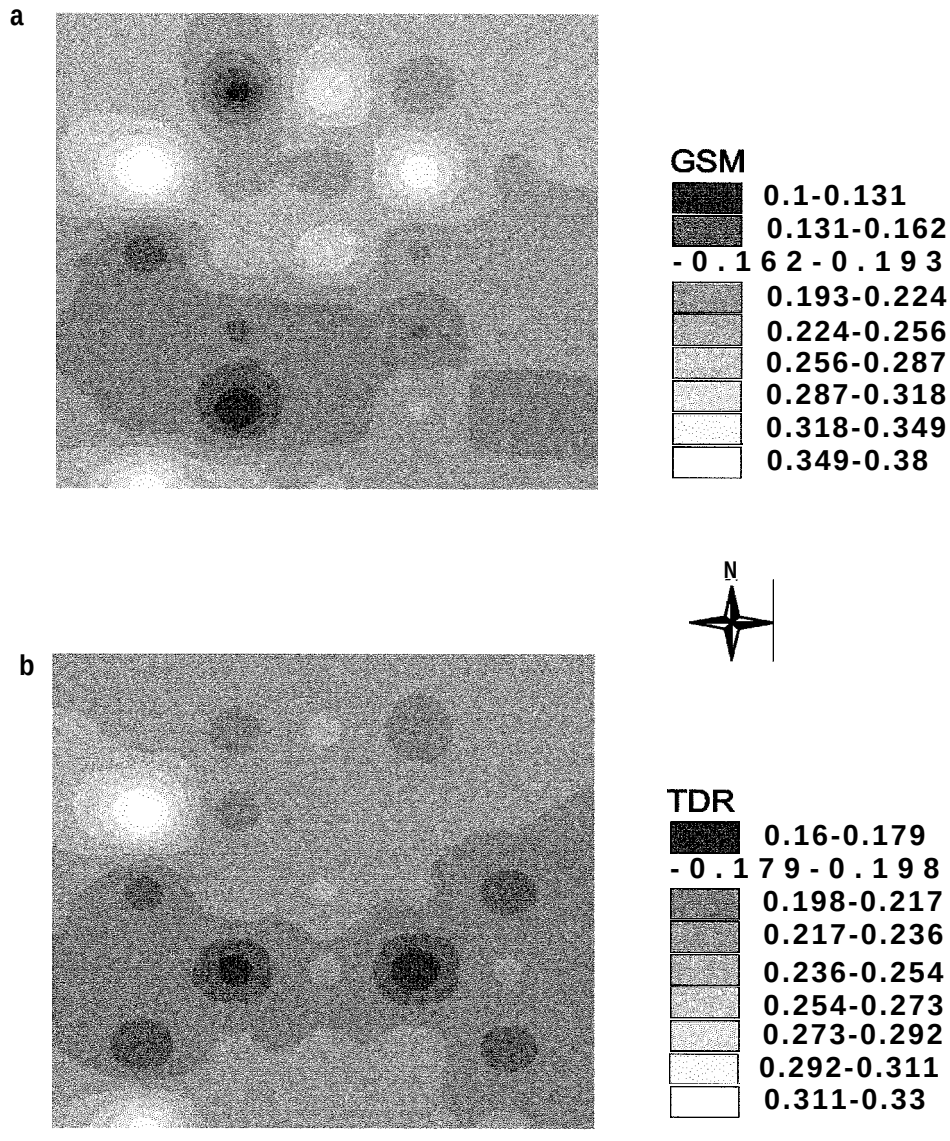


Figure 8. Comparison of the spatial distribution of soil moisture for O -6 cm depth by (a) GSM and (b) TDR measurement techniques for pasture field (sampled June 24, 1997).

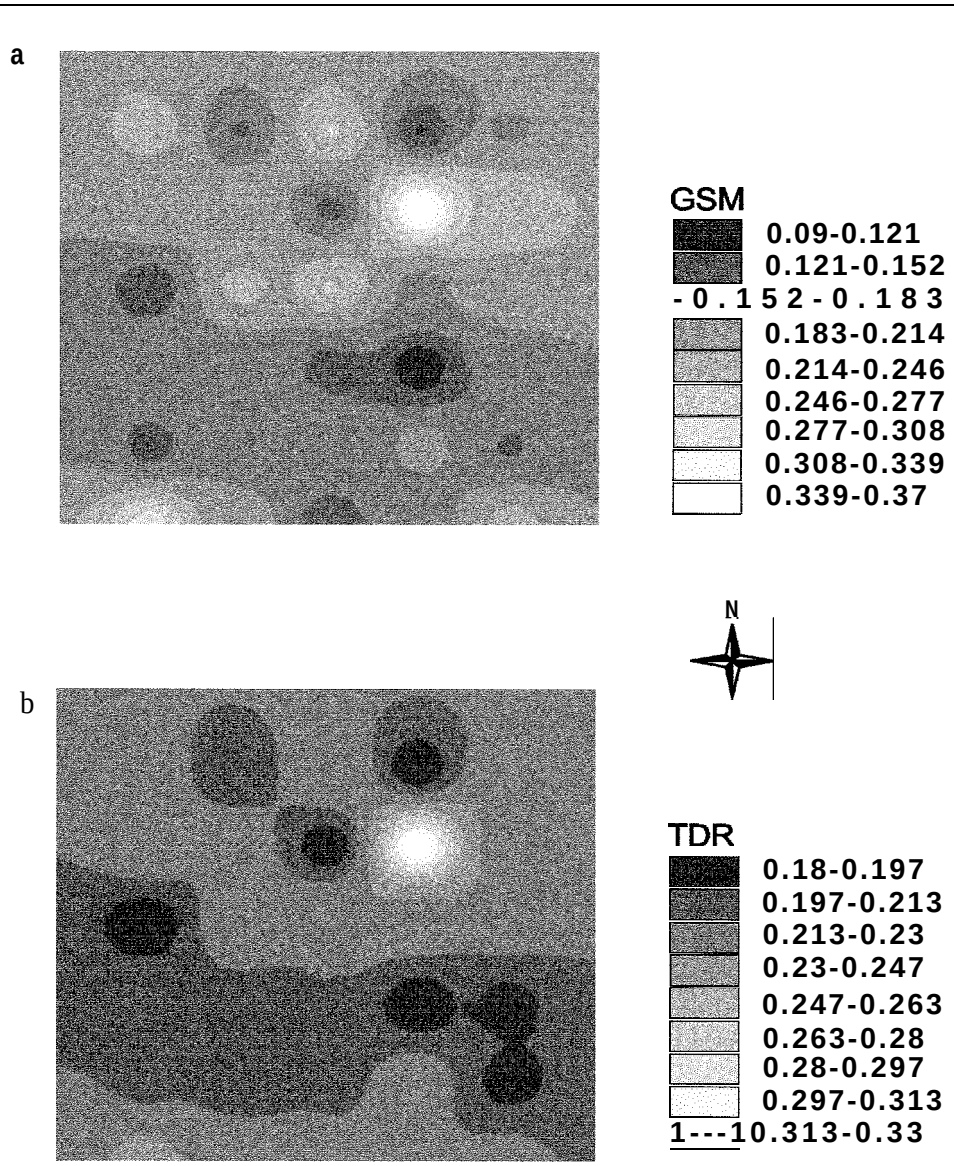


Figure 9. Comparison of the spatial distribution of soil moisture for O -6 cm depth by (a) GSM and (b) TDR measurement techniques for pasture field (sampled June 28, 1997).

Table 1. Statistical comparison of volumetric water content measured by the gravimetric soil moisture (GSM) and the Time Domain Reflectometry (TDR) technique.

Measurement Dates	**21-Jun 1997**		**24-Jun 1997**		**28-Jun 1997**	
Measurement Technique	GSM	TDR	GSM	TDR	GSM	TDR
Mean	0.151	0.174	0.235	0.237	0.207	0.223
Variance	0.0110	0.0062	0.0061	0.0054	0.0075	0.0022
Correlation	**0.86**		**0.81**		**0.91**	

The spatial and temporal distribution of soil moisture measured byTDR and GSM techniques are given in Fig. 7-9. The range of soil moisture measured by the two techniques for each date was slightly different. However, both techniques gave results that are visually comparable throughout the sampling locations. Clear spatial continuity was not observed in this field for all three dates. However, there is strong evidence that the distribution of soil moisture is quite random in this pasture field.

CONCLUSION

Both GSM and TDR techniques were used to measure soil moisture in the upper 6-cm of the soil depth. The two techniques were utilized side by side and the data obtained from these techniques were quite similar. A fourth order polynomial equation was fitted to each probe calibration curve. The GSM technique slightly under estimated the mean soil moisture content as compared to the TDR measurements. We believe that the GSM technique underestimated due to the measurement error while transporting the samples from the field to the lab. There is no clear spatial continuity of soil moisture in this pasture for the three dates that this study examined. The GSM technique can be successfully replaced with such TDR devices as long as site specific calibration is done,

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Center for Hydrology, Soil Climatology, and Remote Sensing (HSCaRS) support staff, Mr. Jhonny Boggs for their valuable time in assisting with the field and laboratory work. Contribution from the HSCaRS Center and Agricultural Expedient Station, Alabama A&M University. "This work was supported by Grant No. NCCW-0084 and NCC8-140 from the National Aeronautics and Space Administration (NASA), Washington, DC. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government

REFERENCES

Crosson, W. L., Laymen, C.A., Manu, A., and **Tsegaye,** T., 1997, Huntsville '96 microwave remote sensing of soil moisture experiment description of soil moisture profiles, *13th Conference on Hydrology, American Meteorological Society*, Long Beach, CA, Feb. 2-8.

Tsegaye, T. D., C. A. Laymen, W. L. Crosson, and T. L. Coleman. 1997. Soil Moisture measurement techniques for validation of remotely-sensed data: Evaluation and performance test of soil moisture sensors. *In Proceedings of the SPIE Conference on Remote Sensing for Agriculture, Ecosystems, and Hydrology. Vol. 3222.22-26* September 1997, London, UK. The European Optical Society (EOS) and SPIE-The International Society for Optical Engineering. Bellingham, WA. (In Press)

Correlation Analysis of Soil Moisture and Soil Temperature under Vegetated Plots

S. Steinfield*, T. Tsegaye#, and A. Fahsi#
*Penn State University
#Center for Hydrology, Soil Climatology, and Remote Sensing
Alabama A&M University

Abstract

Understanding soil properties is crucial to creating alternative techniques to better comprehend crop growth, micro-topography, and global change. Results showed that soil surface temperature, GSM and θ_v varied substantially from one location to another but minimally under different vegetation types. A Pearson correlation revealed a strong negative correlation (-O. 82 to -0.99) between soil temperature and θ_v when average positional values were used, but only a weak (-0.07 to -0.41) negative correlation when all data points were considered. Correlation between temperature and GSM was also negative. However, when using only the average positional values, these positions were positively correlated for the alfalfa and tall fescue plots and moderately-negatively correlated for the corn plot. Since data for this experiment is based on one field and one data collection. further investigation is needed to confirm the results and make conclusive statements about the correlation between surface soil temperature and GSM.

Introduction

Soil moisture and temperature are two dynamic soil physical properties that influence plant growth and crop yield. These two variables vary spatially due to changes in vegetation cover, slope and other related soil and climatic factors. Previous work investigating the spatial variability of soil moisture indicated that the presence of vegetative cover increased the natural variability of gravimetric soil moisture (GSM) (Tsegaye et al., 1997; Belisle et al., 1997).

Several factors influence the variability of soil temperature in the field (Shumway et al., 1989; Davidoff and Selim, 1988). For example, tillage may contribute to change in microtopography on subsurface soil thermal properties. Cloud cover or non-uniform radiation loads at the surface may effect not only diurnal, vertical soil temperature fluctuations, but also horizontal variation in soil temperature (Vauclin et al., 1982) and soil moisture. Moreover, vegetation cover and crop residues left on the soil surface may also result in non-uniform radiation load. Vegetation covers usually have reflective and conductive properties that differ from mineral soil with concurrent changes in the surface net radiation and soil heat-flux density. Several researchers have shown that vegetation cover can modify mineral soil temperature (Gupta et al., 1984; Bristow and Abrecht, 1989) and create soil moisture variability within a given field plot (Tsegaye et al., 1997). Few studies have addressed the positional similarities and dissimilarities between soil surface temperature, GSM, and volumetric water content for the upper 6-cm portion of the soil profile.

Therefore, the objectives of this study are threefold; first, to compare positional similarities and dissimilarities between upper level soil moisture content, gravimetric soil moisture content, and soil surface temperature; second, to compare soil moisture and soil surface temperature variability under different vegetative covers; and third, to examine the statistical correlation between soil moisture and soil surface temperature.

Procedure

The experiment was conducted at the Alabama A&M University, Winfred Thomas Agricultural Research Station (WTARS) in Hazel Green, Alabama. The soil series designation at this site is a Decatur silt. The soil is classified as a Rhodic Paludult. The three experimental plots measuring 50 x 60 meters (Figure 1) were covered with alfalfa, tall fescue, and corn, respectively. All samples were collected on July 10, 1997 between 2:30 and 5:30 PM. In each plot, soil samples for GSM determination were collected from three positions at four locations using a 6-cm soil sampler. The soil samples were collected with special scoops, designed in-house, that take approximately 2 16-cm3 volume of soil.

The traditional GSM determination technique was used to determine soil moisture content. Each collected sample was weighed and then oven dried at 105° C for twenty-four hours. Next, the oven-dry weight of each soil sample was determined and recorded. Finally, the GSM was calculated for each sample by subtracting the sample dry weight measurement from the moist weight measurement, dividing the result by that sample's dry weight and then multiplying that by one hundred to yield a percentage.

Surface soil temperature was measured using an infrared thermometer (Spectrum Technologies, Inc. 1996). An infrared thermometer's lens, pointed at an object collects and directs the energy onto an infrared detector inside

the instrument, producing a signal that the thermometer microprocessor translates and displays on the LCD. The procedure to measure temperature is done by simply pulling a trigger and reading the temperature from the LCD.

Volumetric water content is the ratio between the volume of water present and the total volume of the soil sample. It is a dimensionless parameter, expressed either as a percentage by volume or a ratio in m^3m^{-3}. Volumetric water content (θ_V) was determined by using a four rod ThetaProbe manufactured by Delta-T Devices (Delta-T Devices LTD, 1996). The thetaprobe measures θ_V by the well-established method of responding to changes in the apparent dielectric constant. These changes are converted into a direct current voltage proportional to soil moisture content over a wide working range. The thetaprobe measures soil moisture by applying a 100 MHz signal via a specially designed transmission line whose impedance is changed as the impedance of the soil changes.

The measurement of θ_V using this device is very simple - push the probe into the soil until the four rods are fully covered, connect the hand-held power supply and take readings from the analogue output. Measurement error could be greater than 0.1 m^3m^{-3} unless the installation is carried out properly. The thetaprobe must be inserted perpendicularly into the soil. Once in the soil, several factors can affect thetaprobe readings. Some of these factors include air pockets around the rods, the insertion angle, variation in soil density and composition, stones close to the rods, roots (either nearby or pierced by the rods), earthworm holes, and variability in transpiration and evaporation losses of surface vegetation.

In all, there was a sum of thirty-six samples taken from the three experimental fields. Statistical relationships between the GSM, θ_V, and soil surface temperature variables were examined with the Pearson correlation coefficient. The Pearson correlation coefficient was the statistical method utilized for this experiment because it measures the strength and direction of a linear relationship between two variables (Lethen, 1997). The calculated coefficient has a value from –1 to 1. A result of 1 indicates a very strong positive correlation, Whereas –I represents a very strong negative correlation. The Pearson correlation coefficient value was determined for each sample, and also for the average at each positional row in each of the vegetated fields.

It is important to realize that only one trial, consisting of 12 samples per plot, was carried out for this study. Conditions for laboratory work were not ideal, in that the soil was extremely dry making it difficult to insert the soil moisture probes. Temperature measurements were not taken at the exact same moment (they were taken over a two hour time period), so cloud cover at different moments may have caused fluctuations in temperature. The individual GSM measurements represent an integrated value for a 6-cm soil depth. The infrared thermometer only collected temperatures for the surface. Different results could be anticipated had we measured temperature and moisture over a similar integrated depth.

Results and Discussion

The mean temperature values at three different positions in the corn, alfalfa, and tall fescue grass plots are given in Figure 2. Temperatures for the corn and alfalfa decreased between position one (Pi) and position three (P3), but increased on the tall fescue plot. The alfalfa plot showed the least temperature variation between the three positions. Figure 2 also illustrates that the tall fescue plot has more positional temperature dissimilarities than the alfalfa plot; the mean surface temperature of the soil increased 6 'F between P_1 and P_3. The corn plot had the most temperature changes among positions and demonstrates the most dissimilarities compared to the alfalfa and tall fescue plots. From P, to P_3 the temperature dropped by 13 'F. Soil temperature changed with respect to spatial positioning and these results confirm that temperature varies in space depending on the type of vegetation cover.

Figure 3 compares the mean soil surface temperature by vegetation type. The number shown in parentheses next to its vegetation type represents the average surface temperature (°F) for the plot. This figure compares the influence of vegetation on the average surface soil temperature regardless of position of sampling location. The mean temperatures were all within 7 'F of each other, with the alfalfa plot recording the coolest temperature. In terms of percentages, the tall fescue temperature was 7% greater, and the corn was 6% warmer than the alfalfa's surface soil. The cooler alfalfa plot temperature can be attributed to its dense vegetation cover.

Figure 4 shows the mean GSM(%) values for all positions by vegetation type. The corn, alfalfa, and tall fescue plots each displayed a small amount of spatial variation. The alfalfa plot showed the most variation, but the differences in the mean were still no more than .035%. Position two in the alfalfa plot had the greatest average GSM value of the three positions, and of any other position within the three studied field plots. Note that while the tall fescue GSM values were slightly higher than those for corn, the two vegetation types had similar increases in magnitude of soil moisture content, in that each has a higher GSM at P_3 than P,. The corn plot had the same average GSM value at P_1 and P_2 but increased at P_3. This figure indicates that GSM values vary slightly with respect to spatial position.

Figure 2. Mean Temperature (°F) Values at three Positions Within Corn, Alfalfa, and Tall Fescue Plots

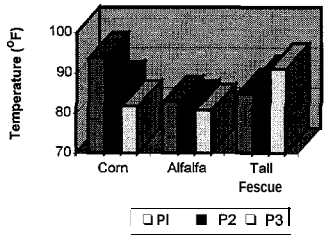


Figure 3, Mean Soil Surface Temperature (°F) for Corn, Alfalfa, and Tall Fescue Grass Plots

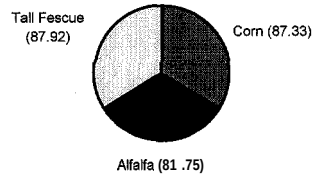


Figure 4. Mean GSM ('%) Values at Three positions Within Corn, Alfalfa, and Tall Fescue Plots

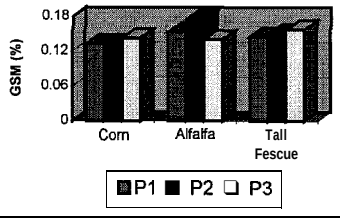


Figure 5. Mean GSM (%) for Corn, Alfalfa and Tall Fescue Grass Plots

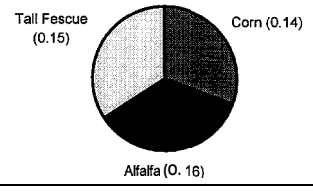


Figure 6. Mean θ_V (m^3m^{-3}) Values at Three Positions Within Corn, Alfalfa, and Tail Fescue Plots

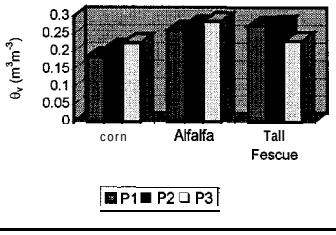


Figure 7. Mean& (m^3m^{-3}) Values for Corn, Alfalfa and Tall Fescue Grass Plots

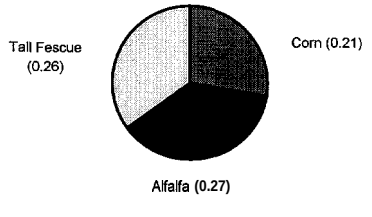


Figure 5 examines the mean GSM for each plot and compares it to the other GSM values. The corn plot had the lowest GSM. GSM values for the alfalfa and tall fescue plots were 12.5% and 6% higher, respectively. Slight differences in GSM were identified based on the vegetation type. One observation that could lead to the differences in GSM is that the corn plot had more bare soil that was exposed to direct radiation, than the alfalfa and grass plots.

Mean volumetric water content values, with respect to spatial position of each plot, are exhibited in Figure 6. The θ_V values observed for each plot varied at all sample positions. The trend for the corn and alfalfa was an increase in θ_V from P_1 to P_3. As the even stair-like appearances on the graph demonstrate, both of these vegetation types' θ_V increased at a similar rate. The tall fecue plot underwent a small increase in θ_V from P, to P, but then dropped by .047 m^3m^{-3} between P_2 and P_3. Figure 6 clearly illustrates a change in θ_V with respect to spatial location for each plot.

Figure 7 compares the overall mean θ_V for each vegetation type. Alfalfa had the highest volumetric water content among the other two plots, followed by tall fescue, then corn. The θ_V was 22°/0 greater for alfalfa and 19°/0 greater for tall fescue than it was for corn. It is not surprising that the corn plot had the lowest value, because the surface was bare and had no crop residue; it experiences the most direct sunlight - and radiation may reduce soil moisture content. Volumetric water content appears to be affected not only by the spatial position (Figure 6), but by the vegetation type as well.

The statistical correlation between soil temperature, GSM and θ_V values for the averages at PI, P_2 and P_3 is given in Table 1. When correlating temperature and moisture of a soil sample, a negative correlation is anticipated. This is due to the fact that it takes five times as much energy (sunlight) to warm water than to warm an equal volume of mineral particles. As a result, wetter soils tend to have cooler temperatures (Plaster, 1992). A negative correlation indicates that the higher the surface soil temperature, the lower the sample's moisture level, and vice versa. The θ_V results fit this trend. The corn and alfalfa plots have the same Pearson value, and came very close to a perfect negative correlation measurement of -1.0. Although the relationship was not as strong as that of the other two vegetation types, the tall fescue temperature versus moisture measurement also had a negative correlation.

Table 1. Pearson correlation between GSM, θ_V, and surface soil temperature for corn, Alfalfa and tall fescue grass plots.

Vegetation type	Corn	Alfalfa	Tall Fescue
Soil property		GSM	
Temperature	-0.8	0.36	0.99
		θ_V	
Temperature	-0.99	-0.99	-0.82

The GSM values in Table 1 show a negative correlation for corn, yet positive correlations for alfalfa and tall fescue grass. The .99 GSM value for tall fescue indicates that the greater temperature, the higher the moisture level. Not only are the alfalfa and especially the tall fescue values unexpected, but based on basic principles of physics, they are also illogical. Since data in this table are based on averages, some important actual sample values with large deviations from the mean do have influence over the Pearson correlation value. Bright sunlight can give high temperature readings with high moisture content. The same is true for low moisture content - low temperature readings due to cloud cover. This may be the reason for the positive correlation observed for the alfalfa and tall fescue GSM data.

Table 2 also shows Pearson Correlation coefficients used between soil temperature, GSM and θ_V values, but this data is calculated using all the measured data without regard to spatial location as in Table 1. Each GSM versus temperature value in this table demonstrates a moderately negative correlation. These results were anticipated and are very dissimilar from those in Table 1. Table 2 shows that the average GSM values do not fit the trend of the overall data. The θ_V versus temperature values were also very different from those in the previous table. When viewed from this perspective, this data had a much weaker negative correlation than when average values were used.

Table 2. Pearson correlation between GSM, θv, and surface soil temperature for corn, alfalfa, and tall fescue grass plots.

Vegetation type	Corn	Alfalfa	Tall Fescue
Soil property		GSM	
Temperature	-0.56	-0.51	-0.42
		θv	
Temperature	-0.41	-0.07	-0.19

A comparison of the two tables indicates that it may be more advantageous to use different approaches when trying to correlate GSM and θ_v with temperature. In terms of applications of the results from this study, a farmer would want to use the average positional values to determine a field's θ_v but all measured data to calculate the GSM of a field. The moderate correlations in Table 2 indicate that other factors, in addition to temperature, influence the GSM and θ_v levels of the soil (i.e., soil structure, bulk density, topography, vegetation density and root structure). Based on some of the results in Tables 1 and 2, and the fact that this experiment was not carried out under ideal circumstances, this experiment should performed again using multiple trials, and focusing particularly on the calculation of the GSM content.

Conclusions

The objectives of this research were to compare positional similarities and dissimilarities between soil moisture, gravimetric soil moisture content (GSM), and soil surface temperature; to compare soil moisture and soil surface temperature variability under different vegetative cover; and to examine the statistical correlation between soil moisture and soil surface temperature.

The results of this experiment lead to the conclusion that surface soil temperature does vary spatially. Positional temperature dissimilarities were identified within all three field plots. GSM measurements demonstrated moderate dissimilarities between positions. θ_v appears to vary spatially. Temperature, GSM, and θ_v varied to a small degree between the alfalfa, tall fescue and corn plots. Measurement dissimilarities between plot types resulted from different amounts of vegetation density in the field.

A strong negative correlation was found between soil surface temperature and volumetric water content for each vegetation type when determined using the spatial means. Pearson correlation coefficients for temperature versus GMC were positively correlated for alfalfa and tall fescue grass. A negative correlation of -0.8 was calculated for the corn plot. The GSM and θ_v versus temperature correlation values were comparable for the corn plot but were dissimilar from each other for alfalfa and tall fescue plots. When the Pearson correlation was repeated using all data values instead of means, moderate negative correlations were revealed. The contrast between coefficients indicates that position means should be used when calculating volumetric water content, but all data values should be used to determine GSM. The field experiment consisted of only one field trial. Repeated trials are recommended to better verify these initial results. Additional research in this area will lead to a better understanding of crop growth, microtopographical changes, and alternative techniques and instrumentation to collect surface measurements.

Acknowledgements

The author sincerely thanks Dr. Teferi Tsegaye who gave support and guidance throughout the entire course of this project. Many thanks to Mr. Harold Anthony for his computer assistance. Gratitude goes to Dr. Ahmed Fahsi of Alabama A&M University and Doug Miller of Penn State University for their aid with revisions. Also, a special thanks to the HSCaRS program at Alabama A&M University for an incredible opportunity to learn from some of the best.

References

Belisle, W., A. Manu, T. Tsegaye, Z. Senwo, A. Fahsi, G. Robertson, J. Boggs, W. Tadesse, P. O'Neill, T. Jackson, W. Crosson, and C. Laymen. 1997. Comparison of soil moisture determination under different crop coverages using remote sensing and conventional methods. *Preprint Vol. Of Conf. On Hydrology* (Feb. 2-7; Long Beach, CA), Amer. Meteor. Sot., pp. 362-365.

Bristow, K.L., and D.G. Abrecht. 1989. The physical environment of two semi-arid tropical soils with partial surface mulch cover. Aust. J. Soil Res. 27:577-588.

Davidoff, B., and H.M. Selim. 1988. Correlation between spatially variable soil moisture content and soil temperature. Soil Sci. 145:1-10.

DELTA-T DEVICES LTD. 1996. *thetaprobe Soil* Moisture Sensor User Manual. Cambridge CB5 OEJ, England. 18 pp.

Lethen, J. 1996. The Pearson Correlation Coefficient. www.stat.tamu.edu/stat30x/node39.html

Gupta, S. C., W.E. Larson, and R.R. Allmaras. 1984. Predicting soil temperature and soil heat flux under different tillage-surface residue conditions. Soil Sci. Sot. Am. J. 48:223-232.

Plaster, E. *Soil Science and Management. New* York: Delmar Publishers. 1992.

Shumway, R. H., J.W. Biggar, F. Morkoc, M. Bazza, and D.R. Nielsen. 1989. Time frequency-domain analysis of field observations. Soil Sci. 147:286-298.

Spectrum Technologies, Inc. 1996. Infrared Thermometer Operation Manual. Plainfield, IL. 17 pp.

Tsegaye, T., T. Coleman, A. Manu, Z. Senwo, A. Fahsi, W. Belisle, W. Tadesse, G. Robertson, Boggs, F. Archer, J. Surrency, L. Birgan, C. Laymen, W. Crosson, and J. Miller, 1997. Spatial and temporal variability of soil moisture with and without vegetation and its impact on remote sensing instruments. *Preprint Vol. Of Conf. On Hydrology* (Feb. 2-7; Long Beach, CA), Amer. Meteor. Sot., pp. 349-351.

Vauclin, M., S.R. Vieira, R. Bernard, and J.L. Hatfield. 1982. Spatial variability of surface temperature along two transects of a bare soil. Water Resour. Res. 18:1677-1686.

A Spatial Classifier for Multispectral Data Using Contextual Information

Chih-Cheng Hung*!, Ahmed Fahsi! and Tommy Coleman!

*Department of Mathematics and Computer Science,
! Center for Hydrology, Soil Climatology, and Remote Sensing
Alabama A&M University, Normal, AL 35762

Abstract

Connectivity describes the spatial relationship among pixels. A spatial classifier which employs the sigma probability concept of the Gaussian distribution and a type of contextual information, connectivity of the pixels, is studied in this paper. This spatial classifier attempts to replicate the kind of spatial synthesis done by the human analyst during visual interpretation or to capture the spatial relationships inherent in an aerial photograph. Several classification results of the Landsat TM data using this classifier with different window sizes for capturing the contextual information are illustrated and compared.

1. Introduction

A typical image classification process consists of two stages: 1) the training stage; a training method is used to determine the optimal number of spectral classes, and these labeled classes (given a set of classes after the training process) are then used for classification in which the unknown pixel should be assigned to one of the labeled classes in the classification stage. 2) image classification; this is done by either a pixel-based or a region-based approach. In the latter case, the image has to be segmented into homogeneous regions, and a set of meaningful features has to be defined. Once defined, image regions (blocks) can be categorized using pattern recognition techniques (Haralick et al., 1973). ECHO (Kettig and Landgrebe, 1976) is a good example of an image classification algorithm that uses the region growing technique to segment the image and classify the regions by applying the maximum likelihood classification algorithm. However, image segmentation has been proven to bean elusive goal (Gurari and Wechsler, 1982). The approach used in ECHO is not suitable for the classification of multispectral images such as LANDSAT or TM data due to the difficulty in segmenting several images efficiently.

In a pixel-based classification, three popular classification techniques have been widely used in remotely sensed data namely, the minimum-distance function, the parallelepipeds function, and the maximum-likelihood function. All of these classifiers use spectral information to classify each pixel in the image. One of the main drawbacks of these methods is that each pixel is treated independently without consideration for its neighbors (Niblack, 1986). In most natural scenes, the objects that have the same attributes tend to cluster. Hence, groups of like pixels are expected to occur together. Spectral, textural, and contextual information are three fundamental pattern elements used in human interpretation of color photographs (Haralick et al., 1973). They are also the distinguishing elements for discriminating objects in image classification. By using these three types of information, more features about an object can be extracted. This should result in a better classification. Several different combinations of these three types of information have been successfully used in image classification algorithms. Different approaches which use contextual information to improve the classification accuracy have been proposed in the literature (Zhang et al., 1988). Most of these approaches take as input the initial classification results which are usually the output of applying the per-pixel classifiers on the original image. The disadvantage of this type of algorithm is that the final classification results depend on the initial per-pixel classification result. In other words, this type of algorithm attempts to reclassify those misclassified pixels based on contextual information. However, one may not be able to correct some misclassified pixels even with contextual information.

A spatial image classifier which incorporates contextual information to classify each pixel in the raw images was proposed by Hung (1995). The procedure incorporates contextual information by using the sigma probability of the Gaussian distribution and connectivity information to classify each pixel into one of several classes. The spatial classifier strives to capture the spatial relationships encoded in the aerial photograph. However, the algorithm proposed did not consider the correlation among the pixels in N-dimensional feature space (a pixel can be defined as a point in n-dimensional feature (spectral) space). In this paper, the correlation among those pixels within the sigma range inside the window was studied. The importance of connectivity and its role in the application of contextual information in multispectral image classification was investigated. Experiments were also made on the impact of the window size in capturing the contextual information.

The organization of this paper is as follows. A review of the spatial classifier is briefly described in the next section. Section 3 discusses connectivity in contextual information. The experimental results are provided in section 4. The conclusion then follows.

2. Review of the Spatial Classifier

The spatial classifier uses the sigma probability of the Gaussian distribution and the connectivity property. The basic idea is to label the pixel being classified by considering those neighboring pixels that have intensity within an adaptive sigma range of the pixel under consideration. Connectivity describes the spatial relationship between a pixel and the other contiguous pixels of the same class; therefore, this ensures that all the pixels within the sigma range are not randomly distributed (noise) pixels. Essentially, this method is looking for majority-support instead of looking at an individual pixel for classification. The proposed algorithm is described in the following section.

Algorithm

In classifying a pixel, any pixel outside an adaptive sigma range most likely comes from a different class and, therefore, should not be included in the class which is being considered. The procedure is described as follows. We assume that a training set has been generated in the training stage and that the number of classes is K in the training set.

The following steps are applied to each pixel to be classified.

Step 1: Establish an intensity interval (X - s, X + s), where $s = ks$, k is a constant and s is the standard deviation of the pixels in the window of size $(2n+1)\times(2n+1)$, n is a positive integer, centered at pixel X. k usually ranges from 0.5 to 2.

Step 2: Mark all pixels which lie within the intensity interval in the specified window.

Step 3: Check the connectivity among the marked pixels using a neighborhood type of 4 or 8. (If the window size 3x3 is used, this step can be ignored).

Step 4: Compute an average pixel value A as follows. If the number of marked pixels is greater than a prespecified value C and they are connected, then A is the average of the marked pixels. Otherwise, A is the average of the immediate surrounding neighbors of X. C depends on the size of window. For a 3x3 window, C can equal 2, and for a 5x5 window, C can equal 3.

Step 5: Label the pixel X with class i if $\|A - Z(j)\| > \|A - Z(i)\|$ for all $i,j = 1,2, \ldots, K, i \neq j$, where K is the number of classes defined in the training stage and Z(i) is the mean of class i. Note that if multibands are used in classification, A and Z(i) are vectors.

Note that step 3 is critical in determining the connectivity for classification. Several cases for connectivity were tested in our experiments.

3. **Connectivity in Contextual Information**

Spatial information in digital images can be quantitatively expressed in many different ways. Gurney et al. (1983) categorizes spatial information into two types: textural and contextual. Texture has received considerable attention. Texture refers to a description of the spatial variability of tones within part of an image. Various methods to extract statistical textural features have been developed (Wu and Chen, 1992). Context is a measure of the relationships between the pixels in a neighborhood. Four basic forms of relationship in contextual information were outlined in Gurney and Townshend, 1983: namely, distance, direction, connectivity, and containment. To develop a general robust image classification algorithm, one should not rely on knowledge-based information heavily. (Distance, direction, and containment are types of knowledge-based information). In other words, this information will vary for different classes of images. Therefore, only the property of connectivity is considered in developing an algorithm. Connectivity describes the spatial relationship between one pixel and other contiguous pixels of the same class. If a pixel is connected to a very large object, or an object of a particular shape, it could be assigned to a particular class (Gurney and Townshend, 1983). This also ensures that all the pixels within the sigma range are not randomly distributed (noise) pixels. Essentially, this method uses majority-support instead of looking at an individual pixel for classification. Connectivity can be applied to any types of images based on the knowledge derived from the spectral information embedded in images.

Connectivity between pixels is of importance in establishing boundaries of objects and components of regions in an image (Gonzalez and Woods, 1992). To determine whether two pixels are connected, a criterion of similarity must be established. Before the criterion is made, the neighbors of a pixel must be determined. A pixel has two horizontal, two vertical, and four diagonal neighbors in a digital image. The 4-neighbors (horizontal and vertical) and 8-neighbors (plus diagonal) are commonly used in digital image processing, For a gray level image, if a region is homogeneous, it is easy to establish the connectivity based on 4- or 8-neighbors concept. However, most regions in an image are not homogeneous, particularly in remotely sensed images. Thereby, the Gaussian distribution is used to identify those neighboring pixels that have intensity within an adaptive sigma range of the pixel under consideration. The 4-connectivity and 8-connectivity can then be used in determining how many neighboring pixels surround a central pixel.

In the analysis of 4-connectivity and 8-connectivity, we found that 4-connectivity may not be able to catch the connectivity in some cases. For example, if the pixels within the sigma range are shown in Figure 1, the 4-connectivity method would not be able to integrate the contextual information in the classification process.

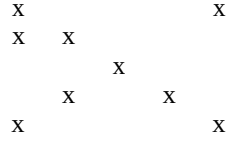


Figure 1: An example of the pixels distribution (the value x of all pixels are in the sigma range) in which the 4-connectivity method failed to generate a group of pixels coming from the same population in a window size of 5x5.

4. **Experimental Results**

The spatial classifier was applied to multispectral remotely sensed images to demonstrate its effectiveness. The experiment used a 429 row by 202 column three bands, For the comparison, both the spatial classifier and the minimum distance per-pixel classifier were implemented to test on the same set of images. A hybrid multispectral classification system consists of unsupervised training and supervised classification. K-means clustering was used to define the appropriate spectral classes in unsupervised training. Several different window sizes of 3x3, 5x5, and 7x7 were used in integrating contextual

information in the classification process. Both 4-connectivity and S-connectivity were used in the experiments. An original TM image (band 4) and its classification results are shown in Figures 2 and 3,

By visually comparing these classified results, in general the spatial classifier generates more connected and smoother classification results. Note that we did not apply any pre-processing technique to the original multispectral images to remove any noisy signals. K was set to 2.0 for the sigma range. One of the advantages for the spatial classifier is the noise in the image can be suppressed during the classification process. We also observed that a small patch of a certain texture can be clearly seen in the classified image by using the proposed classifier. This indicated that by varying the sigma range, a clean, smooth classified image can be obtained. This can release the burden of post-classification processing where the objective is to smooth the classified image. It may not be necessary to run the mode or LIFE filter (Intergraph, 1992) to clean up the classified image if this spatial classifier is used in the classification stage. This smoothing will make the classified image appropriate for vectorization in Geographic Information System (GIS) applications. Hence, remote sensing information can be interrelated with other types of spatially distributed (cartographic) information. The contextual information is highly contingent upon the size of the window. If the window size is much larger than the region, the whole region may be wiped out completely. This also depends on the constant K. The AND function in general has a poor performance due to its restriction on the coverage of the related pixels in 3 bands. The OR function always produces smooth results. Due to the variability of spectral responses, the OR function should be used to determine the connectivity. From our experiments, the window size of 3x3 or 5x5 should be used for the remotely sensed images. If the window is too large, it may mix several different population classes into one class. This leads to misclassification.

5. Conclusion

A simple image classification algorithm based on contextual information was studied for the multispectral TM data classification. The algorithm generates a more smoothly classified image than the minimum distance classifier in our experiments. It should be noted that by controlling the sigma interval and window size, the spatial classifier may generate a different classification result. If one knows the texture of the image for classification, an appropriate sigma range and window size can be selected for optimal classification. From our experiments, we found that a smaller k should be used if the size of the neighboring window is large while a larger k should be used if the size of the neighboring window is small. The OR function also should be used to determine the correlation from multiple bands.

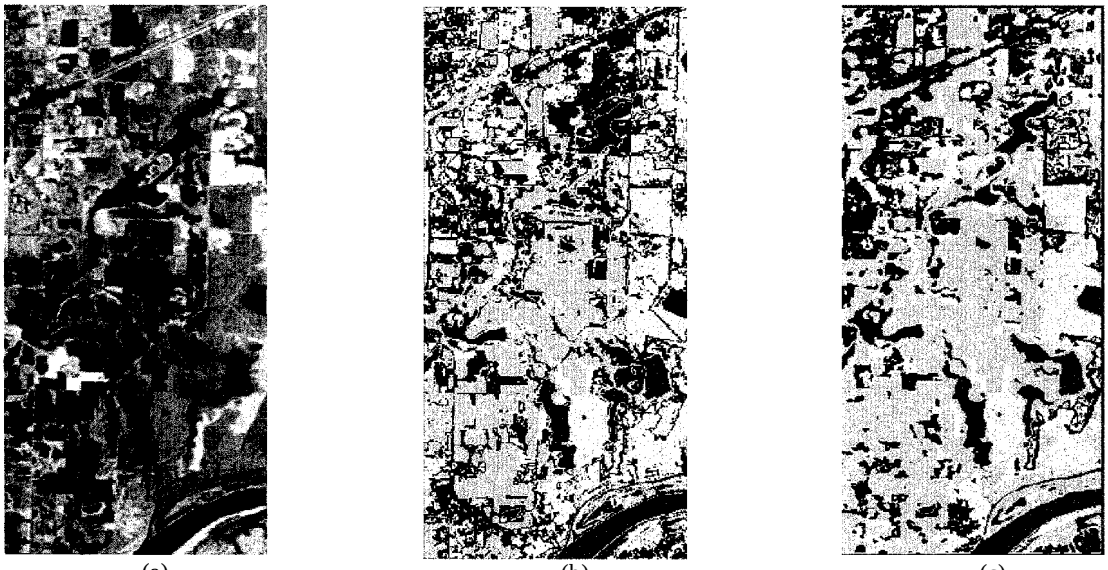
Like most local neighborhood operations in image processing, this spatial classifier is suitable for real-time or near real-time implementation as all of its operations are localized. Several pixels can be processed in parallel without waiting for their neighboring pixels to be processed. The experimental results appear encouraging compared to the results of the minimum distance per-pixel based classifier. This classifier can be more noise immune during the classification process due to the use of the connectivity property and local information. If the noise level is not serious, this spatial classifier not only eliminates the need of noise removing in pre-processing but also reduces the effort needed for smoothing in post-processing. In other words, the smoothing process has been integrated into the spatial classification algorithm. Classification accuracy assessment is under investigation.

Acknowledgment

This work was supported by Grant No. SUB90-206-A12 from the Alabama Space Grant Consortium and NCC W-0084 from the National Aeronautics and Space Administration (NASA), Washington, DC. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The authors wish to thank Dr. John Adams for proofreading the manuscript.

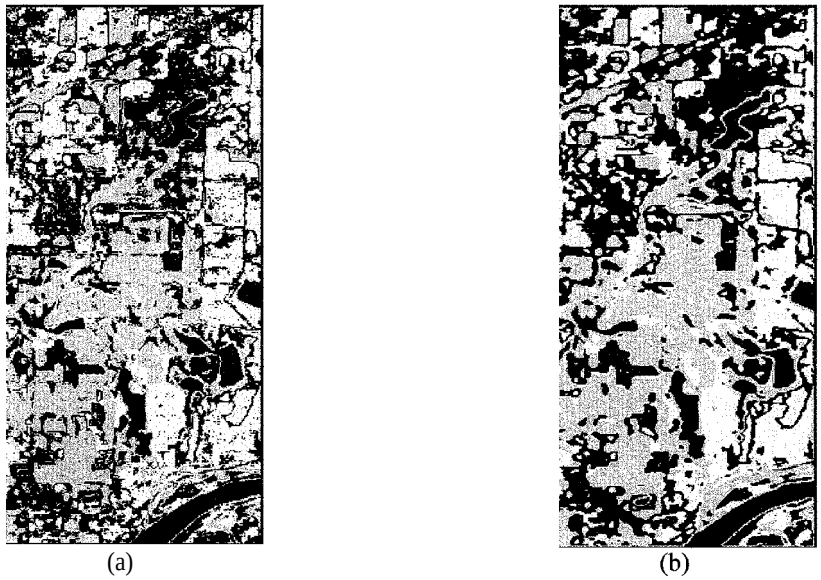
References

Gonzalez, R. C. and R. E. Woods, Digital Image Processing, Addision-Wesley, 1992.



(a) (b) (c)

Figure 2: (a) An original TM image (band 4), (b) the classified image using the minimum distance classifier, and (c) the classified results using 5x5 window with 8-connectivity without consideration of the correlation among the bands.



(a) (b)

Figure 3: (a) and (b) the classified results with the correlation considered using AND logic fiction among bands, with the correlation considered using OR logic function among bands, respectively. A set of bands 4, 5, and 1, the window size 5x5 and 8-connectivity were used in the above classified images.

Gurari, E. M., and Wechsler, H. (1982), "On the Difficulties Involved in the Segmentation of Pictures", **IEEE Trans. Pattern Analysis and Machine Intelligence,** Vol. PAMI-4, No. 3,304-306.

Gurney, C. M. and J. R. G., Townshend (1983),"The Use of Contextual Information in the Classification of Remotely Sensed Data", **Photogrammetric Engineering and Remote Sensing,** Vol. 49, No. 1, 55-64.

Haralick, R. M. and I. Dinstein (1975), "A Spatial Clustering Procedure for Image Classification", **IEEE Transactions on Circuits and Systems,** Vol. CAS-22, **No. 5,** May, pp.440-450.

Haralick, R.M., K. Shanmugam and I. Dinstein (1973),"Textural Features for Image Classification", **IEEE Transactions on systems, man, and cybernetics, vol.** SMC-3, **No. 6,610-621.**

Haralick, R. M. and Shaprio, L. G, (1985), "Image Segmentation Techniques Survey", **Computer Vision, Graphics, and Image Processing 29, 100-132.**

Hung, C. C., "A Multispctral Image Classification Algorithm Based on Contextual Information," **Proceedings of the 13th Annual International Conference - Computer Science Group, The** Association of Management, Vancouver, B.C., Canada, August, 1995.

Intergraph (1992), **Modular GIS Environment/ImageStation Imager-3,** Intergraph **Corporation, Huntsville, Alabama, USA.**

Kettig, R. L. and D.A. Landgrebe (1 976), "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects", **IEEE Trans. on Geoscience Electronics, Vol.** GE-4, **No. 1,** January, pp.19-26.

Niblack, W.(1986), **An Introduction to Digital Image Processing, Prentice/Hall Internal.**

Wu, C.-M. and Y.-C. Chen (1992), "Statistical Feature Matrix for Texture Analysis", **CVGIP: Graphical Models and Image Processing, Vol. 54, No. 5, September, 407-419.**

Zhang, Z. et. al. (1988), "A New Spatial Classification Algorithm for High Ground Resolution Images", **proceedings of IGARSS '88 symposium, Edinburgh, Scotland, Sep. 13-16.**

Effect of land-use practice on soil moisture variability for soils covered with dense forest vegetation of Puerto Rico

T. Tsegaye[1] *, T. Coleman', Z. Senwo[1], D. Shaffer[2], and X. Zou[2]

Center for Hydrology, Soil Climatology, and Remote sensing '

Alabama A&M University, Normal, AL 35862

&

University of Puerto Rico', SAN JUAN, PUERTO RICO 00936

ABSTRACT

Little is known about the landuse management effect on soil moisture and soil pH distribution on a landscape covered with dense tropical forest vegetation. This study was conducted at three locations where the history of the landuse management is different. Soil moisture was measured using a 6 -cm three-rod Time Domain Reflectometery (TDR) probe. Disturbed soil samples were taken from the top 5-cm at the up, mid, and foothill landscape position from the same spots where soil moisture was measured. The results showed that soil moisture varies with landscape position and depth at all three locations. Soil pH and moisture variability were found to be affected by the change in landuse management and landscape position. Soil moisture distribution usually expected to be relatively higher in the foothill (P3) area of these forests than the uphill (PI) position. However, our results indicated that in the Luquillo and Guanica site the surface soil moisture was significantly higher for P 1 than P3 position. These suggest that the surface and subsurface drainage in these two sites may have been poor due to the nature of soil formation and type.

INTRODUCTION

Forests have many values and uses, not the least of which are moderating our atmosphere, increasing the organic matter content of soils, and holding the soil in place, on about 30 percent of the land surface. Thus, forests are the focus of much attention in ecosystem management and sustainable development concepts.

Most of the forest acreage loss in Puerto Rico during the fifties were to agricultural activities to cultivate coffee and sugar cane and to create pasture lands for cattle (Dietz, 1986). A shift to an industrial based economy resulted in the abandonment of agricultural lands (Thomlinson et al. 1996) and an increase in forest cover(Birdsey and Weaver, 1987). Zimmerman et al. (1995) studied forest regeneration in a 60 years old abandoned pastures and coffee plantations and compared them with forested stands. Their results showed that past land use had no effect on species diversity, basal area, or stem density, but it had a strong effect on the species composition.

The environmental heterogeneity of landscape is hierarchical and is controlled by different processes at different spatial and temporal scales (Urban et al., 1987). Landscape also varies across time and space, even without human influence. Knowledge of this variability is extremely useful in determining if the current condition of a land use is sustainable, given historic pattern and process criteria. Description of historic landscape disturbance regimes (e.g., fire magnitude and frequency) and the ecosystem component patterns they maintain (e.g., vegetation composition) provide an initial template for assessing ecosystem health.

To study and manage large forest regions, or around the globe, it is necessary to aggregate information gathered at finer scales. The technical problems of aggregating and integrating small-scale processes at multiple scales seem overwhelming. A major challenge for land use effect is to discover the assembly rules for aggregation of small landscape units and the land use effects into units of larger scales.

Change in land use practices and its effects on ecology represents an interest in large scale phenomena among a growing number of scientists, land managers, and decision makers. Information concerning the soil moisture distribution on a landscape is essential for landscape characterization in the tropical rainforest. This information, in turn, is commonly used in land management to describe spatial and temporal patterns of species and communities, patterns of bio-environments and physical environments at different spatial scales, spatial and temporal patterns in ecosystem composition, structure, and fiction. Very limited work has been done to investigate the spatial distribution of soil moisture in the tropical rain forest of Puerto Rico. The objective of this research is to examine the kind use management and landscape position effects on the spatial distribution of soil moisture and pH on the top 6 cm soil depth.

DESCRIPTION OF THE STUDY SITE

Luquillo

The Sabana site (18 18'N, 65 50'W) is in the northeastern Luquillo Mountains of Puerto Rico. The native forest between 200 and 600 m above sea level was dominated by tabonuco forest (Ewel and Whitmore, 1973). Annual precipitation is approximately 2500 mm with mean monthly temperatures varying from 22 to 26°C. We selected three sites adjacent to each other on the same slope using air photos from 1936, 1951, 1964, 1971, 1977, 1983, 1988, 1991, and 1993 (Zou and Gonzalez, 1997). These sites have similar soil type and microclimate conditions, but they differ in land use history. One site was deforested and converted to pasture prior to 1936. It was abandoned in 1971 and secondary forest vegetation regenerated the area. We refer to this site as the "secondary forest" with an age range from 26 to 33 years (Aide et al., 1995). The second site was also deforested and converted to pasture prior to 1936. However, this site maintained an "active pasture" until 6 months before our field sampling. We refer to this site as "active pasture". The third site has been forested since 1936. We refer to this site as the "mature forest".

Guanica

We visited three sites in the Guanica dry forest: the Fuerte road site, Guitierrez, and the La Entrada site. The Fuerte road site had been subjected to patch-cutting (or selective cutting) about 60 years ago. Some of this wood was buried into pits for charcoal production. The presence of residual charcoal was evident at the time of our soil sampling. Both the Gutierrez and La Entrada sites were completely cleared for agriculture more than 40 years ago. Between 30 and 40 years ago, this agricultural use was stopped and secondary forest was allowed to regenerate naturally without further intervention.

In general, all three sites receive an average of 750-mm rainfall annually, with a pronounced dry season from December through July. This extended dry season in some years is interrupted by Heavy rainfall during the month of May. The average annual temperature is 25 degrees centigrade. In the Holdridge Life Zone system the Guanica dry forest is a subtropical dry forest.

Ciales

This site was located in the karst region of Puerto Rico. The limestone karst region extends along the north coast of Puerto Rico from the town of Aguada to Loiza. The area is underlaid by six limestone formations that range in age from Oligocene to Miocene (Giusty, 1978). The study sites occur in Lares Limestone formation, which is made from many marine fossils and calcium carbonate. It is the second oldest limestone formation in Puerto Rico (Monroe, 1976 Giusty, 1978). The average annual rainfall in the town of Ciales is 2074 mm and ranged from 1563 to 2785 mm during the period of 1970 to 1994 (NOAA).

At this site, vegetation composition varies depending on the topographic position and orientation. The hilltops have a dry evergreen forest and the valleys and sinkholes have evergreen seasonal or possibly lower montane rain forest (Murphy, 1916; Chinea, 1980).

MATERIALS AND METHODS

Soil moisture measurement and soil sampling were taken at the above mentioned three sites (i.e., Luquillo, Guanica, and Ciales). The land use management at the Luquillo site are described as: secondary forest (L 1), active pasture (L2), and mature forest (L3); Guanica site patch cutting site (L1), cleared for agriculture use and secondary forest (L2), secondary forest (L3); and Ciales site Dry vegetation (L 1), evergreen seasonal rain forest (L2), and lower montane rain forest (L3). The landscape positions were divided within each land use as up (Pi), middle (P2), and foothill (P3). Volumetric soil moisture content were determined by inserting a TRIME-FM three 5-cm rods Time Domain Reflectometry (TDR) probe into the top 5-cm soil depth. Disturbed soil samples were taken at the time of soil moisture measurements from the top 5-cm using a specially designed 125 cm^3 scoop sampler and transported to the laboratory to measure soil pH. Soil pH was determined by using a 1:2 soil solution ratio. A pH meter was used to determine the soil pH (Klute, 1986).

RESULTS AND DISCUSSION

The effect of land use management on soil pH and soil moisture distribution in the top 5 cm of the soil surface for all three sites are shown in Figure la-f. Land use practices changed dramatically between 1936 and 1998 in most part of Puerto Rico. One of the most marked changes was the dramatic decrease in sugar cane production. This was matched by a dramatic increase to low and medium density forest, and pasture land. The land use management and rainfall distribution for all three sites were quite different. For example, the land use in Luquillo has changed rapidly between 1936 and 1988 (Thomlinson et al., 1996). Prior to 1988, sugar cane and pasture were dominant in our study area. By 1988, most of the land was covered by dense forests and some with pasture. There was significant difference in soil pH and VWC within each three sites due to land use management. In general, the Luquillo site had lower pH than the Guanica and Ciales site. The type of soil formation and the distribution of rainfall in this site may have caused the soil to be acidic. The soil pH remains above 6 in Guanica and Ciales sites. The VWC at the top 5-cm of the soil depth also indicated significant difference among land use practices. The Luquillo site had also the highest amount of soil moisture compared to the other two sites. The mature forest exhibited high VWC as compared to the secondary forest due to high residue cover on the surface. The Guanica site soils were exposed to direct sunlight and, as a result, the VWC at this site was the lowest. Comparing the land use management in these three sites, the dense forest vegetation cover had a pronounced effect on decreasing the soil pH and increasing the moisture content at the surface of the soil profile.

Landscape position effect on soil pH and soil moisture distribution was also examined within each site. The results are shown in Figure 2 a-f. There is significant difference among means among positions at all three sites. However, there was no consistent pattern in the distribution of soil pH with landscape position. In the Luquillo and Ciales site, pH decreases as the landscape position changes from the up hill position (Pi) to the foothill position (P3). In the Guanica site, the mid-hill position exhibited high pH compared to the up-hill or foothill positions. The distribution of the soil moisture was quite different within each position for each site. The landscape position can have significant effect on the distribution of soil moisture throughout the soil profile. Soil moisture distribution is usually expected to be higher in the foothill position of the landscape due to surface and subsurface drainage. In both the Luquillo and Guanica sites, the up-hill positions (P 1) indicated a higher level of soil moisture regardless of the land use management. This suggests that the surface and sub-surface drainage in these two sites may have been poor due to the nature of soil formation and soil type.

CONCLUSION

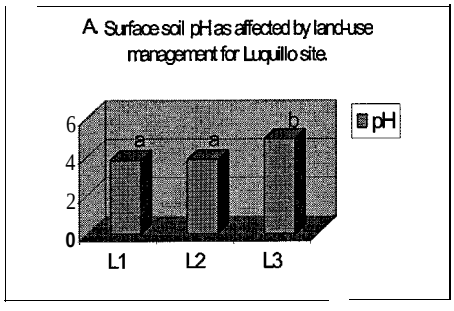
In this study the soil pH was low where the surface soil moisture was quite high at the Luquillo site compared to the Guanica and Ciales sites. In addition to the nature of soil formation and soil type, landscape position and land use management may affect soil pH and the movement and redistribution of soil water in the profile. Land use management and soil formation can have pronounced effect on soil moisture as well as soil pH spatial pattern. Land use management and landscape positions are two important factors to consider in ecosystem management decision making.

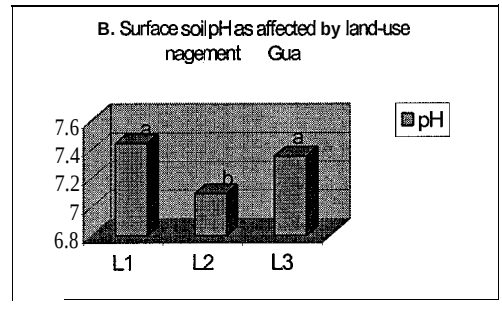
ACKNOWLEDGMENTS

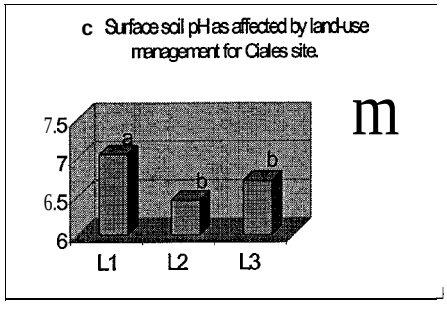
The authors gratefully acknowledge the Center for Hydrology, Soil Climatology, and Remote Sensing (HSCaRS) support staff, Mr. Jacques Surrency for their valuable time in assisting with the field and laboratory work. Contribution from the HSCaRS Center and Agricultural Experiment Station, Alabama A&M University. AThis work was supported by Grant No. NCCW-0084 and NCC8-140 from the National Aeronautics and Space Administration (NASA), Washington, DC. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government

REFERENCES

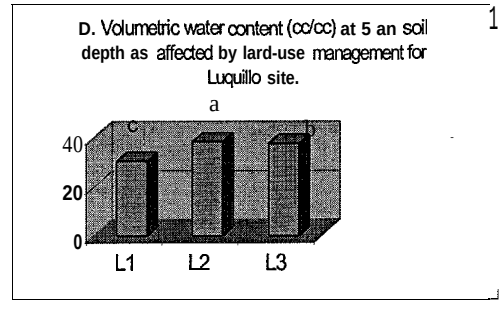
Fig. 1. LAND USE MANAGEMENT EFFECT ON SOIL pH AND SOIL MOISTURE

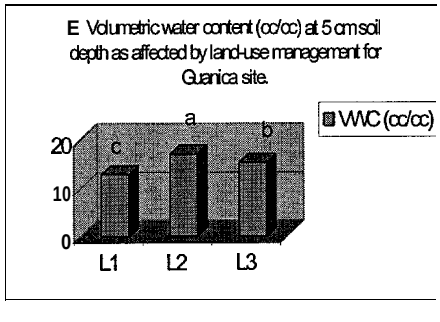






1





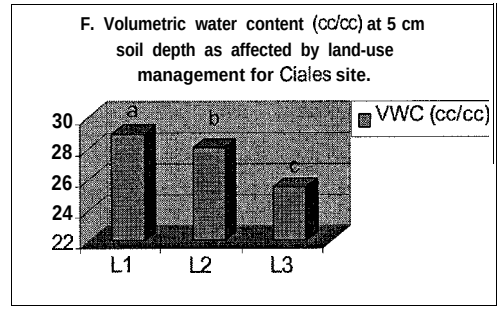
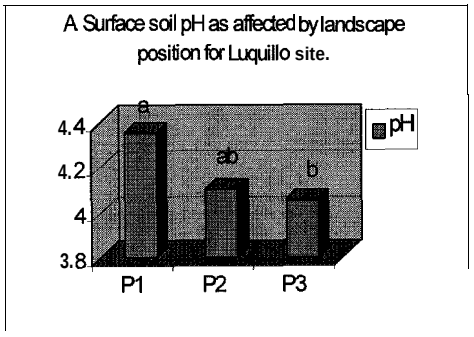
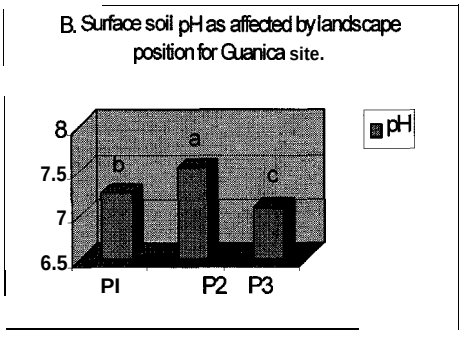
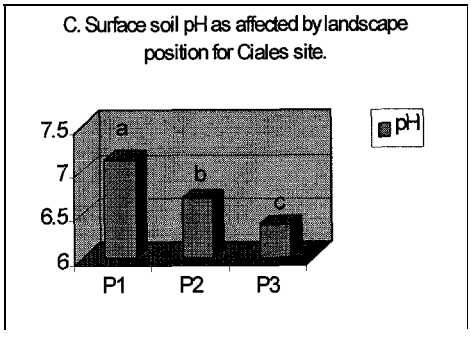
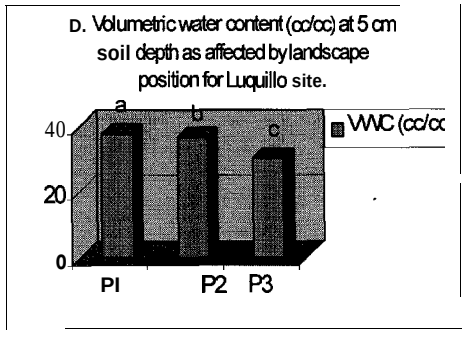


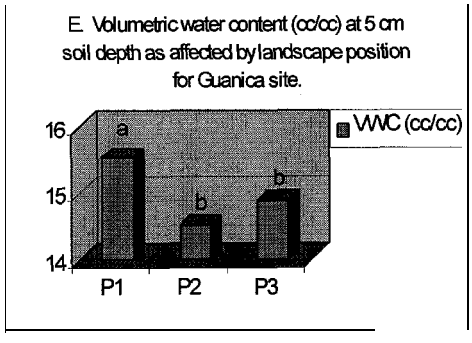
Fig 2. LANDSCAPE POSITION EFFECT ON SOIL pH AND SOIL MOISTURE

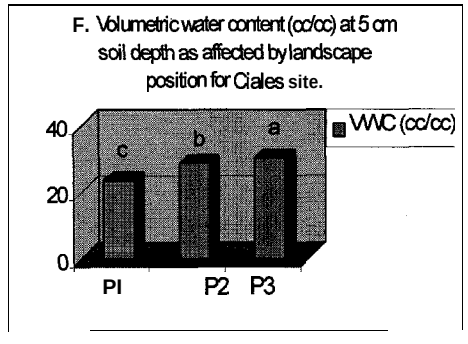












REFERENCES

Aide, T. M., Zimmerman J. K., Herrera L., Rosario M. and Serrano M. 1995. Forest recovery in abandoned tropical pastures in Puerto Rico. Forest Ecology and Management 77:77-86.

Birdsey, R. A., and Weaver, P. L. 1987. Forest area trends in Puerto Rico. Res. Note, SO-331, US Forest Service, New Orleans, LA 13 pp.

Chinea, J. D. 1980. The forest vegetation of the limestone hills of northern Puerto Rico. Cornell University, Master Thesis.

Dietz, J. L., 1986. Economic history of Puerto Rico: Institutional change and capitalist development. Princeton University Press, Princeton, NJ, 337 pp.

Ewel, J. J., and Whitmore J. L. 1973. The ecological life zones of Puerto Rico and the U.S. Virgin Islands. USDA Forest Service Research Paper, ITF- 18. Institute of Tropical Forestry, Rio Piedras, Puerto Rico.

Guisty, E. V. 1978. Hydrogeology of the karst of Puerto Rico. U.S. Geol. Surv. Prof. Paper

Klute, A. 1986. Methods of soil analysis. Part 1. Physical and mineralogical methods. 2 'd ed. Agron. Monogr. 9 ASA and SSSA, Madison, WI.

Monoroe, W. H. 1976. The karst land forms of Puerto Rico. U.S. Geol. Surv. Prof. Paper 899.

Murphy, L. S. 1916. Forest of Puerto Rico, past, present, and future, and their physical and economic environment. USDA Bull. 354, Washington, DC, 99 pp.

Thomlinson, J. R., Serrano, M. I., Lopez, T. M., Aide, T. M., and Zimmerman, J. K. 1996. Land use dynamics in post-agricultural Puerto Rican landscape (1936-1988). Biotropica, 28:525-536.

Urban, D. L. , O'Neill, R V., Shugart, H. H., Jr. 1987. Landscape ecology: a hierarchal perspective can help scientists understand spatial patterns. Bioscience. 37:119-127.

Zimmerman, J. K., Aide, T. M., Rosario, M., Serrano, M., and Herrera, L. 1995. Effects of land management and a recent hurricane on forest structure and composition in the Luquillo Experimental Forest, Puerto Rico. For. Ecol. Manage., 77:65-76.

Zou, X and Gonzalez, G. 1997. Changes in earthworm density and community structure during secondary succession in abandoned tropical pastures. Soil Biology and Biochemistry 29:627-629.

98 URC019

Surface Morphology of Undoped and Doped ZnSe Films

T. George, M. Hayes, H. Chen, K. Chattopadhyay, E. Thomas, S. Morgan and A. Burger
NASA/Fisk Center for Photonic Materials an Devices
Department of Physics
Fisk University
Nashville, TN 37208

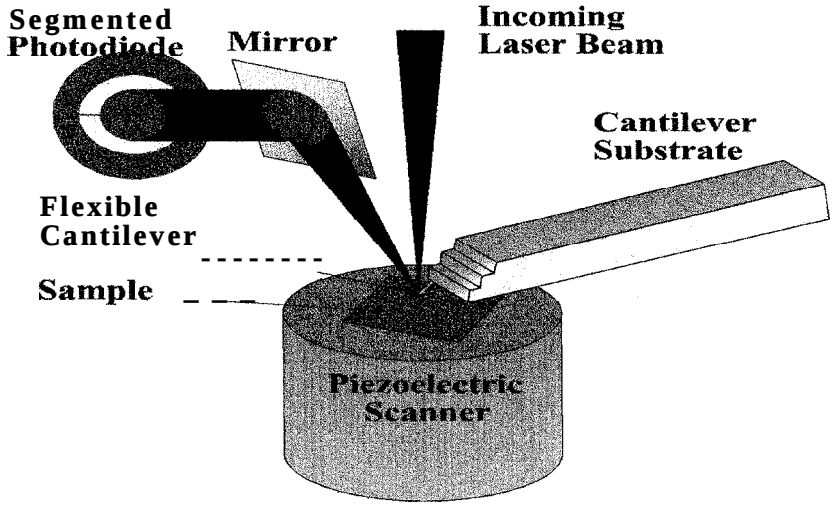
Abstract

Rare-earth doped ions in polar II-VI semiconductors have recently played an important role in the optical properties of materials and devices. In this study, undoped ZnSe and erbium doped ZnSe films were grown by radio frequency (RF) magnetron sputtering method. Atomic Force Microscopy (AFM) was used together with optical microscopy and UV-Vis spectroscopy to characterize the films. Doped samples were found to have higher surface roughness and quite different surface morphology compared to that of undoped samples. The grown films generally show a relatively smooth and uniform surface indicating that they are of overall good quality. The impact of plasma etching on ZnSe:Er film examined under AFM is also discussed.

Introduction

ZnSe with room temperature bandgap of 2.67 eV is a II-VI compound semiconductor which has gained a lot of attention during the last decade due to its promising potential for blue light emitting diodes and blue laser diodes. In the past, ZnSe films were usually grown by epitaxial techniques such as metal organic vapor phase epitaxy (MOVPE)[1], molecular beam epitaxy (MBE)[2] or gas source molecular beam epitaxy (GSMBE)[3] on GaAs substrate as well as liquid phase epitaxy[4]. Recently, rare-earth doped and transition metal doped in H-VI compounds have been realized to play an important rol in the optical properties of materials and devices[5,6]. Most reports on doped ZnSe however, have been either nitrogen doped ZnSe films by MBE and MOMBE (metal organic molecular beam epitaxy)[7,8] or Er-doped on bulk ZnSe crystal[9]. Limited works have been reported on doped and undoped ZnSe films which were prepared by sputtering technique despite the fact that this technique has been successfully applied to grow CdS[10] or Er-doped SiO_x[11] films.

In this investigation, we have grown undoped ZnSe and Er-doped ZnSe fires on Si substrates by radio frequency (RF) magnetron sputtering technique. The prepared films were examined by atomic force microscopy (AFM) (Fig.1) together with optical microscopy and UV-Vis spectroscopy. AFM was used because it is a real-time imaging technique which can produce topographical images of surfaces with atomic resolution in three dimension and the information derives from AFM images can usually be correlated to material processing conditions. Variation in surface morphology after plasma etching will also be discussed.



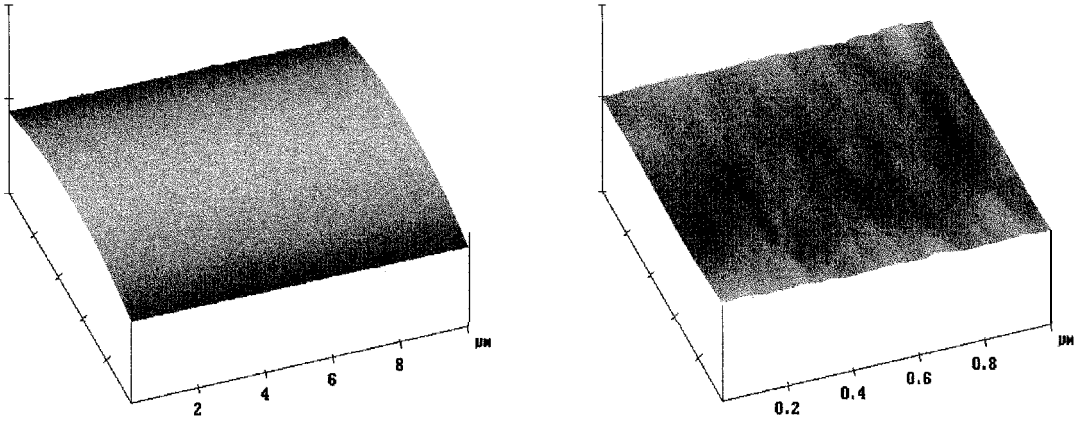
Atomic Force Microscope (AFM)
Fig. 1. Experimental arrangement of Atomic Force Microscopy.

Experimental

Doped and undoped ZnSe films were grown on (111) Si wafers 1 inch in diameter and on 1 x 2 inch micro-slides. The Si substrates were initially etched in HF for 2 minutes, rinsed in distilled water and then blown dry before being loaded into the vacuum chamber. The films were produced in a RF magnetron sputtering system (1 3.56 MHz, 300W Manitou system from Kurt J. Lesker Inc.) with base pressure of 10^{-6} Torr. The target was 2 inches away from the substrate which is heated at 350^{0}C by tungsten-halogen lamps with a temperature controller during sputtering. All runs started with presputtering of the target for 10 minutes to ensure the good quality of the films. Grains of erbium were co-sputtered with the ZnSe target onto the heated substrate to produce Er-doped ZnSe film. After the "annealed-sputtering growth", the samples were kept in vacuum chamber until they were cooled down to room temperature.

Atomic force microscopy was employed to reveal the surface morphology of the films. The instrument was a Digital Instruments Nanoscope E which was previously calibrated using mica and a gold grid. A piezoelectric scanner tube was used with an effective scan range of about 0.5 to 14μm. The cantilevers were commercial Nanoprobes made of gold coated silicon nitride having a force constant of 0.12 N/m. All images were taken in ambient air and processed by available software in the Nanoscope E computer.

UV-Vis spectrums were taken for the films produced on glass slides. The instrument is a Hitachi Model U-2000 Double-beam UV/Vis Spectrometer. In addition, a MEIJI metallurgical microscope model-E48405-20 was also employed to microscopically examine the films.



(a) **(b)**

Fig. **2.** Ambient AFM images of ZnSe films sputtered on Si wafers: (a) 10 x 10 μm, **(b)** 1 x 1 μm (Z = 10 nm).

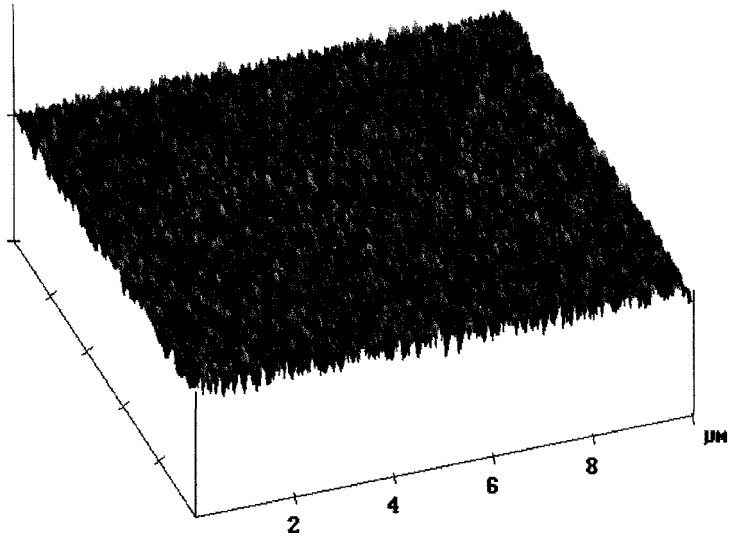


Fig. **3.** Ambient AFM image (10 x 10 pm) of Er-doped ZnSe film sputtered on Si wafer (90W, 350^{0}C, 170 militorr, 2 hours) (Z= 200 inn).

Results and discussion

Shown in Fig. 2 (a) is the 10 x 10 μm ambient AFM image of ZnSe film sputtered on Si wafer for 2.5 hours at 80W power and 150 militorr pressure while the substrate was being annealed at 350^{0}C. Fig. 2(b) is a closer look of a 1 x 1 μm zooming size ZnSe film surface sputtered for 1 hour at 50 W, 140 militorr and 350^{0}C. As can be seen, the grown films are generally very uniform. Roughness analysis showed that the films have a mean roughness of 0.3 nm which is in the same order of bare Si wafer (O. 1 nm). Compared to the overall thickness of the film which is in the order of several hundreds nanometers as measured by scanning electron microscopy (SEM), they are considered relatively very smooth. Microscopically, this is confirmed by examining the films under optical microscope. The Er-doped sample on the other hand, appears to have much rougher surface (Fig. 3, mean roughness = 4.5 nm) and quite different surface morphology compared to that of the undoped sample. We attribute this to the presence of erbium which may create small surface clusters.

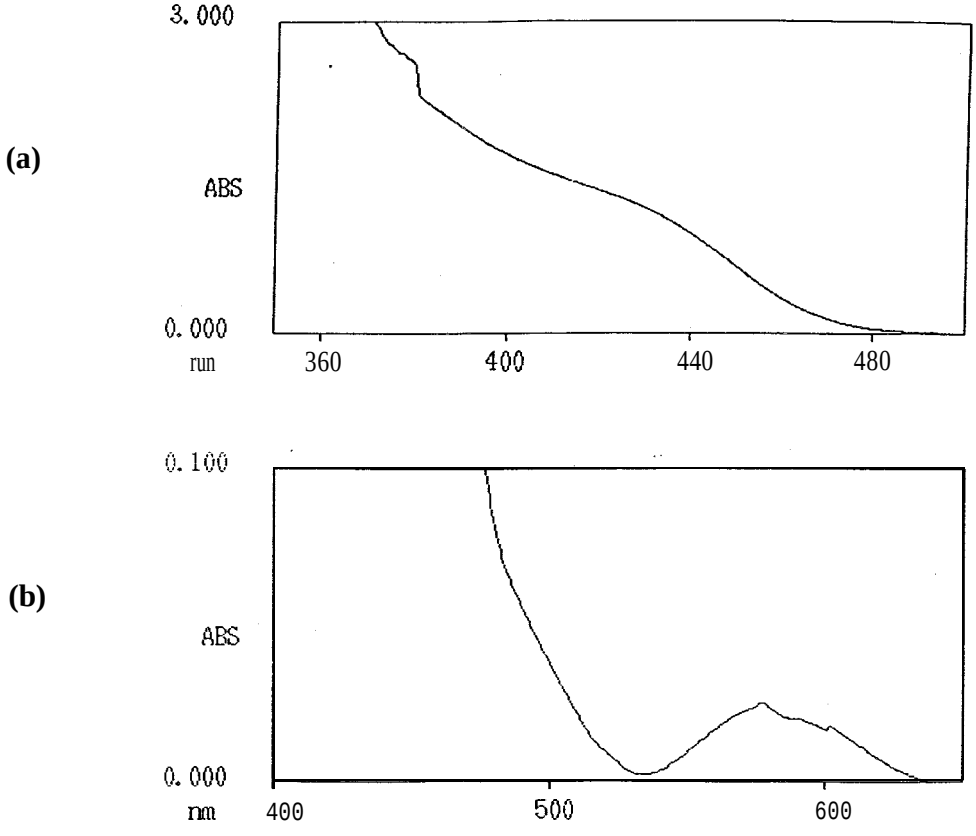


Fig. **4.** UV-Vis spectra of ZnSe:Er film showing absortion peaks of Er around 375 nm (a) and 575 nm (b).

Nevertheless, the image shows a homogeneous and uniform distribution of grain structures thus once again indicating that the grown film is of structurally good quality.

UV-Vis was then applied to detect the presence of ZnSe films as well as Er within the doped films. This is based on the fact that for an electron to move from the valence band to the conduction band it must acquire an energy greater than or equal to the band gap energy, from the absorbed photon. This would result in a sharp absorption edge in the UV-Vis spectrum. Ideally, the spectra of a uniform film would show a sharp absorption edge. However, if the film is not perfectly uniform it is possible for the overall spectrum to have a gradual slope as opposed to the abrupt change in the absorbance in the spectrum at the bandgap location. Our UV-Vis results of ZnSe films do show a rise of absorption edge around 461 nm which corresponds to that of ZnSe's whereas Fig.4 shows additional shifts at around 375 nm (4 (a)) and 575 nm (4 (b)), indicating the presence of Er in the film since Er has two main absorption peaks around 375 nm and 575 nm. Hence, our desired films were verified and confirmed.

Finally, it is also interest to look at the change in surface morphology after plasma etching. This is because reactive ion etching is a good method for the production of low dimensional structures[12] and nanocrystalline ZnSe film has recently capture interest due to its application in thermally stable and frequency selective lasers as well as photodetectors[13]. We have studied and compared the the AFM surface images of ZnSe:Er before and after 10 minutes of Ar plasma etching. The density of the plasma source is $10^8/cm^3$. The electrons temperature (kinetic energy) is in the order of 5-10 eV whereas that of ions is about 0.1 eV. The details of the plasma etching experiment and the AFM images are presented elsewhere[14]. The results showed that the plasma has significantly altered the surface structure of the ZnSe:Er film. In addition, roughness analysis showed that the mean roughness has dropped from 4.5 nm to 0.17 nm after ions etching indicating that a significant change has taken placed within the film as expected.

Nevertheless, not until a lot more characterization techniques such as auger electron spectroscopy (AES), x-ray diffraction analysis (XRD), photoluminescence (PL)... being done can we actually confirm the composition as well as the optical properties of the films. The study however, has shown the possibility of growing ZnSe and especially Er-doped ZnSe films by the RF magnetron sputtering technique with potential application in light emitting semiconductor devices due to the 4f-4f optical transition[9,15] in rare-earth doped semiconductors.

Summary

In conclusion, thin ZnSe films with and without erbium doping have been successfully grown by RF magnetron sputtering method. UV-Vis is a useful tool to identify semiconductor ZnSe film due to its absorption spectrum change at the bandgap. This is also true for ZnSe:Er since Er has several absorption peaks at wavelengths different from ZnSe's. From AFM analysis, doped samples were found to have higher surface roughness and quite different surface morphology compared to that of the doped samples. In general, the grown films have a relatively smooth and uniform surface indicating that they are of overall good quality. Although ZnSe:Er were successfully produced, more in-dept

studies and characterizations will be needed to determine the optimizing growth condition of this film to make it useful for light emitting devices application.

Acknowledgment

The authors acknowledge the support of NASA through the Fisk Center for Photonic Materials and Devices, by Grant NCC8- 133, of the Department of Energy and Lawrence Livermore National Laboratories, by Grant B336497.

References

1) A. Schneider, K-P. Geyzers, F. E. G. Guimaraes and M. Heuken, Adv. Mat. Opt. Elec. Vol 3,261 (1994).

2) J. Ren, K. A. Bowers, B. Sneed, D. L. Dreifus, J. W. Cook, Jr., J. F. Schetzina and R. M. Kolbas, Appl. Phys. Lett. 57 (18)29 October, 1901 (1990).

3) M. Imaizumi, Y. Endoh, K-i. Ohtsuka and T. Isu, J. Crystal Growth 136,366 (1994).

4) S. Fuijta, H. Mimoto and T. Noguchi, J. Appl. Phys. 50(2), February, 1079 (1979).

5) M. Godlewski and K. Swiatek, J. Crystal Growth, 117,634 (1992).

6) S. H. Sohn and Y. Hamakawa, Jpn. J. Appl. Phys. 31, Pt. 2, No. 7B, L963 (1992).

7) J. Saraie, K. Yamawaki, N. Matsumura and A. Ikehara, J. Crystal Growth, 159,334 (1996).

8) I. S. Hauksson, J. Suda, M. Tsuka, Y. Kawakami, Sz. Fujita and Sg. Fujita, J. Crystal Growth, 159,329 (1996).

9) J. Dzieiaty, St. Muller, R Boyn, Th. Buhrow, A. Klimakow and J. Kreissl, J. Phys.: Condens. Matter 7,4271 (1995).

10) C. T. Tsai, D. S. Chuu, G. L. Chen and S. L. Yang, J. Appl. Phys. 79 (12), 15 June, 9105 (1996).

11) H. K. Kim, C. C. Li, X. M. Fang, J. Solomon, G. Nykolak and P. C. Becker, Mat. Res. Sot. Symp. Proc. Vol. 301,55 (1993).

12) E. Wirthl, H. Straub, M. Schmid, H. Sitter, P. Bauer and G. Brunthaler, J. Crystal Growth, 159,746 (1996).

13) S. Banerjee, R. Pal, A. B. Maity, S. Chauhuri and A. K. Pal, Nanostructured Materials, Vol. 8, No. 3,301 (1997).

14) G. Lampkin, E. Thomas, Jr., M. Watson, K. Wallace, H. Chen and A. Burger, NASA URC-TC, Feb. 21-26, Hunstville, AL (1998).

15) K. Lobe, R. Boyn and H. Zimmermann, J. Phys.: Condens. Matter 5,6335 (1993).

Biography

Timothy George has recently joined Fisk University as a graduate student majoring in Physics. He did his undergraduate study at Southern University in Baton Rouge, LA but has been actively involved in the Summer Research Program for undergraduates at Fisk University the last three summers in the NASA/Fisk Center for Photonics Materials and Devices. He is interested in solid state physics, thin film preparation as well as different techniques of materials and devices characterization.

1.54 μm Emission from Erbium implanted GaN for Photonic Applications

Myo Thaik arid U. Hömmerich*
Research Center for Optical Physics,
Department of Physics, Hampton University,
Hampton, Virginia 23668

R.N. Schwartz and R. G. Wilson
Hughes Research Laboratories
Malibu, California 90265

J. .M. Zavada
U. S. Army Research Office
Research Triangle Park, North Carolina 27709

Abstract

The development of efficient and compact light sources operating at 1.54 μm is of enormous importance for the advancement of new optical communication systems. Erbium (Er) doped fiber amplifiers (EDFA's) or semiconductor lasers are currently being employed as near infrared light sources. Both devices, however, have inherent limitations due to their mode of operation. EDFA's employ an elaborate optical pumping scheme, whereas diode lasers have a strongly temperature dependent lasing wavelength. Novel light emitters based on erbium doped III-V semiconductors could overcome these limitations. Er doped semiconductors combine the convenience of electrical excitation with the excellent luminescence properties of Er^{3+} ions. Electrically pumped, compact, and temperature stable optoelectronic devices are envisioned from this new class of luminescent materials. In this paper we discuss the potential of Er doped GaN for optoelectronic applications based on temperature dependent photoluminescence and photoluminescence excitation studies.

Keywords for Indexing:

III-nitride semiconductors, optical communications, photoluminescence, optical spectroscopy, photonic materials

1. Introduction

Erbium doped semiconductors are currently being developed as materials for electrically pumped, compact, and temperature stable light sources used in photonic applications [1,2]. Er^{3+} exhibits sharp and temperature stable luminescence at 1.54 μm due to the intra-4f-shell transition $^4I_{13/2} \rightarrow {}^4I_{15/2}$. Since the 1,54 μm emission from Er^{3+} coincides with the minimum absorption region of silica based optical fibers, Er doped semiconductors are suitable light sources for optical fiber communications [1-3], The photoluminescence (PL) of Er doped Si and other III-V compound semiconductors have been studied intensively, however, only weak Er^{3+} PL was observed at room temperature. It was noted by Favennec et al. that the Er^{3+} PL intensity was strongly dependent on both the bandgap energy of the semiconductor and the host temperature [4]. Figure 1 illustrates the variation of the Er^{3+} emission intensity as a function of temperature for various III-V semiconductors. The rapid decrease of the luminescence intensity with temperature is often referred to as thermal quenching. As can be seen from Fig. 1 the thermal quenching of the Er^{3+} PL decreases with increasing bandgap of the semiconductor host. Therefore, recent research activities focus on studying Er^{3+} in wide gap semiconductors [2].

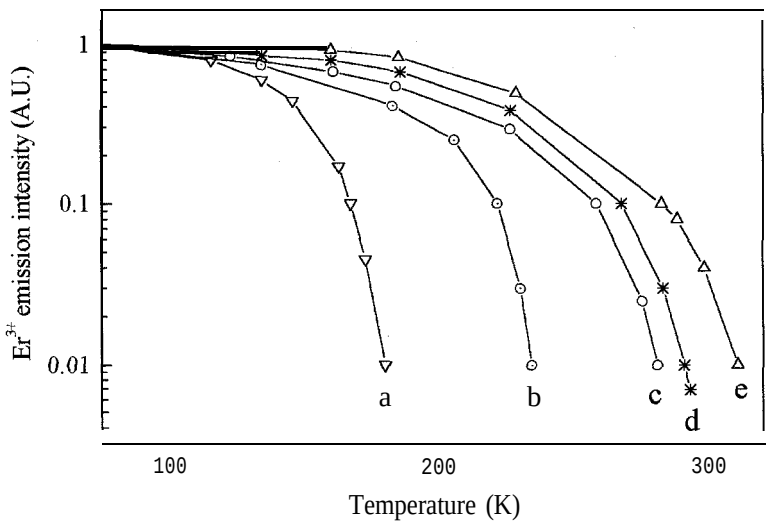


Figure 1: Variation of the Er^{3+} luminescence intensity as a function of temperature for various III-V compound semiconductors, (a) $Ga_{0.38}In_{0.62}As_{0.84}P_{0.12}$ (E_g = 0.807 eV), (b) Si (E_g = 1.12 eV) (c) InP (E_g = 1.27 eV) d) GaAs (E_g = 1.43 eV) (e) $Al_{0.17}Ga_{0.83}As$ (E_g = 1.67 eV) [4].

HI-nitride semiconductors - such as GaN, AlN, etc. - are currently being studied world wide as high brightness blue light emitting diodes (LEDs) and blue laser diodes [5,6]. The wide bandgap of III-nitrides ranging from 1.9 eV to 6.2 eV makes them also interesting candidates for Er^{3+} doping. Intense room temperature luminescence at 1.54 μm has been demonstrated from Er doped GaN and AlN [7-11]. These first results indicate that III- nitride semiconductors are suitable host material for Er^{3+} ions to emit intense 1.54 μm luminescence at room temperature. In this paper we present the optical properties of Er implanted GaN and discuss its potential for optoelectronic applications.

2. Experimental Details

A GaN thin film was grown by molecular beam epitaxy (MBE) and was Erbium implanted at an energy of 300 keV and a fluence of 2x1 014 cm^{-2} [7]. Photoluminescence measurements were performed using the 488 nm and 514.5 urn lines of an Ar ion laser as well as the 32.5 nm line of a HeCd laser as the excitation sources. An optical parametric oscillator (OPO) pumped by a Q-switched Nd:YAG laser was employed as the tunable excitation source in photoluminescence excitation (PLE) measurements. 1.54 μm luminescence from Er doped GaN was dispersed by a 1m monochromator and detected with a liquid nitrogen cooled Ge detector. The signals were processed with a lock-in amplifier for PL measurements and with a boxcar averager for PLE measurements and recorded in a personal computer. For temperature dependent PL measurements between 15 and 300K, the GaN:Er sample was placed onto a cold finger of a closed-cycle helium refrigerator. A home made heating element was used for above 300 K PL measurements.

3. Results and Discussion

Photoluminescence spectra

Above bandgap excitation of GaN:Er was achieved using the 325 nm line from a HeCd laser. This excitation wavelength leads to the creation of electron-hole pairs which subsequently recombine and transfer their energy nonradiatively to Er^{3+} ions. Subsequently, the relaxation energy of the excited Er^{3+} 4f-shell electrons results in 1.54 μm luminescence. The creation of electron-hole pairs is important for future device development because it simulates the operation of an electrically pumped photonic device. Figure 2 shows the high-resolution PL spectrum of GaN:Er for above bandgap excitation. Characteristic 1.54 μm emission was observed which can be attributed to the intra-4f-shell transition $^4I_{13/2} \rightarrow {}^4I_{15/2}$ of Er^{3+}. The PL spectrum had a full width half maximum (FWHM) of - 80 nm which suggests that the PL spectrum is inhomogeneously broadened [10,11]. The broadening points out that Er^{3+} ions occupy a wide range of sites in the GaN host 'with slightly different atomic configurations [12]

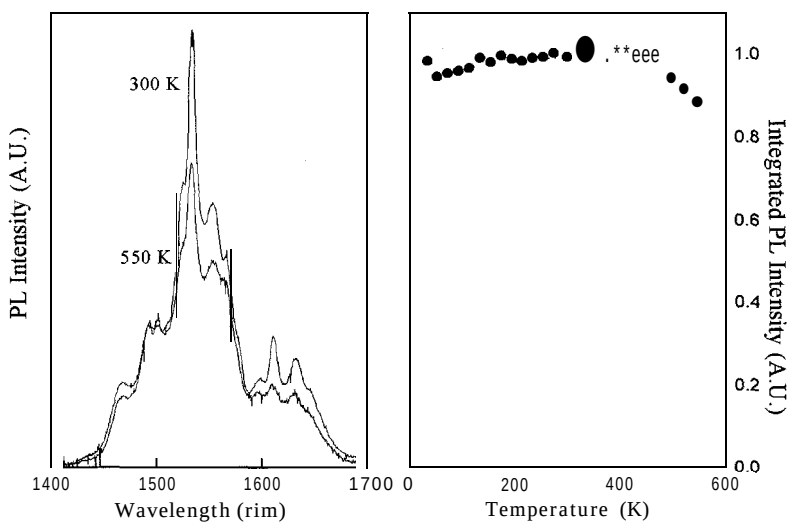


Figure 2: Above bandgap photoluminescence spectra of GaN:Er at 300 K and 550 K. The integrated PL intensity was found to be nearly temperature independent up to 550K.

Figure 3 depicts the room temperature PL spectrum of GaN:Er recorded for below gap excitation using the 488nm line of an Argon laser. The below gap Er^{3+} spectrum had a FWHM of ~50 nm. It was noted that below and above gap excitation led to different Er^{3+} PL spectra which indicates that depending on the excitation scheme, different subsets of Er^{3+} ions with distinct Er^{3+} PL properties are excited. The integrated Er^{3+} PL intensities from 15 to 550 K for above and below gap excitation are also illustrated in Figure 2 and 3. Relative to its value at 15 K, the integrated Er^{3+} PL intensity at 550 K decreased by 10 % for above gap excitation and by 50 % for the below gap excitation. The temperature stability of the Er^{3+} PL intensity from GaN:Er provides further evidence that thermal quenching of Er^{3+} PL is less in wide bandgap semiconductors [4]. It also demonstrates that GaN:Er is an attractive material for high temperature optoelectronic applications.

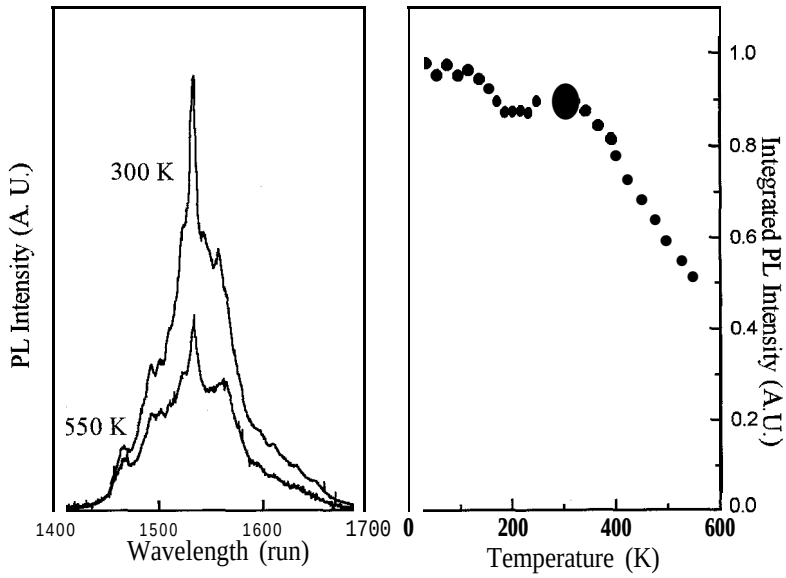


Figure 3: Below bandgap photoluminescence spectrum of GaN:Er at 300 K and 550 K. The integrated PL intensities was found to be nearly temperature independent up to 5 50K.

Photoluminescence Excitation (PLE)

Photoluminescence excitation spectroscopy is a powerful tool to identify below gap absorption bands in thin film materials [9]. The PLE spectrum of GaN:Er is shown in Figure 4. The PLE spectrum consists of a broad band and sharp peaks located at 495 nm, 525 nm, 553 nm, and 651 nm. The sharp peaks correspond to energies resonant with intra-4f Er^{3+} absorption bands. The broad band spanning from 420 nm to 680 nm in the PLE spectrum is most likely due to carrier-mediated, non-resonant, Er^{3+} excitation processes involving defects in the GaN host. It is interesting to note that the PLE result of GaN:Er is similar to that of AlN:Er [9]. The PLE result suggests that resonant and non-resonant excitation schemes are inherent characteristics of Er^{3+} ions in III-nitride wide bandgap semiconductors.

Pump Power Dependent Photoluminescence

In order to gain more insight in the Er^{3+} PL properties of Er implanted GaN pump power dependent PL measurements were carried out. Figure 5 shows the Er^{3+} PL intensity as a function

of pump power at different temperatures. At each temperature the 1.54 μm Er^{3+} PL intensity initially increases linearly with pump power followed by the start of saturation at higher powers. Saturation indicates that a large fraction of the optically active Er^{3+} ions are excited. That means the maximum obtainable Er^{3+} luminescence intensity is limited by the number of optical active Er^{3+} centers in the GaN film. It can also be inferred from Figure 5 that both the saturation level and the pump power at which saturation occurs depend on the temperature. Temperature dependent lifetime measurements (not shown) reveal that the Er^{3+} decay time decreases slightly with increasing temperature due to nonradiative decay. However, the change in lifetime does not completely account for the observed difference in the saturation level at different temperatures. Preliminary modeling of the pump power data suggest that besides the onset of nonradiative decay also the Er^{3+} excitation efficiency decreases at higher temperatures. More detailed investigations on the pump power dependent PL behavior of Er^{3+} are currently in progress.

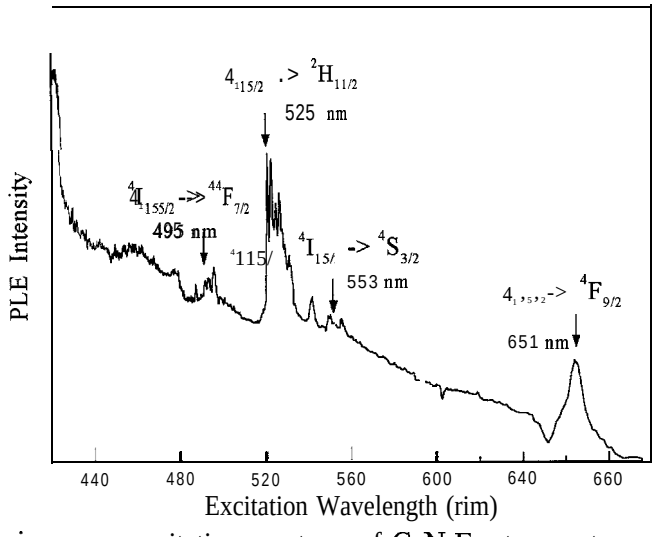


Figure 4: Photoluminescence excitation spectrum of GaN:Er at room temperature.

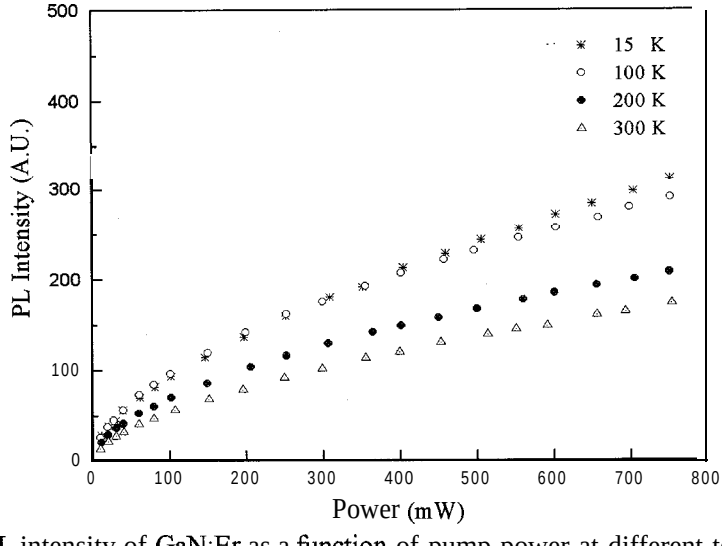


Figure 5: Er^{3+} PL intensity of GaN:Er as a function of pump power at different temperatures

4. Conclusions

Results of photoluminescence and photoluminescence excitation studies of Er implanted GaN were presented. Strong 1.54 μm Er^{3+} luminescence was observed not only at room temperature but also at high temperature (550 K) for above and below bandgap excitation. The integrated Er^{3+} PL intensity was found to be nearly constant up to 550 K which indicates the potential of Er implanted GaN for high temperature optoelectronics. PLE studies revealed that Er^{3+} ions can be excited through direct optical excitation of intra-4f-shell electrons or through indirect energy transfer processes involving defects in the host. Preliminary pump power dependent PL studies suggest that the Er^{3+} PL intensity is limited by the number of optical Er^{3+}, the Er^{3+} excitation efficiency, and the onset of nonradiative decay at higher temperatures. More studies of the Er^{3+} PL properties as a function of material synthesis parameter will be carried out to optimize Er doped GaN films for future device development.

Acknowledgment

The authors from Hampton University would like to acknowledge the financial support by NASA through Grant No. NCC- 1-251 and the Army Research Office through Grant No. DAAH04-96-1-0089.

References

1. G. S. Pomrenke, P. B. Klein and D. W. Langer (Eds.), Rare Earth Doped Semiconductors, *Mat. Res. Sot. Symp. Proc.* Vol. 301 (1993).
2. S. Coffa, A. Polman, and R. N. Schwartz (Eds.), Rare Earth Doped Semiconductors, *Mat. Res. Sot. Symp. Proc.* Vol. 422 (1996).
3. J. M. Zavada and Dahua Zhang, Solid State Electron. 38, 1285-1293 (1995).
4. P. N. Favennec, H. L'Haridon, D. Moutonnet, M. Salvi, and Y. Le Guillou, Electron. Lett. 25,718-719 (1989).
5. S. Nakamura, T. Mukai, and M. Senoh, J. Appl. Phys. 76,8189-8191 (1994).
6. Shuji Nakamura, Masayuki Senoh, Shin-ichi Nagahama, Naruhito Iwasa, Takao Yamada, Toshio Matsushita, Yasunobu Sugimoto, and Hiroyuki Kiyoku, Appl. Phys. Lett. 70, 868-870 (1997).
7. R. G. Wilson, R. N. Schwartz, C. R. Abernathy, S. J. Pearton, N. Newman, M. Rubin, T. Fu, and J. M. Zavada, Appl. Phys. Lett. 65, 992-994 (1994).
8. C. H. Qiu, M. W. Leksono, and J. I. Pankove, J. T. Torvik, R. J. Feuerstein, and F. Namavar, Appl. Phys. Lett. 66, 562-564 (1995).
9. X. Wu, U. Hömmerich, J. D. MacKenzie, C. R. Abernathy, S. J. Pearton, and J. M. Zavada, Appl. Phys. Lett. 70,2126-2128 (1997).
10. C. H. Qiu, M. W. Leksono, J. I. Penkove, R. J. Feuerstein, and F. Namavar, Appl. Phys. Lett. 66,562-564 (1995).
11. J. T. Torvik, C. H. Qiu, R. J. Feuerstein, J. I. Penkove, and F. Namavar, J. Appl. Phys. 81, 6343-6350 (1997).
12. T. D. Culp, U. Hömmerich, J. M. Redwing, T. F. Kuech, and K. L. Bray, J. Appl. Phys. 82, 368-374 (1997).
14. S. Kim, S. J. Rhee, D. A, Turnbull, E, E. Reuter, X. Li, J. J. Coleman, and S. G. Bishop, Appl. Phys. Lett. 71,231-233 (1997).

Purification and Crystal Growth of Lead Iodide by Physical Vapor Transport Method

G. W. Wright, M. Cole, Y.-F. Chen, K.-T. Chen, H. Chen, K. Chattopadhyay and A. Burger
NASA/Fisk Center for Photonic Materials and Devices,
Fisk University, Nashville, TN 37208

Abstract

Lead iodide (PbI2) is a layered compound semiconductor being developed as room temperature x- and gamma-ray detector. Compared to the more studied material, mercuric iodide, PbI2 has a higher melting temperature and no phase transition until liquid phase which are indications of better mechanical properties. In this study, the source material was purified by the zone-refining process, and the purest section was extracted from center of the the zone-refined ingot to be grown by physical vapor transport (PVT) method. The zone-refined material and as-grown crystals were characterized by optical microscopy and differential scanning calorimetry (DSC) to reveal the surface morphology, purity and stoichiometry. The results shows that both materials are near-stoichiometric composition, with the purity of the as-grown crystals higher than zone-refined materials. The resistivity of the as-grown crystal (10" Ω·cm) was derived from current-voltage (I-V) measurement, and is 10 times higher than the zone-refined materials. Detail results will be presented and discussed.

INTRODUCTION

Physical Vapor Transport Method utilizes the vapor pressure of sublimed source material to transport source material through a decreasing temperature gradient until the sublimed source material reaches region inside the furnace where the temperature is cool enough for the sublimed source material to solidify into a orderly arranged crystal. A temperature profile is obtained for our furnace. The temperature profile illustrates the distribution of thermal heat flow inside the furnace. obtained with a region constant at the sublimation temperature. From the temperature profile of the furnace the region inside the furnace where there is a constant decreasing temperature gradient, (in the direction of the negative slope) is where the crystal formation will occur.

Crystal solidification is driven by the dissipation of heat. The steeper the slope of the temperature gradient the faster the crystallization process can occur and the greater the chance a single crystal can be grown, The steep temperature gradient allows for the latent heat of crystallization to be conducted through the bulk of the crystal and faster in a radial direction through the growth interface. The growth interface is the surface at which phase transformation crystal growth takes place.

If the temperature gradient is not steep enough the latent heat of crystal formation may not be conducted fast enough through the bulk; thereby, part of the latent heat may remain in the growth interface increasing the temperature of the growth interface, thus not allowing solidification of the vapor phase to take place. In this case the center of the growth interface would be at a higher temperature than the edge around the circumference of the growth interface; thus allowing the crystal to grow faster along the edge resulting in a concave growth interface.

Experimental

First, the ampule was fabricated out of quartz material utilizing a blow torch. Next, we cleaned the ampule with aqua regia solution (25% HNO_3 and 75% Hcl). The ampule was outgassed to to get rid of any gaseous species trapped inside the quartz wall of the ampule. The zone-refined PbI_2 source material was placed in the ampule. The growth ampule was sealed under a vacuum at 10-6 torr. The transparent horizontal furnace was calibrated utilizing a variable autotransformer was held at10% of 120 Volts. The temperature profile of the furnace was obtained utilizing a thermocouple. The growth ampule is placed inside the furnace at the position that would locate the source material at a constant temperature such that the source material would sublime with enough vapor pressure that will allow source material to pass through the temperature gradient. The position ing of the ampule is determined from the temperature profile of the transparent horizontal furnace.

Optical Microscopy

We have utilized a microscope that has a resolution of approximately 5 microns. Any item smaller than 5 microns loses its definitive dimensions. Optical microscopy is utilized to characterize surface morphology of crystal up to the 5 micron level. It allows us to obtain some information about the actual shape of the crystal, whether the shape is platelet, rhomboidal, hexagonal, etc.; whether we have a single crystal or polycrystal; allows us to see mechanical defects due to cleavage and precipitates and inclusions larger than or equal to the 5 micron level.

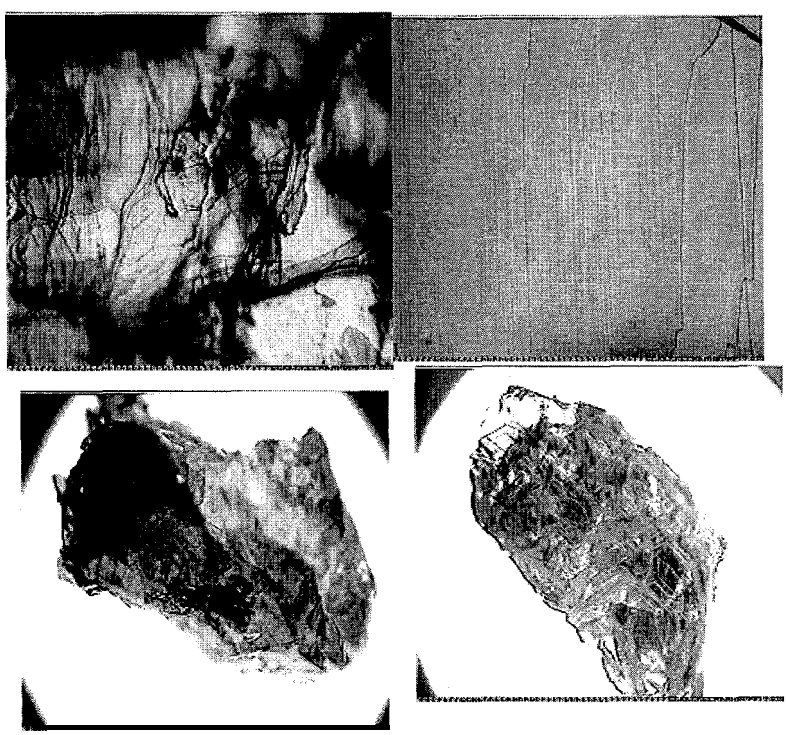
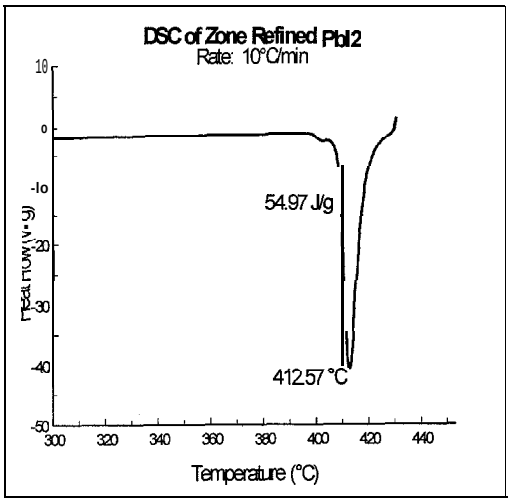


Figure 1. Optical micrographs of as-grown and cleaved lead iodide platelets. Scan size is 1 .2X1.2 mm^2



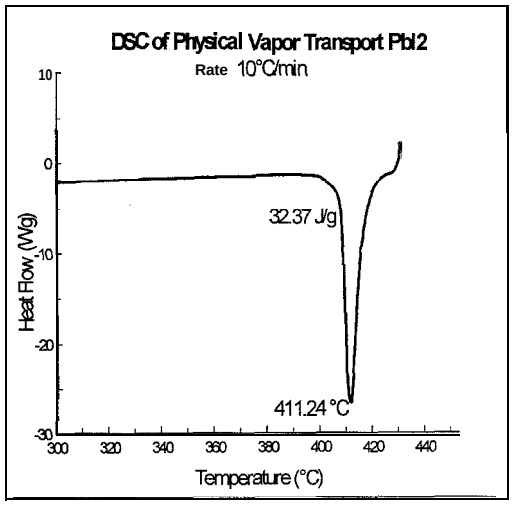
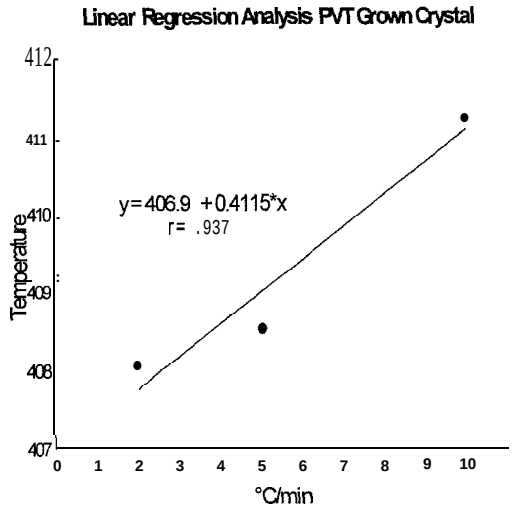


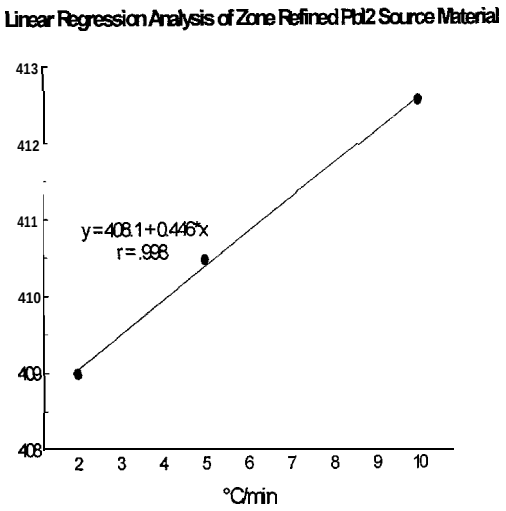
Figure 2. DSC thermograms of zone refined and PVT grown crystals

Differential Scanning Calorimetry

Differential Scanning Calorimeter measures the difference in heat flow supplied between an

Figure 3. Linear regression of DSC results





empty calorimetric cell and a calorimetric cell containing the sample as a function of temperature. Through the difference in measure a graph of heat flow vs temperature is obtained. In this graph the melting point or temperature at which a phase transition is occurs is represented by peaks, and the heat of transition is measured by the area contained within the peaks.

Current -Voltage (I-V) Characteristics

The detectors are fabricated and connected to a voltage source. The voltage is increased in increments of 10 volts until a maximum of 180 volts for PVT grown crystal and 200 volts for zone refined source material. At each increment of increased voltage the current is measured in picoamperes. A graph of current (I) vs. voltage (V) is obtained. The resistivity is calculated as the inverse slope of the I-V curve.

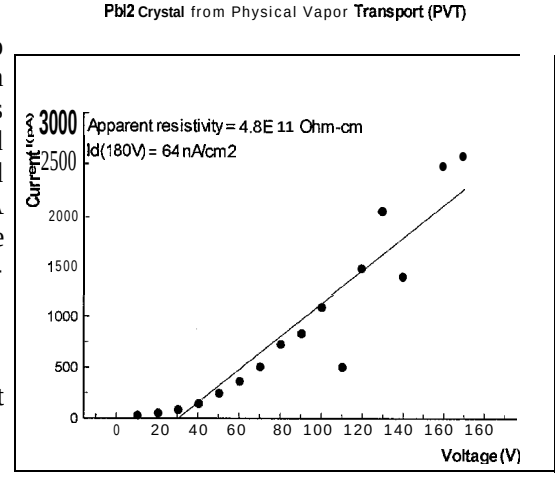


Figure 4 I-V characteristics of a lead iodide platelet grown by Physical Vapor Transport

The resistivity is an intrinsic property independent of the sample size. The maximum current density (I_d) is an extrinsic property (size dependent) is calculated by obtaining the current at the maximum voltage that exhibits a linear relationship between the current and voltage, then dividing the maximum current by the area thickness of the crystal.

Results and Discussion

The zone refined source PbI_2 and PVT PbI_2 crystal where characterized by optical microscopy, differential scanning calorimetry, and I-V methods.

The optical microscopy showed that the surface morphology of the crystal where not smooth and some inclusion were identified in the bulk. No precipitates of impurities were located on the surface indicating that the stoichiometry of both crystal were pure and did not change,

The DSC for both crystals were ran at 2,5,10 ^{0}C All three DSC graphs showed one peak with decreasing intensities as the temperature rate was decreased, however, for all three different temperature rates there were a high degree of sharpness for each peak produced by the respective temperature rates. The fact that each DSC showed a peak, and only one peak, at approximately the same temperature with approximately the same heat of fusion indicates that both the zone refined source material and PVT PbI_2 grown crystals exhibit a high degree of stoichiometry. Furthermore, we did not obtain any peaks at the melting point of lead(Pb) [327^{o}C] or the melting point of iodide(I_2) [113^{o}C] indicating the zone refined PVT crystal have good

stoichiometry. The sharper the peak the greater the purity of the sample. A measure of sharpness is to obtain FWHM (Full Width Half Maximum) of the peak. Linear regression analysis was done for each crystal to obtain the true melting temperature. Equation obtained from linear regression analysis are as follows:

Zone Refined PbI_2	PVT PbI_2
$y = (408.1 \pm .12) + (0.4450 \pm .018)x$	$y = (406.9 + .69) + (0.4115 \pm .106)x$

when x=0 we obtain the y-intercept which represents the true melting temperature at 0°C/min. Utilizing the statistical error of the y-intercept value plus an experimental error of 2°C there is overlap in the temperature interval for zone refined source material and PVT PbI_2. However, when utilizing FWHM we obtain a width value of 5 .8°C for zone refined source material and a width value of 5 .0°C for PVT PbI_2. The width value for PVT PbI_2 is 16°/0 smaller than the width value for zone refined source material; thereby, we can conclude that the PVT grown crystal has a higher purity than the zone refined source material.

The I-V plot method gave the following results

Zone Refined PbI_2	PVT PbI_2
Resistivity = 6.4×10^{10} Ohm-cm	Resistivity = 4.8×10^{11} Ohm-cm
I_d (200V) = 4nA/cm^2	I_d (180V) = 64nA/cm^2

Where x represent the rate: °C/min and y represents the melting temperature, respectively,

Since resistivity is an intrinsic property we can compare the zone refined PbI_2 with the PVT grown PbI_2. The PVT PbI_2 grown crystal has a higher resistivity than the zone refined; thereby indicating that the PVT grown PbI_2 crystal could make abetter detector than the zone refined material. Thus, more electrons can be excited up to the conduction band via electromagnetic radiation; thereby allowing a considerable difference in the amount of current produced in the crystal by the electromagnetic radiation.

Conclusion

The zone-refined PbI_2 has been used as a source material for PVT growth. The DSC results show that the PVT PbI_2 is purer than the zone-refined material by 16% as revealed by FWHM. The resistivity of PVT PbI_2 is higher by 10 times than the zone-refined material Therefore, we conclude that the PbI_2 crystals obtained from PVT were more suitable for detector application than the PbI_2 zone refined source material. The PVT process we utilized improved the purity and resistivity, and maintained the stoichiometry of the crystal. The improved results were verified through peak analysis of DSC and resistivity obtained through current-voltage measurement. The next step is to do a spectrum analysis of the crystal to see what wave length of electromagnetic radiation affects the crystal; thereby, allowing us to determine what kind of radiation can this crystal detect and which wavelength range is the best detected by the crystal.

ACKNOWLEDGMENTS

This work was supported by NASA through Fisk/Center for Photonic Materials and Devices, grant #NAGW-2925.

REFERENCES

[1] S. Roth and W. R. Willling, Appl. Phys. Lett 18, p. 328 (1971).
[2] C. Manfredotti, R. Murri, A. Quirini and L. Vasanelli, IEEE Trans. Nucl. Sci. **NS-24**, p. 126 (1977).
[3] J. C. Lund, K. S. Shah, M. R. Squillante and F. Sinclair, IEEE Trans Nucl. Sci. **NS-35** p. 89 (1988).
[4] J. C. Lund, K. S. Shah, M. R. Squillante, L. P. Moy, F. Sinclair and G. Entine, Nucl. Inst. And Meth. **A283**, p. 299 (1989).
[5] J. C. Lund, K. S. Shah, F. Olschner, J. Zhang, L. P. Moy, Medrick and M. R. Squillante, Nucl. Inst. And Meth. **A322**, p. 464 (1992).
[6] J. Zhang, K. S. Shah, F. Olschner, J. C. Lund, L. P. Moy, K. Daley, L. Cirignano and M. R. Squillante, Nucl. Inst. And Meth. **A322**, p. 499 (1992).
[7] D. C. Dominique, R.B. James H. Feemster, R. Anderson, A. J. Antolak, D. H. Morse, A. E. Pontau, H. Jayatirtha, A. Burger, X. J. Bao, T. E. Schlesinger, G. S. Bench and D. W. Heikkinen, Mat. Res. Sym. Proc. 302, p. 335 (1993).
[8]T. E. Schlesinger, R. B. James, M. Schieber, J. Toney, J. M. Van Scyoc, L. Salary, H. Hermon, Goorsky, Nucl. Instr. And Meth. **A380**, p. 193 (1996).
[9] K.-T. Chen, D. Shi, S. H. Morgan, W. E. Collins and A. Burger, Proceeding of NASA University Research Centers (URC-TC '97), edited by M Jamahidi et al., 1, p. 157, February, Albuquerque, New Mexico (1997).
[10] H. Chen, J. Tong, D.T. Shi, G. H. Wu, K.-T. Chen, M. A. George, W. E. Collins, A. Burger, R.B. James, C. M. Stable and L. M. Bartlett, J. Appl. Phys. 80 (5), p. 1 (1996).

Biography

Gomez Wright is a graduate student at Fisk University in the Department of Chemistry. He has been involved in the purification, crystal growth, characterization, and detector fabrication process of material science. He is married with three children and plans to pursue a Ph. D degree upon graduation from Fisk.

NITRIDE SEMICONDUCTOR THIN FILMS GROWN BY N-RADICAL-ASSISTED PULSED LASER DEPOSITION

Félix E. Fernández and Manuel Pumarol
Department of Physics, University of Puerto Rico, Mayagüez, PR 00681-9016
and
Antonio Martinez
Department of Physics, University of Puerto Rico, San Juan, PR 00931

Abstract

Wide bandgap semiconductor materials, particularly the Group III - Nitrides (AlN, InN, GaN), have extraordinary potential for a wide range of technological applications. In the last few years, development of thin films of these materials has attracted rapidly increasing numbers researchers. One of the key issues of concern is that of the required growth temperature for films of sufficient quality in terms of stoichiometry, crystallinity, crystal orientation, and defect concentration. The most successful growth techniques, of those so far utilized, require high substrate temperatures. Deposition techniques providing higher kinetic energies of incident species offer an alternative route which might allow growth of good quality films at lower temperatures. Pulsed Laser Deposition can provide higher kinetic energies than most thin film growth methods. However, III-nitride thin films grown by PLD are often nitrogen deficient. We have been able to obtain good stoichiometry for aluminum nitride films even at room temperature by providing atomic nitrogen at low (thermal) energies during growth. Very good orientation was obtained on sapphire substrates at moderate temperatures (~ 500 C). We also report on preliminary work by the same method with GaN film growth from a liquid Ga target.

Introduction

Wide bandgap nitride semiconductor thin film materials offer extraordinary possibilities for microelectronic and optoelectronic applications. High working temperature and radiation-hard electronic devices, UV and blue-green laser diodes and LEDs, and solar-blind detectors are some of the possibilities, in some cases alresdy demonstrated, which make these materials attractive.

While diamond films received more attention at least up to a couple of year ago, the nitrides AlN, GaN and InN, having direct-transitions, hold greater promise as optoelectronic materials. In particular, direct bandgaps are required for active media in laser diodes. Potential applicability of these materials is augmented by the fact that while their bandgaps are quite different (6.3 eV for AlN, 3.4 eV for GaN, and 2.0 eV for In N), their structures are however very similar, much more so for AlN and GaN, and they form a solid solution. Thus, for the ternary $Al_xGa_{1-x}N$ it is feasible to tailor the bandgap. Superlattices are also a possibility with this pair of materials. It should be noticed that crystalline SiC wafers are now available commercially, although still quite expensive. The wurtzite-structure (6H) SiC basal plane has corresponding lattice constants close to those of Al N, which in turn are close to those of GaN. SiC shares desirable thermal properties with these materials. Also, both p and n conductivity can be achieved for SiC. Thus, while SiC has an indirect bandgap of 2.9 eV, and is thus less interesting for optoelectronic applications, this can be a very suitable substrate for fabrication

of some Al N/GaN devices.

In a very exciting developement, just a year ago, the first GaN blue laser was demonstrated by S. Nakamura.[1] Many important scientific and technical issues must be resolved before any of these materials can be common l y employed in devices, however. Film crystal l in ity and orientation, stoichiometry, defect concentration, doping, and metal ization are al l important concerns presenting difficult problems. Of the many techniques which have been employed to date for thin film growth of wide bandgap nitrides, metal-organic vapor-phase epitaxy (MOVPE) and molecular beam epitaxy assisted by reactive nitrogen (RMBE) have produced the best quality materials. Both require the use of fairly high substrate temperatures (- 1,000' C for MOVPE and over 700° C for RMBE) in order to surmount the kinetic barrier for formation? The lower temperatures in RMBE are possible through provision of well controlled reactive-N species. A very recent review of RMBE results for GaN growth has been given by Mohammad et al.[3] The thermal energies available for the evaporated metal species in MBE imply that high substrate temperatures must still be used. An alternative which might allow further lowering of the growth temperature is the provision of higher energy of the metal species, such as can be achieved by pulsed laser deposition (PLD). While PLD growth of AlN in particular has been attempted in the past,[4,5] these efforts, in which AlN targets were ablated in vacuum or in a N_2 atmosphere, resulted in nitrogen-deficient films. More recent work has led to high quality AlN films grown by PLD,[6] but relatively high temperatures were required. Clearly, a more reactive source for the nitrogen is desirable. For this purpose, similar approaches to those recently employed in RMBE could be suitable. One of the better alternatives in RMBE appears to be the provision of a low-energy atomic nitrogen beam during growth.[7,8] Accordingly, we have initiated an exploration of wide bandgap nitride growth by PLD assisted by atomic-N beam. The current report offers a summary of our initial results for AlN and GaN films.

Film growth description

In order to provide reactive nitrogen to the growing film, we use an atomic nitrogen source (Oxford Applied Research MPD21) which breaks up the nitrogen molecules by coupling RF energy to the N_2 gas flowing through a pyrolytic BN crucible. In order to maintain stable operation of the source, the RF power is feedback controlled by monitoring the $3s^4P \rightarrow 3p^4S^o$ nitrogen plasma emission line. This type of source has been shown to generate an appreciable flux on neutral N atoms.[9] Furthermore, the ion content of gas exiting the source is very small.[10] Flow of ultra-high purity N_2 into our source is through an mass flow controller with digital readout.

For this work, a dedicated vacuum chamber was fabricated. The chamber is pumped by a 550 l/s turbomolecular pump, which permits a relatively low pressure (- 10'4 Torr) even at the largest gas flow rates possible for the N source (up to 15 seem). The targets lie on a horizontal plane, which allows the use of liquid targets. Pressure is measured with a nude-type BA ion gauge. Gas composition can be monitored with a residual gas analyzer detecting masses up to 200 amu. A shutter is used to cover the substrate while the target is initially ablated. Substrates can be heated up to 1,000" C using a feedback-controlled molybdenum heater.

The laser used in our system is an excimer laser (Lambda Physik Compex 110) operating with KrF (248 nm emission), and incident at 45° to the target normal, after passing the focussing lens (f = 350 mm) and chamber window, both of which are fused quartz. In order to further

reduce losses, the lens is AR-coated for 248 nm.

A ceramic AlN target (99.8 % pure) was used for most AlN depositions. Some were done using an Al sputtering target (99.9995% pure). In these cases the target was rotated by means of a stepping motor in order to avoid crater formation. All GaN depositions were executed with a liquid Ga target (99.9999% purity before use). The gallium was kept in an alumina crucible placed directly under the substrate. To avoid freezing of the gallium, the crucible was kept warm by means of an external heating filament.

The substrates used for our depositions were (111) Si and (001) sapphire. Some samples were deposited on pyrolytic graphite substrates in order to determine concentration of Al, N, and O by RBS analysis. Substrates were attached to the substrate heater with a silver suspension in ethanol. The target-substrate distance was 5 cm.

Background pressures for all depositions were in the 10'8 Torr range. Before each deposition N_2 was admitted to the N-source, the plasma in the N-source was ignited and stabilized with the RF impedance matching circuit, after which power control was transferred to the plasma monitor. For all deposition runs reported here, N_2 flow was 10 seem, forward power was ~250W and reflected power was ~ 2 W. The laser fluence at the targets were in the range of 6 to12J/cm^2 for AlN, over 15 J/cm^2 for Al, and - 10 J/cm^2 for Ga. Pulse frequency was 20 Hz or 10 Hz for most depositions. Ablation of liquid Ga produces a deep purple plume.

Film characterization results

Thin films of AlN with good stoichiometry were obtained at room temperatures. The RBS scan shown on Fig. 1 corresponds to a sample grown on pyrolytic graphite, for which the composition was 50 $\pm$ 1 at. % for Al and N. The oxygen in this sample appears to be mainly on the interfaces, though no doubt there should be some in the film itself, particularly since the purity of available AlN ceramic targets is not very high. In at least one sample there was excess nitrogen in the film.

XRD analysis of the films was undertaken with a θ-2θ diffractometer and, in order to observe off-plane reflections, with a gated area array detector. At the lower growth temperatures, partial orientation can be observed mainly in off-plane directions, as shown in Fig. 2 for an AlN film grown on (1 11) Si at room temperature. At higher temperatures orientation is improved, but Si, as expected, cannot provide a good template for growth of oriented films. Basal plane- cut sapphire, although not well lattice-match to AlN, is known to be a better substrate because the AlN crystal can partially accommodate through a rotation about the c axis. Fig. 3 corresponds to an AlN film grown on (001) sapphire at 600' C. The film is well oriented with the substrate, as indicated by the strong AlN (002) reflection observed. The peak can naturally be observed also in a θ -20 scan, as shown in Fig. 4 for a similar film grown at 500' C. The sapphire (006) reflection, which is very weak, can be barely noticed in both instances.

The surface of the films was very smooth and transparent, with some particulate inclusions. Films grown from with an Al target had small Al droplets observable at X400 magnification. The refractive index in the visible range, and the film thickness were determined for the AlN films with a 44-wavelength ellipsometer (Woolam M-44) through model fits for the film-substrate system. In spite of the film thickness non-uniformity (of -6 %) a reasonably good fit was obtained

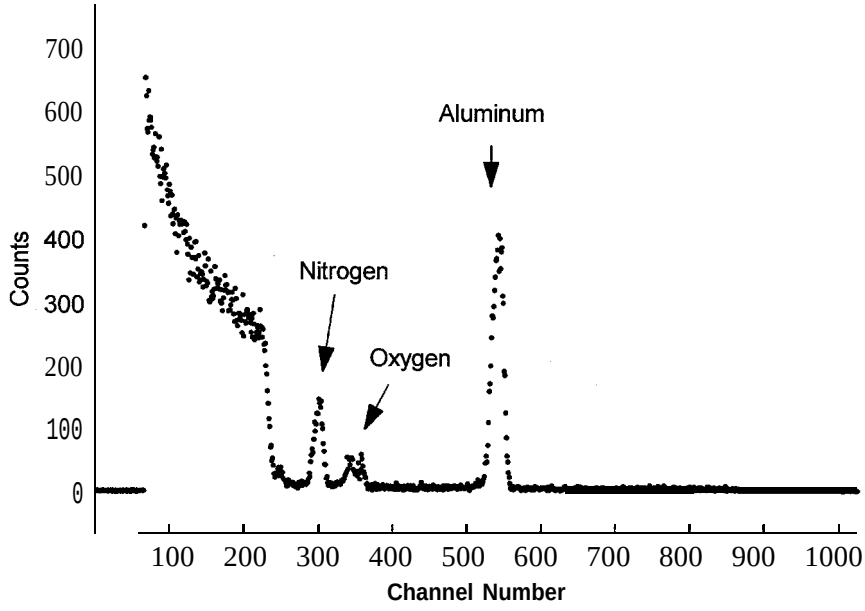


Figure 1 RBS spectrum of AlN film. Oxygen content is mainly on the interfaces.

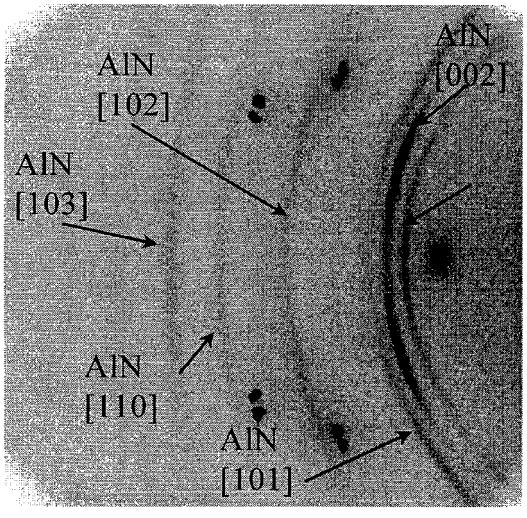
between the measured and calculated ellipsometric parameters as a function of wavelength. The refractive index of the films can be represented by a Cauchy function $n(\lambda) = A + B/\lambda^2 + C/\lambda^4$ (λ in μm), with A = 2.0306, B = 0.0165, and C = 0.0002, This compares reasonably well with known values for AlN. Determination of film thickness through the ellipsometric fits leads to an estimate for growth rates of - 0.14 Å / pulse for films grown from AlN targets at a fluence of - 6 J/cm^2. This is the same value obtained from the RBS results if bulk density is assumed, which along with the refractive index values confirms that the films have high densities.

For the GaN films, the initial results obtained indicate a similar pattern to those obtained for AlN. However, film stoichiometry, as determined by RBS for a film grown on graphite, was marred by some oxygen inclusion in the film, probably originating from a not entirely satisfactory arrangement to warm up the gallium. Very small (sub-micron) particles were observed on the films, apparently ejects from the gallium. It may be possible to reduce this by optimizing the fluence. XRD study of GaN films on (111) Si (Fig. 4) do not show much preferential orientation. Preliminary results from model fits to ellipsometric measurements indicate refractive index values somewhat lower than known values for GaN. This is presumably due to the oxygen content in the films.

Conclusions

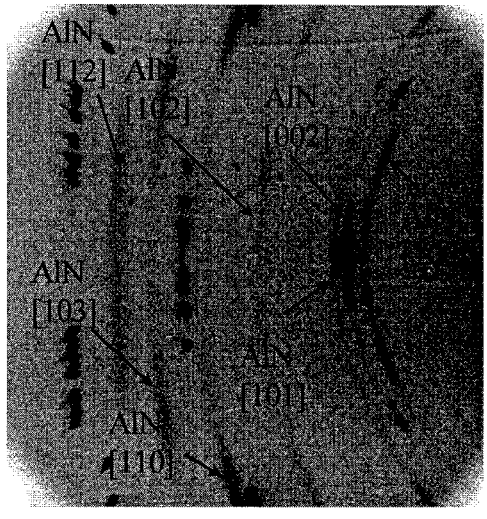
AlN thin films with good stoichiometry even at room temperature has been obtained by

providing atomic nitrogen at low (thermal) energies during PLD growth. Very good orientation was obtained on (001) sapphire substrates at moderate temperatures (- 500 C). The refractive index as a function of wavelength was determined for the films by ellipsometric studies. Preliminary work by the same method on GaN was also reported.



N117 AlN on (111)-Si T = 25 C

Figure 2 XRD for AlN on Si.



N119 AlN on AlO_3 T=600C

Figure 3 XRD of AlN on sapphire.

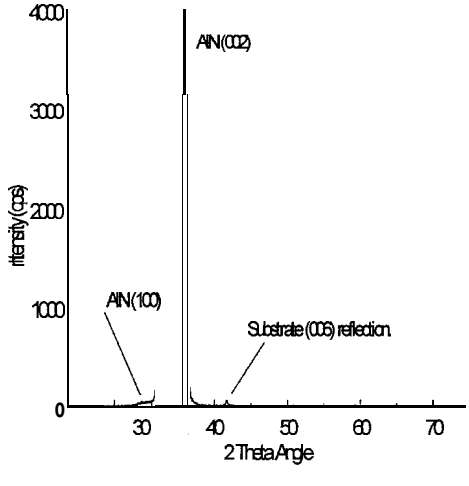
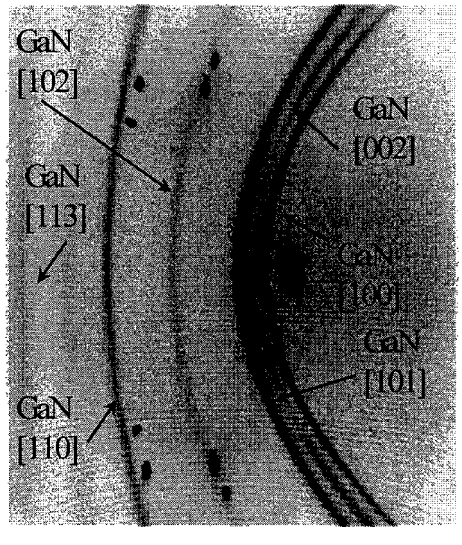


Figure 4 Diffractometer scan for AlN film on (001) sapphire substrate.



GaN on (111)-Si

Figure 5 XRD of GaN on Si. Growth temperature: 400 C.

References

1. S. Nakamura and G. Fasol, The Blue Laser Diode, Springer , Berlin (1997).

2. N. Newman *et* al., Mat. Res. Sot. Symp. Proc. 339 (1994) 483.

3. S.N. Mohammad *et al.,* MRS Bulletin, February 1997, Pag. 22.

4 G.F. Restrepo, Thesis, Univ. of Puerto Rico-RUM, Mayagüez (1994).

5. P. Battacharya and D. Bose, Jpn. J. Appl. Phys. 30 (1991) L1750.

6. R.D. Vispute, H. Wu, and J. Narayan, Appl. Phys. Lett. 67 (1995) 1549.

7. W.E. Hoke *et al.,* J. Cryst. Growth 111 (1991) 1024.

8. H. Liu *et al.,* J. Appl. Phys., 74 (1993) 6124.

9. R.P. Vaudo, J.W. Cook, and J.F. Schetzina, J. Crystal Growth 138 (1994) 430.

10. W.E. Hoke, P.J. Lemonias, and D.G. Weir, J. Cryst. Growth 111 (1991) 1024.

98 URC023

Radiation Effects on Resonant Tunneling Diodes: Preliminary Results

R. Wilkins, Shojah S. Ardalan and Eric Jackson
Center for Applied Radiation Research
Prairie View A&M University
Prairie View, TX 77446

J. Gryko
Jacksonville State University
Department of Physical and Earth Sciences
Jacksonville, AL 36265

ABSTRACT

We have observed enhanced bistability (hysteresis) in the current-voltage characteristics of gamma irradiated resonant tunneling diodes. This bistability occurs in the negative differential conductance region for voltages larger than 1.1 V. We discuss possible mechanisms for the observed phenomena and also discuss the measurements sensitivity to the external circuit.

INTRODUCTION

Resonant tunneling diodes (RTD) are quantum electronic devices with highly non-linear current voltage (I-V) characteristics. The I-V characteristics have a distinctive shape where the current rises to a peak with increased voltage followed by a sharp decrease in the current. This negative differential conductance (NDC) region is a consequence of electrons tunneling through a quantum well region between the two terminals of the device. The voltage where the peak current occurs is associated with the Fermi energy of the emitter being in resonance with a quantized electrons energy level in the well. Since the device action results from quantum tunneling, it is envisioned that RTDs can be used for fast electronic circuits. RTDs have been the subject of a recent review [1].

We have performed preliminary total dose radiation studies on two, InGaAs well RTD at room temperature. A hysteresis effect produces a current bistability in the I-V curve which is well known for these devices. We observe an enhancement of this bistability in the I-V curve of one RTD. Prior to irradiation the bistability had a width of 40mV, but increased to 200mV upon irradiation. After irradiation, the bistability width became smaller with a final value of about 80mV. No significant radiation effect was observed for another, smaller, device.

Numerical simulations show that enhanced hysteresis may be explained by a small shift of positive ions from the AlAs barrier into the InGaAs well. This shift occurs at some external threshold voltage and results in a sudden lowering of the resonant energy level in the well below the edge of the conduction band producing a corresponding sharp drop in current.

EXPERIMENTAL PROCEDURE

RTD samples were furnished to Prairie View A&M University (PVAMU) from Texas Instruments Corporation via collaborators at Texas A&M University (TAMU) and the University of Texas, Dallas. Two in-line packaged InP substrates with three RTD devices each were provided. The I-V characteristics of each device were first measured at room temperature and 4.2°K at the NanoFab Labs at TAMU. The I-V characteristics were then remeasured (at room temperature) in conjunction with TAMU using a parametric analyzer at PVAMU. The PVAMU test circuit is shown in Figure 1. Data taken with both test circuits were comparable.

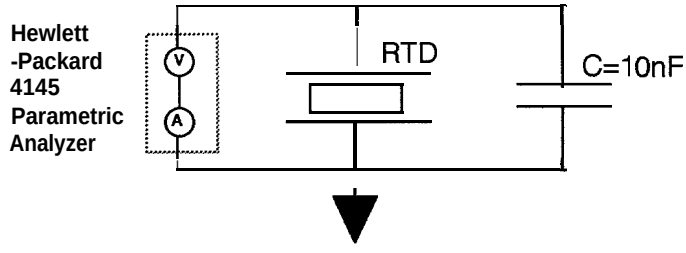


Figure 1: Schematic of the PVAMU test circuit. The parallel capacitor helps maintain constant voltage across the RTD on a time scale determined by the inductance of the measurement leads. The HP4 145 varies the voltage across the device and measures the resulting current.

The devices have been radiation tested for total dose effects using a low dose rate Co-60 source at the Texas A&M University Veterinary school. For our tests, the sample is placed under the source at a known distance, with material placed over the samples to assure proper dose for the materials in use. In the tests conducted to date, the dose rate has been held constant at 50rad/minute.

The testing procedure has been as follows:
1. A baseline I-V characteristic is taken on-site prior to irradiation.
2. The device is then given a specified radiation dose.
3. The I-V characteristic is remeasured just after irradiation and periodically thereafter.

RESULTS AND DISCUSSION

The energy band diagram of the RTD device used in these studies is shown schematically in Fig. 2. The experimental I-V characteristics are shown in Fig. 3. The I-V curve for the device of interest has a well known shape, with an exponential increase of current at low voltage, followed by a sharp decrease of the current in the NDC region. The NDC region shows two step-like decreases in current at approx. 0.68 V and 1.1 V respectively, separated by a flat region of almost constant current (constant with respect to the response time of the measurement circuit). Before irradiation the current-voltage curve shows a narrow hysteresis loop (Fig. 3A), with an approximate width of 40 mV. This device was irradiated with a total dose of 7 krad. After irradiation, the hysteresis loop becomes much wider, up to 0.2 V (Fig. 3B). The width decreases

slowly with time, but after two months it is approximately two times larger than before irradiation (Fig. 3C).

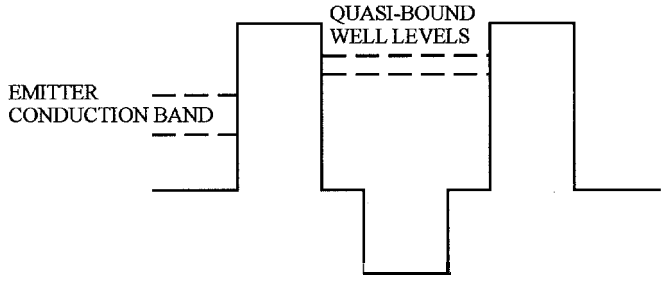


Figure 2: The band energy diagram of the RTD device. The emitter and collector layers are Sn doped InGaAs. The barrier is 2.3 nm thick AlGa, the well is 1 nm thick (undoped) InGaAs, and the 2 nm thick notch is undoped InAs. The AlAs barriers are approximately 0.87 eV high, and the notch is approx. 0.36 eV deep.

As shown in Fig. 3, the gamma radiation does not change the overall structure of the I-V curves significantly. The only changes that take place are in the region of flat current between 0.68 V and 1.1 V, with the loop at 1.1 V becoming much more pronounced. We concentrate here on changes that take place in the bistability region. The bistability of RTD devices is well known and it has been reviewed recently by several authors [1,2,3]. Although the origin of intrinsic bistability is still being debated, it is generally attributed to electron charge buildup in the well region. This charge buildup can be observed experimentally in magnetoconductance measurements [4,5].

Our measurements indicate that the bistability of the RTD device is significantly enhanced by gamma radiation. The vertical current transition in the hysteresis loop is very sharp, similar to those observed in unirradiated double barrier structures at low temperatures [6]. This suggests that the inhomogenous broadening of the upper edge of the quasi bound states in the well is very small. It is observed that the transition from the low to high current takes place in a very small voltage range of about 0.001 V. At an applied voltage of 1.195 V the current is very small. It remains small until the voltage reaches approximately 1.1 V. However, if the measurement is done starting at voltage of 1.194 V, the current is much higher, and remains approximately constant till 0.68 V. The difference in the current between high and low states corresponds to approximately 3x1 016 more conducting electrons present in the state of high current. This large difference in the number of current carriers is caused by only a - 1 mV change in the applied voltage. This effect is reproducible and represents an intrinsic property of the device; its significance will be discussed below.

The observed gamma radiation enhancement of bistability in this device cannot be filly explained by electronic charge buildup in the well region. The charge buildup in the well region has a dynamical character and is probably the same in irradiated and unirradiated devices. The magnitude of the charge buildup in irradiated and unirradiated devices will be checked in magenetoresistance measurements that we plan to carry out in the future. It may be that the

gamma radiation creates charged structural defects in the RTD device by breaking chemical bonds between atoms. These defects are analogous to surface atoms and therefore they should easily undergo transitions analogous to a surface reconstruction. The experimental data suggest that these defects have a long lifetime. The charged defects have two main effects on the electron states: they can trap conduction electrons and also they change the band bending in the accumulation layer of the emitter. Therefore, the emitter bands are bent more in the irradiated device and overlap stronger with the well quasi bound states. This in turn requires higher voltage to break the overlap and therefore the current in the device is high, until electron charge buildup in the well breaks down the overlap between the emitter and the well states. However, this description does not explain why the hysteresis loop vertical lines are so sharp at room temperature. In fact, one can expect that the charge buildup will gradually decrease current, resulting in a "round" hysteresis loop. This is not observed experimentally at room temperature.

To explain the sharpness of the hysteresis loop we postulate that charged atoms with broken chemical bonds, created by gamma radiation are shifted slightly at the threshold voltage in the interface region between the AlAs barrier and the InGaAs well. These are probably aluminum atoms swinging into the well region, or negatively charged arsenic atoms swinging into the barrier. This creates an additional attractive potential in the well causing the energy of the quasi bound well states to fall sharply below the lower edge of the emitter conduction band. Therefore, the slight shift of the charged defects in the barrier/well interface region causes a sharp decrease of the current. This shift in atomic positions is analogous to the surface reconstruction. When the applied voltage decreases, the charged ions move back to their original positions, at some lower threshold voltage, and the current in the system increases sharply. This model predicts that the width of the hysteresis loop will decrease with time after irradiation because the number of charged defects in the interface region will decrease with time, as is observed experimentally. However, the number of ions that can be shifted into the well at the threshold voltage is limited by the strong electrostatic interaction between charged defects. Therefore, after a second irradiation we observed that the device showed the same hysteresis loop as after the first irradiation, because additional defects that are created by the gamma radiation simply cannot be shifted into the well region. This measurement suggests that 1 krad radiation dose was sufficient to create almost the maximum possible number of interface defects. Finally, this model predicts that the width of the hysteresis loop should be larger than the width of the bistability originating from the electronic charge buildup only. This is simply because a higher voltage is required to shift charged defect than to buildup additional charge in the well.

The RTD device investigated above was relative large in size, as indicated by a large value of peak current through the device, approximately 12 mA. We also irradiated a much smaller RTD device with a peak current of approximately 50 microamps. The I-V characteristics of the small device is almost the same before and after irradiation, and it does not show hysteresis loop. This can be probably explained by the small lateral size of the quantum well, making it less likely to build up electron charge.

Our preliminary simulations, based on a simple, non-selfconsistent 1D model [7], seems to support the above conclusions about the nature of the hysteresis loop in irradiated RTD device. We have found that the simulated current strongly depends on the additional attractive potential

added to the well region. At present we are conducting more realistic simulations that we hope will help us to understand the nature of the radiation enhanced bistability.

Interpretation of these results is ongoing. There is some concern that the effects on the bistability are related to the test circuit, but we emphasize that:
1. A baseline characteristics was measured before any irradiation at the test site.
2. The only change in circumstance for the device represented in Figure 3 was that the in-line package holding the substrate was removed from the test circuit and irradiated, then placed back into the test circuit. The test circuit was not disturbed or changed during this process.

However, recent measurements in our own laboratory, with an improved test set up, have verified that small changes in the external circuit result in significant changes in the I-V characteristics. The priority for our future work is to obtain convincing data on the "true" or intrinsic I-V characteristic for our devices.

We have obtained additional devices from Texas Instruments. PVAMU will work with the TAMU team and Texas Instruments to address all concerns about the measurement of the I-V characteristics, then proceed with addition radiation tests and modeling of the observed effects.

ACKNOWLEDGMENTS

This work is partially supported by NASA CARR and the Air Force Office of Scientific Reseach. Texas Instruments Corporation provided the devices studied in this work. We would like to thank Wiley P. Kirk, Greg Spencer and Kevin Clark of the Texas A&M University NanoFab Center for device characterization and useful discussions. We also thank D. D. Wu for his work toward interpreting the data.

REFERENCES

[1] H. Mizuta and T. Tanoue, The Physics and Applications of Resonant Tunneling Diodes, (Cambridge Univ. Press, Cambridge, 1995).
[2] K. Ferry and C. Jacoboni (Eds.) Quantum Transport in Semiconductors, (plenum Press, New York, 1992).
[3] L. Chang, E. E. Mendez, and C. Tejedor, Resonant Tunneling in Semiconductors, (Plenum Press, New York, 1991).
[4] J. Goldman, D. C. Tsui and J. E. Cunningham, Phys. Rev. B35, 9387 (1987).
[5] A. Payling, E. S. Alves, L.Eaves, T, J. Foster, M. Henini, O. H. Hughes, P. E. Simmondas, F. W. Sheard, G. A. Toombs, and J. C. Portal, Surf. Sci. 196,404 (1988).
[6] J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. Lett. 58, 1256 (1987).
[7] R. Wilkins *et. al.* July 1997.

The current-voltage characteristics for the RTD device before irradiation. The hysteresis loop width is approximately 40 mV.

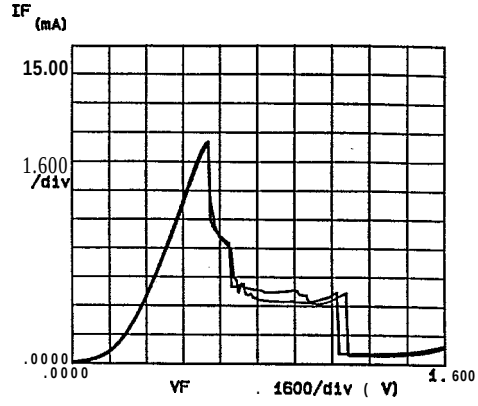


Fig. 3A

The current-voltage characteristics for the RTD device after a total dose of 7 krads. The hysteresis loop is now about 200 mV.

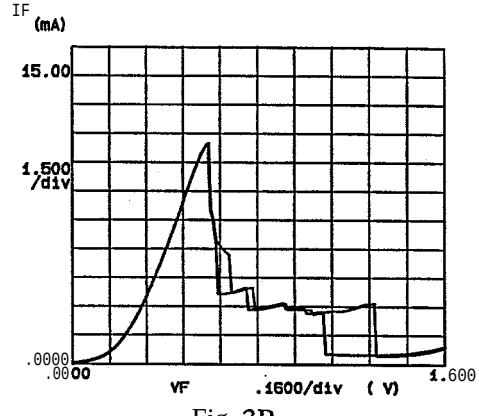


Fig. 3B

The current-voltage characteristics for the RTD device two months after the 7 krads total dose. The hysteresis **loop is about 80 mV wide, twice the** unirradiated **width.**

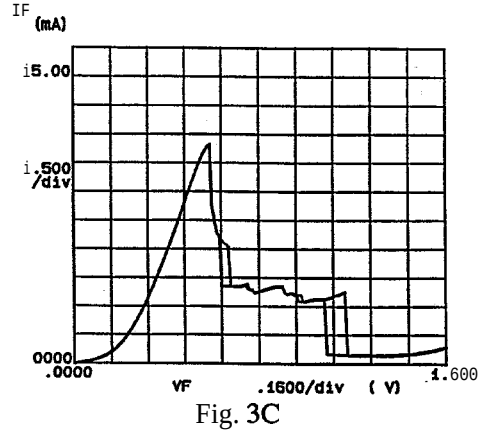


Fig. 3C

Figure 3

"Temperature Dependence of Diffusion and Reaction at a Pd/SiC Contact"

D.T.Shi, W.J.Lu, E. Bryant, K. Elshot, K. Lafate, H. Chen, A. Burger and W.E. Collins

NASA/Fisk Center for Photonic Materials& Devices
Fisk University, Nashville, TN 37208

Schottky diodes of Palladium/SiC are good candidates for hydrogen and hydrocarbon gas sensors at elevated temperature. The detection sensibility of the diodes has been found heavily temperature dependent. In this work, emphasis has been put on the understanding of changes of physical and chemical properties of the Schottky diodes with variation of temperature. Schottky diodes were made by depositing ultra-thin palladium films onto silicon carbide substrates. The electrical and chemical properties of Pd/SiC Schottky contacts were studied by XPS and AES at different annealing temperatures. No significant change in the Schottky barrier height of the Pd/SiC contact was found in the temperature range of RT-400 "C. However, both palladium diffused into SiC and silicon migrated into palladium thin film as well as onto surface were observed at room temperature. The formation of palladium compounds at the Pd/SiC interface was also observed. Both diffusion and reaction at the Pd/SiC interface became significant at 300 'C and higher temperature. In addition, silicon oxide was found also at the interface of the Pd/SiC contact at high temperature. In this report, the mechanism of diffusion and reaction at the Pd/SiC interface will be discussed along with experimental approaches.

Keywords: XPS and AES, Pd, SiC, Schottky Barrier.

1, INTRODUCTION

Hydrogen and hydrocarbons exist in a wide variety of fuels and their by-products. Sensitive detection of hydrogen and hydrocarbon is vital to the safe and effective use of these materials. Silicon-based semiconductor gas sensors, for example, Pd/Si system[1,2], have been investigated for over a decade for room or low temperature applications. The detection mechanism for hydrogen and hydrocarbons is as follows: the dissociation of hydrogen or hydrocarbon on the Pd metal surface leads to the formation of a dipole layer composed of hydrogen at the metal-semiconductor or metal-insulator interface. The electronic properties (for example Schottky barrier height) of a semiconductor device will be altered by the dipole layer according to the amount of hydrogen and other working gases present in the surrounding ambient atmosphere.

However, many applications involving hydrogen and hydrocarbon gases encounter severe environments, such as high temperature and high mechanical stress, in which silicon-based semiconductor devices fail to operate. Detection of these gases at elevated temperatures is especially important for monitoring and controlling emissions from, for example, an aircraft engine. The wide bandgap nature of silicon carbide (SiC) makes the SiC-based electronics a choice for high temperature applications. Besides, SiC, as a" semiconductor, also has many other superior properties such as good chemical stability, high thermal conductivity, high break down electric field, and extreme mechanical and radiation hardness. The rapid improvement of the technologies for growing high quality SiC bulk crystals and epilayers in the recent years has made the SiC-based electronics an emerging technology. Similar to the Pd/Si system, the Pd/SiC system has been explored for the detection of hydrogen and hydrocarbons at elevated temperatures[3]. Chen etc. [3] studied the Pd/SiC and Pd/SiO_2/SiC (with introducing a thin, 50 Å, SiO_2 layer into the Pd/SiC interface) diodes. They found that both diodes were sensitive to hydrogen at low temperature. The Pd/SiC system was still sensitive to the presence of hydrogen even after extended annealing (140 hours) at 425 "C although the detection sensibility was getting lower, while the Pd/SiO_2/SiC system became nearly insensitive to hydrogen after the same annealing. The decrease of detection sensibility of both Pd/SiC and Pd/SiO_2/SiC systems after the extended annealing was tentatively attributed to the formation of SiO_2 layer on the Pd surface and the formation of palladium silicides (Pd_xSi) throughout the palladium film (mainly for the Pd/SiC system). It is still unclear how the interracial diffusion and reactions change with temperature and how the changes affect the electric properties of the Pd/SiC system.

This paper presents a study of the diffusion and reaction mechanism at the interfaces of Pd/SiC contacts at Fisk University. The electrical property and chemical compositions of the Pd/SiC system were measured by x-ray photoelectron spectroscopy (XPS) and Auger electron spectroscopy (AES). The surface and interracial morphologies were obtained by atomic force microscopy (AFM) and ultrahigh vacuum scanning tunneling microscopy (UHV-STM), which will be presented in a separate paper. In order to directly probe the Pd/SiC interfaces by XPS/AES, ultra-thin palladium films were fabricated on the SiC surfaces. The use of ultra-thin Pd

films eliminates the needs of using ion sputtering or plasma etching that may change chemical compositions at the interfaces because of preferential sputtering or etching effects[3].

2. EXPERIMENT

Schottky diodes of Pd/SiC were fabricated by depositing ultra-thin palladium films onto a SiC substrate. A commercially available n-type 3.5° off-axis (0001) polished alpha 6H-SiC silicon-face wafer was diced for the substrate. A high doping level ($2.6x10^{18}/cm^3$ of nitrogen) was selected for STM study and I-V measurements. Prior to the coating of palladium, the SiC substrate underwent a series of chemical and thermal cleaning procedures to remove surface contaminants. Several chemical cleaning procedures were tested and the effectiveness of the cleanings was examined by XPS on silicon wafers and then on SiC wafers for comparison. The optimized cleaning procedure for SiC was as follows: diced SiC substrates were first washed with distilled water, then subsequently immersed and/or washed in $NH_4OH:H_2O_2:H_2O$ (1: 1:2) at 85 "C for 5 minutes, $HCl:H_2O_2:H_2O$ (1:1:2) at 85 "C for 5 rein, $HF:H_2O$ (20°/0 solution) at RT for 5 rein, HNO_3 (50°/0) at 90 "C for 5 rein, and $HF:H_2O$ (1 0°/0 solution) for 2 min. Between each step above, a sample was rinsed by de-ionized water. After chemical cleaning, the sample was kept immersed in methanol to prevent contamination.

An rf sputtering system (from Kurt J. Lesker Company) was employed to fabricate the Pd ultra-thin on a SiC substrate. The chemically cleaned sample was put in the sputtering chamber and baked at 350 "C (the maximum baking temperature allowed for the system) for 30 minutes at a base pressure of 1 xl 0^{-6} tom. The Pd film was fabricated with the following parameters: 5 inches' sample to Pd source distance, 35 w rf power, 20 second sputtering time, and about 70 "C of sputtering temperature. The purity of the Pd source was 99.95%. The resulting thickness of the Pd film was about 30 Å based on XPS measurements.

The sample was then studied with XPS and AES at the room temperature after a brief exposure to air during transportation and after different temperature annealing. A Kratos XSAM 800 spectrometer that contains both XPS and AES was employed for the sample analysis. The spectrometer with a resolution of 0.1 eV uses Mg Kα ($h\upsilon$ = 1253.6 eV) x-rays to create photo-emissions. The operating vacuum was about $1x10^{-8}$ torr. Annealed samples were obtained by sequentially putting the Pd/SiC sample into an oven which was maintained at 300 "C or 400 "C for 30 minutes each. Those annealing temperatures were chosen based on the observation of the existence of chemical reactions at 425 °C[3]. UHV-STM and AFM analyses of the samples were conducted after XPS/AES studies.

3. RESULTS AND DISCUSSIONS

A. Surface and Interface Compositions

Atomic and chemical compositions of Pd/SiC sample surfaces and interfaces were obtained from the quantitative analyses of XPS/AES spectra of different samples. Due to the variation of charging of different samples, the peak positions of different elements and chemicals were identified by using NIST XPS standards[4] and the peak position (284.6 eV) of the adventitious carbon on the sample surfaces, Table 1 lists the peak positions of C 1s (in SiC), Si 2p (in SiC) and Pd 3d (both metallic and compound), Wide scans of XPS showed that the major chemical species on the sample surface and interface were C, Si, O, and Pd (for Pd coated samples). A trace of F was found on the surface of the SiC substrate (less than 0.1 % of Si), which was the residue of the HF etchant. The detailed-scan XPS spectra of carbon, silicon and palladium and spectra of oxygen Auger peaks are shown in Fig. 1 for the samples of the SiC substrate, un-annealed Pd/SiC, 300 "C annealed and 400 "C annealed Pd/SiC at a take-off angle of 90° of photoelectrons.

Table 1. Binding Energies of Si(2p), C(1s) and Pd($3d_{5/2}$) and Schottky Barrier Heights(SBH).

XPS Peak	SiC substrate	unannealed Pd/SiC	annealed Pd/SiC at 100°C*	annealed Pd/SiC at 300°C	annealed Pd/SiC at 400°C
C (1s)(eV)	283.18	283.26	283.13	283.24	283.23
Si (2p) (eV)	101.14	101.21	101.12	101.18	101.16
Pd (3d)[a] (eV)		335 on	335.47	335.34	
Pd (3d)[b] (eV)		337.59	337.76	336.86	336.67
SBH		1.07	1.20	1.09	1.10

*) a different Pd/SiC diode for ref., a) metallic Pd, b) Pd compound, C and Si peaks in SiC.

The angular resolving technique (AR) of XPS was employed to distinguish the surface chemicals from interracial and bulk chemicals. Fig. 2 shows the ARXPS spectra of the 400 "C annealed Pd/SiC sample. Two sets of spectra in figure 2 correspond to two different take-off angles of the energy analyzed photoelectrons from the sample surfaces. The spectra in solid lines and in dashed lines correspond to the take-off angles of 90° and 60°, respectively. Because at small take-off angles the signal (the number of photoelectrons coming into the energy analyzer) from a sample surface becomes greatly enhanced relative to that from the bulk, the depth profiles of samples can be obtained by varying the take-off angles and comparing the normalized relative peak intensities. For example, on the carbon spectra in Fig. 2, the intensity of the highest component decreases with the decrease of the take-off angle, which means that component is mainly from the bulk. After energy calibration, that component indeed belongs to the SiC underneath the Pd coating. The second strongest carbon peak on the left shoulder (higher binding *energy)* of the highest peak shows an opposite behavior, that is, the peak intensity increases with the decrease of the take-off angle, which means that it came from the sample surface and actually, it was from the adventitious carbon (on sample surfaces). Two silicon species were clearly demonstrated by the Si spectra in Fig. 2. They were assigned to SiC (the larger peak) and SiO_2/SiO_X (the smaller one). The interesting behavior of the SiO_2/SiO_X peak that did not change with the variation of take-off angles indicates that this component existed in both surface and interface regions. The oxygen peaks did not change with take-off angles, either. However, the oxygen peak shown was actually the convolution of oxygen 1s peak and palladium 3p peaks. Both the oxygen peaks and palladium peaks were complicated themselves so that no quantitative data has been extracted from the spectra. Palladium signal was rather weak after sequential annealing at 300 "C and 400 "C, but Fig.2 still shows that more (larger density) palladium was on the surface than that in the deeper region of the sample, which indicates the diffusion of Pd.

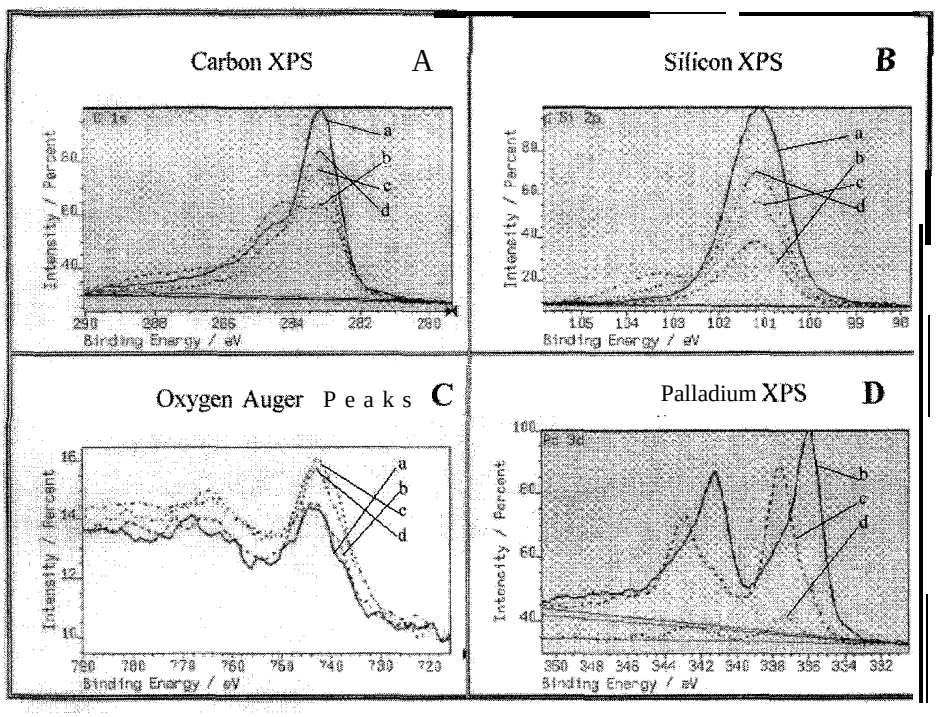


Fig. 1. C, Si, O and Pd Spectra of Different Samples (refer to the text for detail).

Comparison of the spectra of *an* individual element from different samples helped in understanding the impact of annealing at different temperatures to the surface and interface properties of Pd/SiC diodes. Fig. 1 A shows the normalized carbon spectra of samples: the SiC substrate (a), the un-annealed Pd/SiC (b), 300 'C annealed Pd/SiC (c), and 300 'C/ 400 "C annealed Pd/SiC (d). The sample labels were used for the spectra of other elements in Fig. 1 as well. It is clear that the carbon species from SiC was the strongest for the substrate (of course) and smallest for the un-annealed sample. It increased with extending annealing at higher temperatures. This means that SiC was

approaching the surface with the higher temperature annealing, which also indicates the diffusion of the palladium surface layer into the substrate. Annealing also reduced the amount of adventitious carbon on the sample surfaces.

The same conclusion can be drawn from the Si spectra (as shown in Fig. 1B) of the samples. The SiC species behaved the same way as that in the carbon spectra (Fig. 1A). Besides, an oxidation peak showed up and increased steadily with annealing. This oxidation peak was attributed to SiO_2 with some SiO_X mixing possible based on peak deconvolution and identification. As mentioned earlier, this oxidation existed not only on the surface, but also at the interface and maybe within the diffused palladium layer. The oxidation became apparent after annealing at the temperature as low as 300 'C, and that is unusual. It is well known that SiC is quite chemically stable and apparent oxidation usually requires a high temperature (well over 1000 'C). The "early" oxidation may be due to the catalytic nature of palladium.

The intensities of oxygen 1s peaks (mixed with the Pd 3p peaks) with the smallest peak from the SiC substrate increased with annealing in the oxygen spectra (not shown here). Instead of showing oxygen 1s spectra, oxygen Auger spectra were displayed in Fig. 1 C to eliminate the interference of palladium. It can be seen that oxygen signal increased with annealing.

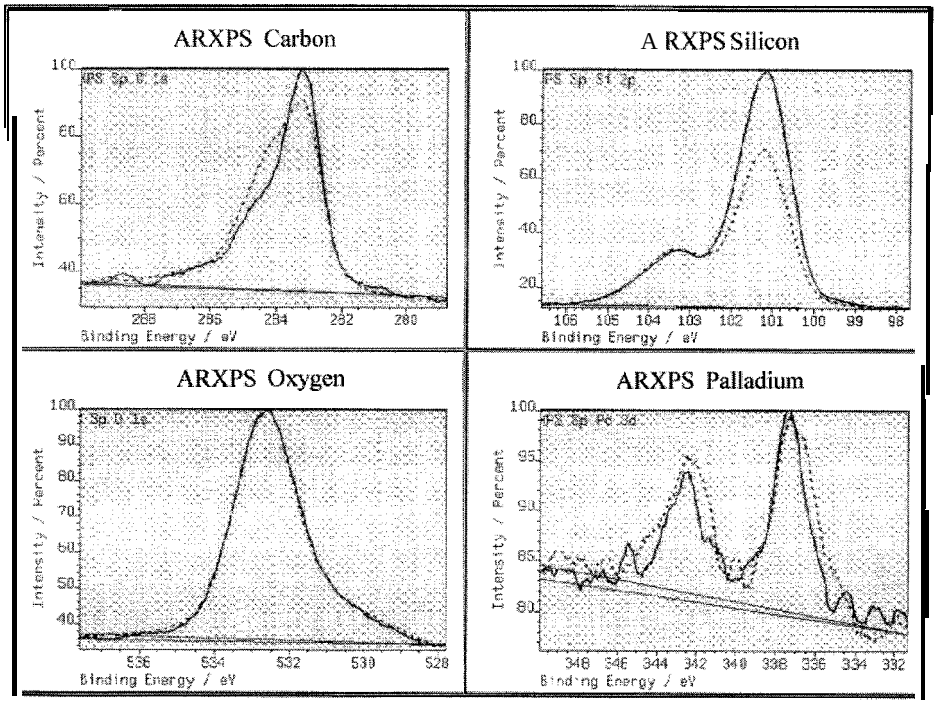


Fig. 2. ARXPS of C, Si, O and Pd Spectra (solid lines---90°, dashed lines---60°)

Palladium spectra had undergone the most dramatic changes (as shown in Fig. 1D). For the as-deposited sample (spectrum b), most of palladium were metallic. After 30 minute annealing at 300 "C the metallic component decreased dramatically, and most of palladium were from palladium compounds with the reduction of overall Pd signals. After another 30 minute annealing at 400 'C, the metallic part completely disappeared and overall Pd had been greatly reduced. Schleich etc. [5] found the similar behavior of palladium in the $Si/SiO_2/Pd$ system. They concluded that Pd atoms from the top palladium film (few nanometer thick) diffused through the SiO_2 (about 2nm thick) layer at elevated temperatures $T \geq 550$ K via oxygen vacancies. As a result, Pd reacted with bulk Si and created a thin layer of Pd_2Si[6] between th, oxide and the underlying silicon substrate. However, no clear silicide signals from the claimed Pd_2Si can be found in the silicon XPS spectra in this study, nor in ref. [5]. Based on ref. [4] and the amount of oxygen, the Pd compound found in this study was assigned to palladium oxides. Further studies is needed to clarify this assignment.

It should be noticed that diffusion of Pd through the Pd/SiC interface into the bulk SiC can be justified from the spectra in Fig. 1 which indicated the diminution of the top Pd layer. Table 2 lists the relative concentrations of selected chemical species.

Table 2. Relative Atomic Concentrations obtained by XPS.

	SiC substrate	unannealed Pd/SiC	annealed Pd/SiC at 300°C	annealed Pd/SiC at 400°C
Pd^d/Si^a	o	0.15	0.030	
Pd^e/Si^a	o	0.093	0.073	0.014
C^b/Si^a	0.81	0.86	0.88	0.96
Si^c/Si^a	0.076	0,17	0.30	0.39

a) Si in SiC, b) C in SiC, c) SiC in SiO_X, d) metallic Pd, and e) Pd compound.

B. Electrical Properties of the Pd/SiC Interface

When a metal comes into a contact with silicon carbide, the respective Fermi levels have to be aligned. The resulting contact can be ohmic or rectifying (also called Schottky) depending on the materials of the contact and fabrication procedures. For palladium and SiC, the Pd/SiC contact is usually rectifying. The electrical behavior of the Pd/SiC contact maybe determined by the Schottky barrier height (SBH or Φ_B). Based on Mott's ideal model[7], the SBH (for an n-type semiconductor) is equal to the difference of the work fiction of the metal (Φ_M) and the electron affinity of the semiconductor (χ_S),

$$\Phi_B = \Phi_M - \chi_S \qquad (1)$$

which is called the Schottky-Mott limit. This relationship implies that SBH for a certain semiconductor can be controlled only by the choice of metal.

However, SBH found in common metal semiconductor contacts is either weakly dependent on or independent of Φ_M[8]. SBH is found strong dependent on the properties of the metal semiconductor interface, that is, dependent on the density of surface states. This leads to the so-called Bardeen limit[9] that states the density of surface states can be so high that they predetermine the Schottky barrier height, such that Φ_B is completely independent of Φ_M. Usually the actual relationship between Φ_B and Φ_M falls somewhere between these limits (pinning)[8].

Experimental measurements of Schottky barrier heights using XPS can be expressed as follows[10]:
For a p-type SiC,

$$\Phi_B = E_f - {}^{SiC}E_{C1s} - ({}^{SiC}E_{C1s} - E_v) \qquad (2)$$

where E_f is the Fermi energy. In XPS the binding energy scale is set to be zero at the sample Fermi energy E_f. ${}^{SiC}E_{C1s}$ is the C(1s) binding energy in SiC at a metal/SiC interface. $({}^{SiC}E_{C1s} - E_v)$ is the difference between C(1s) binding energy and valence-band maximum binding energy in SiC, a bulk material constant. A value of $({}^{SiC}E_{C1s} - E_v) = 281.30$ eV[11] was adopted for the SBH calculation in this study. Waldrop etc. reported similar values of $({}^{SiC}E_{C1s} - E_v)$ as 281.26 eV[12] and 281.45[13].

For an n-type semiconductor,

$$\Phi_B = E_G - E_f \qquad (3)$$

where E_G is the band gap energy. For 6H-SiC, the band gap energy is 3.03 eV[14]. The Fermi energy (E_f) can be measured from Eq.2. The obtained SBHs in this study are also listed in Table 1.

Clearly, the SBH remained unchanged for the Pd/SiC contact after different temperature annealing, although apparent diffusion and chemical reactions had happened at the Pd/SiC interface. This means that after successive annealing at 300 “C and 400 “C, the Pd/SiC diode will still be capable of sensoring hydrogen or hydrocarbons. This conclusion is in consistence with the observation of Chen etc.[3]

4. SUMMARY

The temperature dependence of diffusion and chemical reactions of a Pd/SiC contact have been investigated with XPS/AES in this work. The interracial properties can be directly monitored by employing ultra-thin coating for the Pd/SiC contacts. Chemical reactions are evident even at the room temperature. Silicon oxidation became apparent and palladium oxidation became dominant after 30 minute annealing at 300 ‘C. Silicon oxidation was

found not only on the surface but also at the interracial region after 400 "C annealing. Up to the annealing temperature of 400 "C the rectifying character of the Pd/SiC contact remained unaffected as evidenced by the stable Schottky barrier height. The formation of silicon oxides could degrade the detection sensibility of a Pd/SiC diode to hydrogen as proposed by Chen etc. Diffusion of Pd into SiC was evident after 300 "C annealing, and dramatically increased after a sequential 400°C annealing. This explains the smearing of Pd at the Pd/SiC interface after a extended 425 "C annealing reported by Chen. Further investigations are necessary to clarify the real nature of the Pd compounds created after annealing. This program was supported by NASA through the grants NAGW-2925, NAG3-1430 and NAG3-20 12.

5. REFERENCES

1. H. Dannetun, I. Lundstrom, and L.G. Petersson, *Appl. Phys.* **63**(1), *207(1988).*
2. I. Lundstrom, M. Armgarth and L. Petersson, *CRC Critical Rev. in Solid State and Mater. Sci., 15, 201(1989).*
3. L.Y. Chen, G.W. Hunter, P.G. Neudeck, G. Bansal, J.B. Petit and D. Knight, *J.V.S.T.* **A15**(3), **1228(1997).**
4. NIST XPS, NIST Database 20, National Institute of Standards and Technology
5. B. Schleich, D. Schmeisser and W. Gopel, *Surf. Sci.* **191, 367(1987).**
6. P.J. Grunthaner, F.J. Grunthaner, A. Madhukar and J.W. Mayer, *J.V.S.T.* 19, 649(1 981), *ibid. 20, 680(1982).*
7. N.F. Mott, *Proc. Cambr. Philos. Sot., 34, 568(1938).*
8. L.M. Porter, R.F. Davis, *Mat. Sci. and Eng.*, **B34**, *83(1 995).*
9. J. Bardeen, *Phys. Rev.*, **71, 717(1947).**
10. J.R. Waldrop, *J. Appl. Phys., 75(9), 4548 (1994).*
11. L.M. Porter, R.F. Davis, J.S. Bow, M.J. Kim, R.W. Carpenter, and R.C. Glass, *J. Mater. Res., 10(3), 668 (1995).*
12. J.R. Waldrop, R.W. Grant, Y.C. Wang, and R.F. Davis, *J. Appl. Phys.*, 72(10), 4757 (1992).
13. J.R. Waldrop, R.W. Grant, *Appl. Phys. Lett. 56(6), 557(1990).*
14. Cree Research, Inc., 2810 Meridian Parkway, suite 176, Durham, NC 27713

Optimal Scheduling of a Multi-User, Multi-Tasking Automatic Robotic Telescope

RN. Tantaris, Center for Automated Space Science and Center of Excellence in Information Systems, Tennessee State University, Nashville, TN 37203

ABSTRACT An algorithm is introduced to generate the optimal schedule of observations for a multi-user, multi-tasking automatic robotic telescope. The schedule is optimal m that it will take the highest possible quality data, perform the largest number of tasks, and fairly serve multiple users in terms of data quality, allocated telescope time, and observation priorities. In addition, the algorithm addresses problems due to overloading and underloading the telescope. The algorithm, which employs dynamic-programming-based decision theory, is analytically proven to produce the optimal schedule using set theory. Simulations conducted using actual astronomical data are presented which show significant improvement over existing scheduling techniques. Once completed the new scheduling algorithm will be tested on the Tennessee State University/Harvard-Smithsonian Center for Astrophysics 32-inch automatic Photoelectric Telescope at Fairborn Observatory, Arizona.

Historically, astronomical observations have been conducted by astronomers who travel to a telescope and conduct observations m person, either manually operating the telescope and associated instrumentation themselves or directing assistants to help them. More recently, as electronic communication networks have improved, some astronomers have had the luxury of remaining at their home institutions and overseeing the step by step progress of their programs via their computer workstations connected to the remote observatory by the internet. In either case, the astronomer works throughout the night, checking the quality of the data being taken, monitoring the weather conditions, deciding what observations to make next, and conducting the observations. Today, astronomers at the Tennessee State University (TSU) Center for Automated Space Science are using completely automated telescopes that free them from both the necessity of traveling to a remote observation site and the requirement to manually oversee the observing process. These automatic telescopes (Figure 1) make night-long observations more efficiently, more economically, and more precisely than manual telescopes.

Figure 1: **TSU/Harvard-Smithsonian Center for Astrophysics 32 inch Automatic Robotic** Telescope, **Fairborn** Observatory, **AZ.**

An important step in the development of automatic telescopes was the creation of software techniques to determine the order in which to make the observations. This order, referred to as a schedule, is critical since the quality of the observations (affected by the air mass through which the observation is made) is time dependent. This was no easy task as, considering only 150 observations per night, the astronomer would have to choose from 150! ≅ 5.7× 10262 possible schedules (assuming that all observations are possible the entire night). Currently, the TSU automatic telescope uses the Automatic Telescope Instruction Set (ATIS) protocol [2,3, 10,11,12,13]. ATIS does not determine a schedule a priori. Instead, when an observation is to be made, a set of hierarchical rules are used to winnow the set of all observations down to a single choice. This method, although a huge improvement over manually determining the schedule, is far from optimal. Researchers at NASA Ames Research Center are currently developing a new scheduling technique as part of the Associate Principle Astronomer (APA) project [4,7,8]. The Ames scheduler does determine a schedule m advance. It starts with a schedule generated by ATIS (through

simulation) and searches for a schedule that improves the performance of the telescope. The Ames scheduler also improves on ATIS by incorporating loading and multi-user concerns.

The scheduler presented here will generate the optimal schedule in terms of data quality, loading, the multi-user issues of data quality and allocated telescope time, and observation priorities. In addition, the algorithm will take into account the problems associated with (1) the moon obscuring the observation, (2) multiple requests for the same observation, (3) calibration observations (intentionally poor data quality), (4) observation windows, and (5) time constraints.

Data quality is affected by the mean air mass at which the observation is made. Air mass is a measure of the distance light from an object must travel through the earth's atmosphere. This distance is not fixed. For example, a star near the horizon has a much higher air mass than one that is directly overhead. Therefore, as the star moves through tie sky, the air mass changes (i.e. air mass is a function of time). Figure 2 is an illustration of the air mass curve of a star. In position 1, the distance through the earth's atmosphere (solid portion of the line from the star to the telescope) is much greater than the distance in position 2. Thus the corresponding point on the air mass curve is greater for position 1. Air mass is accurately estimated as the secant of the angle between the star and the point in space directly overhead. Since each observation is made over an interval of time, the mean air mass is used as a measure of data quality. The mean air mass is defined as the mean value of the air mass curve taken over the duration of the observation;

Loading a telescope consists of selecting the specific observations to be made. Typically, more observations are desired than can be performed on a given night (i.e. the telescope is ***overloaded).*** Therefore, there must be some mechanism to select a subset of the requested observations, which can then be scheduled. In the unlikely event that there are not enough observations to fill the night, some of the observations can be duplicated to minimize the idle time of the telescope. Again a subset of the requested observations must be chosen. Each observation m this subset will then be duplicated. In both cases, this subset should change on a nightly basis so that all the observations are made in a uniformly distributed manner. For example, if an observation is not observed tonight, it must be observed tomorrow night.

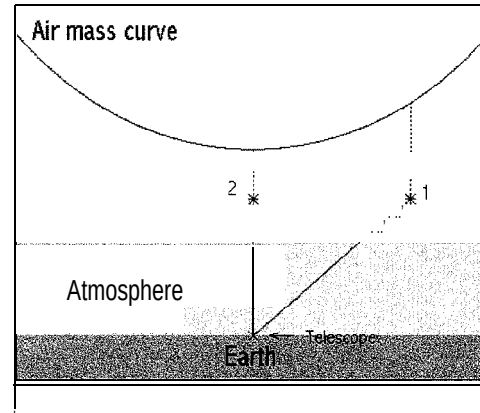


Figure 2: Air mass curve.

Multi-user concerns have arisen as a result of the high cost of modem telescopes. Today multiple institutions combine their resources to purchase a telescope. Thus there must be a method to insure the proper proportion of telescope time for each user and that data quality is comparable. For example, assume that users A and B have purchased a telescope. User A paid 60 percent of the cost and user B paid 40 percent. User A would naturally want 60 percent of the time on the telescope each night (leaving 40 percent for user B). In addition, it would be unfair if all of user A's observations were low quality (high air mass) while user B obtained high quality data (low air mass).

Perhaps the most important issue to be addressed concerns the priority of the observations. The priority of an observation is a measure of the relative importance of the observation to the astronomer. One of the biggest factors influencing priority is the need to make calibration observations in order to standardize the observations. If these calibration observations are not made, the rest of the observations may be meaningless. Therefore, a mechanism to incorporate the differences in priority is essential.

The new algorithm must address several other problems. The first problem is due to the moon. At certain times, the light of the moon may obscure or completely block an observation. The scheduler must insure that this situation does not occur by scheduling the observation at another time during the night. The second problem occurs when a user requests that the telescope make more than one observation of a given star. There must be a way to incorporate this situation m the scheduler. The third problem is that sometimes, particularly for calibration observations, the user intentionally requests observations at high air mass (i.e. poor data quality). Since the algorithm is designed to maximize data quality (i.e. minimize mean air mass), these high air mass observations must be excluded from the optimization process for data quality. Another problem is due to the specification of observation windows. Although an observation maybe possible the entire night, the astronomer may only wish to make the observation during a small portion of the night. This portion is referred to as an observation (or enablement) window. Typically the observation window is used to ensure that the observation is made at or below a

specified air mass. The last two problems involve time. Since the scheduler must produce a (possibly) different schedule every night, the scheduling algorithm must be fast enough to run in the time between the end of one night and the beginning of the next. Also, the schedule will typically break one or more times during a night. The time required to make an observation (the observation duration) is stochastic. Thus if a schedule is generated using specified nominal observation durations, at some time in the night the next scheduled observation may become unobservable. When this occurs, the schedule is said to have broken. When a schedule breaks, anew schedule must be generated for the remainder of the night or the telescope must be synchronized with the current schedule. During the generation of a new schedule (if the first method is used), the telescope will default to the ATIS scheduler until a new schedule is generated. If this occurs, the schedule will not be optimal. Also, since the set of observations is dynamically changing as ATIS makes each new observation it may be difficult to determine the set of observations to be scheduled using the new algorithm.

Both current scheduling methods result m schedules which are not guaranteed to be optimal, thus there is a need for an algorithm to generate the optimal schedule incorporating data quality, multi-user concerns, and loading in a timely fashion. The proposed algorithm is based on dynamic programming [1,5,6]. There are two modifications made to the traditional dynamic programming algorithm

(1) the use of a weighted objective function to determine individual arc costs (cost of moving between one node and another) and

(2) cost comparisons are only made for ***similar*** paths.

The arc cost corresponds to making a particular observation at a specified time. A path is defined as a sequence of connected arcs. This sequence corresponds to a partial schedule. Two paths are similar if they consist of exactly the same observations.

The use of an objective function serves two purposes. First it provides a mechanism to incorporate the time dependent nature of the problem (due to air mass). At each point in time where the arc cost of making a particular observation is needed, the objective function is evaluated at the specified time. Second, it allows for multiple attribute (multi-variable) optimization. The attributes in the objective function are data quality, loading, user allocation, user fairness, and priority. These are incorporated as weighted terms in the objective function. The other problems which must be addressed (moon obscuration, multiple requests, etc.) are handled either by preprocessing or by conditional statements in the algorithm

The use of dynamic programming has made it possible to determine the ***optimal*** schedule. The only other technique currently available to determine the optimal schedule is an exhaustive search. Simulations were conducted to test the algorithm against ATIS and the Ames scheduler using the twelve stars in Table 1. The twelve

Star Index	Catalog Number
1	108076
2	110313
3	1 2 8 4 2 8
4	132142
5	124570
6	106516
7	104988
8	114174
9	115043
10	122742
11	120066
12	124533

Table 1: Twelve test stars.

observations were chosen as a worst case subset (in terms of computational complexity) from the set of stars for the TSU/Harvard-Smithsonian Center for Astrophysics 32-inch Automatic Photoelectric Telescope, Fairborn Observatory, AZ. A simulator maintained at Fairborn generated schedules for ATIS and the Ames scheduler. The proposed algorithm, called the Optimal Schedule Algorithm (OSA), was generated on a 233 MHz Pentium PC. For the purpose of testing, only the air mass attribute of the objective function is considered. Since ATIS does not incorporate multi-user concerns, comparison would not be on an equivalent basis. In addition, the results obtained using all attributes can only be evaluated by the astronomer. The exclusion of other attributes m this test does not in any way affect the results.

The schedules were generated for the night of April 7, 1997. In Figure 3, the air mass curves of the twelve observations are plotted versus local sidereal time (LST). The two vertical lines represent the beginning and end of the night. Table 2 contains the resulting schedules and the corresponding air mass for each of the three scheduling techniques (X denotes mean air mass). The schedules for ATIS, the Ames scheduler, and OSA are shown graphically in Figures 5,4, and 6, respectively. The dotted lines represent the air mass curves for each of the twelve stars. The solid line represents the schedule by following the appropriate air mass carve for each of the twelve observations. The vertical portions of the solid line represent the transitions from one observation to the next. The area under the curve (solid line) is proportional to the total mean air mass for the night. Thus the optimal schedule will have the minimum area under the solid line. The average air mass for tie night is the area under the solid line divided by the number of observations, in this case twelve. Comparison of the areas under the solid line in Figures 4 through 6 is difficult, but comparison can be made for individual observations. In Figure 5, ATIS observes star 12 as the sixth observation. This is the highest air mass observation. In fact, this is the worst possible time to observe 12. Following the curve for observation 12 down to its minimum value in Figure 5, it is obvious that star 12 should be observed last. Comparing Figures 4 through 6, star 12 is observed ninth for Ames and eleventh for OSA, with correspondingly lower air masses for each one. The question now arises that, if the minimum air mass for star 12 is obtained by observing it last, why does OSA observe it eleventh? The answer is that star 3 also has a minimum air mass when observed last. Referring to Table 2, one can see that star 3 is observed eleventh, tenth, and twelfth by ATIS, Ames, and OSA, respectively. Thus the reason OSA does not observe star 12 last is that by observing it second to last and star 3 last, a lower total air mass is achieved.

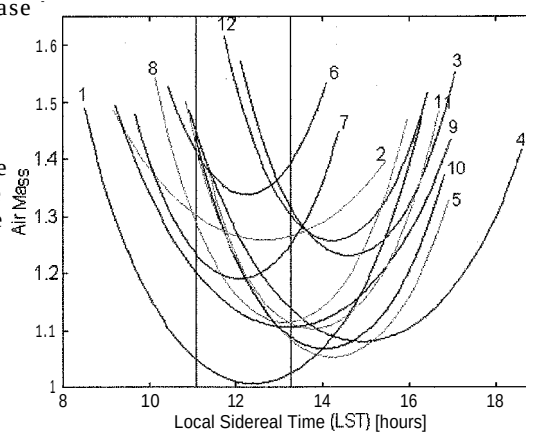


Figure 3: Air mass curves with night indicated.

An overall comparison of the three algorithms is accomplished by comparing the total mean air mass (Total X) for each method in Table 2 ATIS has the highest value at 14.7214, Ames is slightly less at 14.6255, and OSA has the minimum at 14.5239.

Observation Order	Scheduling Algorithms					
	ATIS		Ames		OSA	
	Index	X	Index	X	Index	X
1st	6	1.0137	2	1.0256	2	1.0340
2nd	7	1.2808	7	1.2971	" 1	1.2971
3rd	2	1.3595	I 1	1..3927	I 7	1.3308
4th	8	1.1461	6	1.1609	6	1.2163
5th	1	1.1424	8	1.1650	9	1.1228
6th	12	1.4015	10	1.3541	8	1.3541
7th	11	1.2136	9	1.2136	11	1.2038
8th	10	1.1927	5	1.1733	4	1.1570
9th	5	1.1099	12	1.1250	10	1.1418
10th	9	1.1605	3	1.2111	5	1.1400
11th	3	1.1915	4	1.1176	12	1.1915
12th	4	1.5092	11	1.3893	3	1.3346
Total X		14.7214		14.6255		14.5239
Average X		1.2268		1.2188		1.2103

Table 2: Schedules and resulting air mass.

The Optimal Schedule Algorithm (OSA) performs as expected m that it does generate the optimal schedule. In the simulations conducted, it consistently outperformed both ATIS and the Ames scheduler, In each case, the Ames algorithm took only about 40 seconds to compute a schedule, even though it was allowed as much as ten minutes. The OSA algorithm consistently took about seven minutes to generate an optimal schedule. For comparison, an exhaustive search for the same 12 observations will take more than three months! Another way to judge the efficiency of OSA is to consider the number of times the objective function is called. In the exhaustive

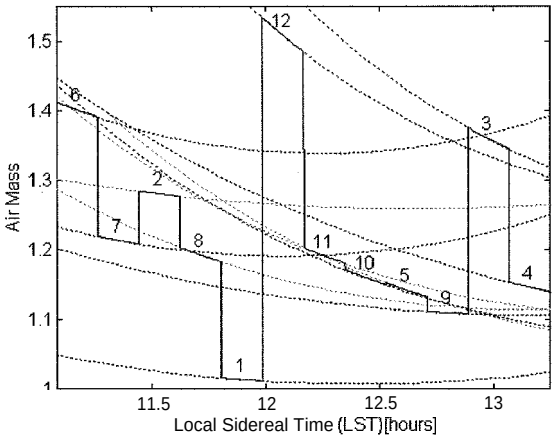


Figure 5: ATIS schedule.

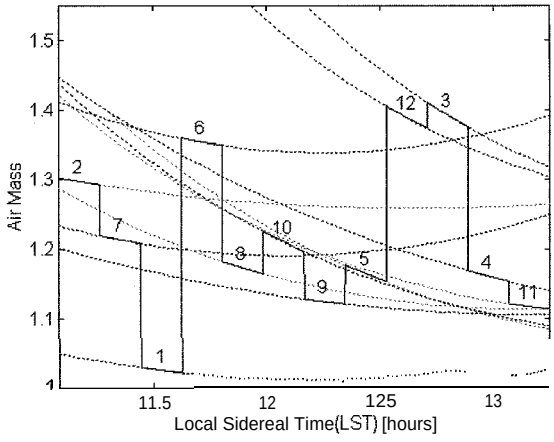


Figure 4: Ames schedule.

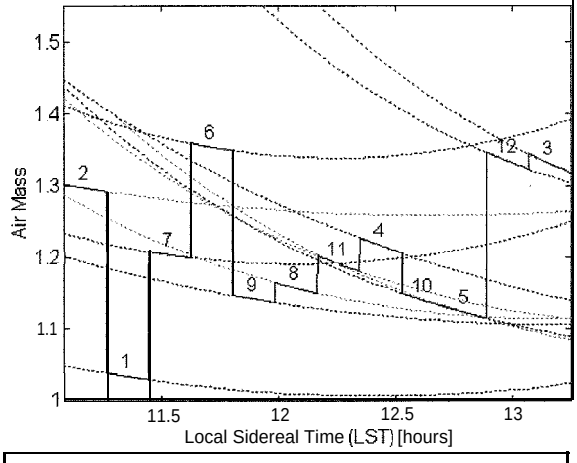


Figure 6: OSA schedule.

search, it will be called over five billion times. The OSA algorithm calls the objective function only about twenty thousand times, an improvement by a factor of 287 million. Data for ATIS and the Ames scheduler are not presented since neither algorithm produces the optimal schedule. Thus comparison would not be on an equivalent basis.

There are, however, still some concerns with OSA. The most significant concern is the speed of the algorithm. For only twelve observations, the algorithm took seven minutes. Since the execution time increases factorially (in tie general case) with a unit increase in the number of observations, the computational time can quickly become too large to be practical. Of course one could increase the computational speed of the hardware, but that increases the cost and there is a limit to how much the speed can be increased. However, there are **several things that can be done to improve the computational time.** First, since the algorithm was written in Matlab, converting the algorithm to C will result in cutting the execution time at least by half Second, the algorithm has not been optimized for computational speed. Thus the application of computer science techniques to optimize the code for speed may significantly reduce execution time. Finally, since the data for the simulations was obtained directly from the ATIS input file, the hour angle enablement windows were adjusted for ATIS. Due to the characteristics of ATIS, these widows are wider than necessary for OSA. The wider the windows, the more observations are enabled at a given level. The factorial increase in computational time is due to the large number of observations enabled at each level. Thus a decrease in the number of enabled observations will result in a factorial decrease in execution time. Another problem with the algorithm is that the weighting of the attributes in the objective function must be adjusted through the experience of the astronomer. This may be a tedious and time-consuming process. Of course the weights only have to be determined once for a given project. This problem maybe solved through the use of artificial neural networks.

The problem of incorporating a time-dependent cost function in an optimization problem has been solved with respect to the telescope scheduling problem. However, this algorithm may be applied to any discrete event system where the cost function is a function of time. In the past, time-dependent cost functions have been converted by approximation to a single cost value. Using this algorithm eliminates the need to do so. As an example, consider a supermarket checkout line. Here the cost function represents the customer's satisfaction with service. It is assumed that the customer's satisfaction is a fixed quantity. However, the longer the customer waits in line, the lower the satisfaction of the customer. Thus the cost function is time dependent. This algorithm allows this cost-

dependency to be incorporated into the optimization process. Further details (including a proof of optimality) on the Optimal Schedule Algorithm can be found in [15].

This research was supported by NASA grants NCC5-228, NGT40021, and NCC2-977. Additional funding was provided by the Center of Excellence in Information Systems; Dr. M. Busby, Director. This research was accomplished under the guidance of Dr. L. Keel and Dr. S. Sathananthan. Mr. G. Henry's work with automatic telescopes provided the inspiration for this project. The astronomical background for this work was provided by Mr. G. Henry and Dr. J. Eaton. Mr. D. Epand's help was essential in conducting the simulations. Much of the approach taken in this research (especially concerning the objective function attributes) was based on the work of Mr. J. Bresina and Mr. W. Edgington.

REFERENCES

(1) Bellman, R., Dynamic Programming, New Jersey: Princeton University Press, 1957

(2) Boyd, L., Epand, D., et al., "Automatic Telescope Instruction Set 1993", IAPPP Communication No. 52, Summer 1993.

(3) Boyd, L., Genet, R, and Hall, D., "APT's: Automatic Photoelectric Telescopes", Sky and Telescope, July, 1985.

(4) Bresina, J., Drummond, M., et al., "Automated Management and Scheduling of Remote Automatic Telescopes", Optical Astronomy from the Earth and Moon Astronomical Society of the Pacific, 1994.

(5) Denardo, E., Dynamic Programming: Models and Applications, New Jersey Prentice Hall, 1982.

(6) Dreyfus, S., Dynamic Programming and the Calculus of Variations, New York Academic Press, 1965

(7) Drummond, M., Bresina, J., et al., "The Associate Principle Astronomer Telescope Operations Model", Proceedings of the Third International Symposium on Artificial Intelligence, Robotics, and Automation for Space, Oct. 18-20, 1994, JPL publication 94-23.

(8) Edgington, W., Drummond, M., et al., "Improved Scheduling of Robotic Telescopes", New Observing Modes For The Next Century, Astronomical Society of the Pacific, 1996.

(9) Hall, D. and Genet, R., Photoelectric Photometry of Variable Stars. 2nd. cd., Virginia: Willmann-Bell, Inc., 1988.

(1 O) Henry, G., "ATIS Dispatch Scheduling of Robotic Telescopes", New Observing Modes For The Next Century, Astronomical Society of the Pacific, 1996.

(1 l) Henry, G., "The Fairborn/TSU Robotic Telescope Operations Model", Telescopes: Current Capabilities, Present Developments, and Future Prospects for Automated Astronomy, Astronomical Society of the Pacific, 1995.

(12) Henry, G., "The Development of Precision Robotic Photometry", Robotic Telescopes: Current Capabilities, Present Developments. and Future Prospects for Automated Astronomy, Astronomical Society of the Pacific, 1995.

(13) Henry, G. and Hall, D., "ATIS and the Evolution of Automatic Telescopes", IAPPP Communication No. 52, summer 1993.

(14) Paxachoff, J. and Kutner, M., University Astronomy, Philadelphia: W. B. Saunders Company, 1978.

(15) Tantaris, R. N., Development of an Optimal Scheduling Algorithm for a Multi-User, Multi-Tasking Automatic Telescope, Tennessee State University Master's Thesis, Nashville, TN, 1997.

Title: Multi-Objective Decision Making and the Motion Control of an Autonomous Mobile Robot

Earl B. Smith' and Reza Langari
Department of Mechanical Engineering, Texas A&M University

1.0 ABSTRACT

When a mobile robot is moving through an unstructured site, a premium must be put on it avoiding obstacles. In addition to that the robot must move efficiently through the site. A teleoperated robot can be used to control its velocity through a site. However, this can be very fatiguing, especially mentally, to a person. Therefore, researchers have been working more with autonomous mobile robots. In fuzzy logic IF-THEN rules can be used where the antecedent is the distance the robot is from the obstacles. The consequent would be the speed of the robot. In this project, multi-objective decision making is being explored to obtain the speed. Multi-objective decision making has several alternatives or decisions where each alternative satisfies an objective to a certain degree, and each objective has a preference. The alternatives, for this case, will represent difference speeds that the robot is to move. The process of selecting the speed, via multi-objective decision making, will commence when the optimum path is selected. The optimum path will also be selected by use of multi objective decision making. The process in selecting the speed is to be used when the vehicle has chosen a path while in the presence of multiple obstacles. Fuzzy logic control will be used, along with multi objective decision making, in getting to the desired speed.

2.0 INTRODUCTION

When someone drives through traffic they usually want to move it as quickly as possible. However, one must obey the traffic laws and observe the traffic conditions, so they can get to their destination as safely as possible. The way a person drives will be dictated by how heavy or light the traffic is. In other words the lighter the traffic, there will be a greater chance that the person will drive at a faster speed and vice versa.

In this research a few of the same techniques will be used to select the appropriate speed of a mobile robot when it is in an unstructured site. The speed of the robot will be dependent on how an obstacle(or obstacles) is from the robot. If a pair of obstacles is close to the robot, then it will be desired that the robot will move slower and vice versa.

Research have used mobile robots in unstructured environment when it pertained to navigation, surveying, and the collection of samples(see Tunstel [1]). In order to move through the site, they, for the most part, have used goal seeking, collision avoidance and motion control techniques.

For instance Wong, Kortenkamp, and Speich[6] used a neural network architecture on a mobile robot where it would find individuals then locate and recognize their faces. Ruspini, Saffioti, and Konolige[5] developed a fuzzy controller for a mobile robot that has, among other things, blended multiple goals, Their work involved the motion control, solving of complex tasks, and mapping functions of a mobile robot.

This paper will discuss how multi-objective decision making, a technique out of fuzzy logic, will be used to find the appropriate speed when a mobile robot is in the presence of multiple obstacles. The multi-objective decision making process for motion control will work in conjunction with the decision making process for navigation. Once the decision making process for navigation ends, the decision making process for motion control will begin.

3.0 MULTI-OBJECTIVE DECISION MAKING

A simple decision can be made to satisfy a single objective. One decision could be made to satisfy objectives such as saving money, decreasing time, or conserving energy. However, a single decision can effect more than one objective where each objective is satisfied to a certain degree. For example when a car is manufactured, the objectives could be to increase safety and reliability, and reduce cost. In this case let the increased safety and reliability be the most important objective. Then in the multi-objective decision making process(for instance see Ross[11]), functions will be set up to where the decision or alternative would be chosen to satisfy a certain objective to a certain degree within the interval [0, 1]. The objectives would be assigned weights of importance or preferences.

During the multi-objective decision making process, there will be a set of alternatives, $A=\{a_1, a_2, \ldots a,\}$ or decisions to make. Also there also exist a set of objectives, $O=\{O_1, 0_2, \ldots . O_r\}$, that the decision making

[1] NASA ACE Fellow

process must consider. Each alternative will satisfy each objective to a certain degree, and the objective is a fuzzy set in the form

$$O_i = \{\frac{\mu_1}{a_1} + \frac{\mu_2}{a_2} + \ldots\ldots + \frac{\mu_n}{an}\}, \quad (3\text{-}1)$$

where a_i represents the alternative and μ_i is the degree to which that alternative satisfies that objective. The decision maker establishes a weight of importance or preference, $P = \{b_1, b_2, \ldots b.\}$, to each objective. Note that there is the same number of preferences and objectives. Preferences can be linguistic values like many, low, or perfect or intervals such as [0, 1], [-1,1] or [1, 10].

To get a decision function that satisfies all of the objectives one can use the decision function for each alternative

$$D = O_1 \cap O_2 \cap \ldots\ldots \cap O_r \quad (3\text{-}2)$$

or

$$\mu_D(a) = \min[\mu_{O1}(a), \mu_{O2}(a), \ldots\ldots \mu_{Or}(a)] \quad (3\text{-}3)$$

To get the optimum decision, *, we use the following function

$$\mu_D(a^*) = \max(\mu_D(a)). \quad (3\text{-}4)$$

When the objective is computed with its respected preference we have

$$D = M(O_1, b_1) \cap M(O_2, b_2) \cap \ldots\ldots \cap M(O_r, b_r) \quad (3\text{-}5)$$

where $M(O_i, b_i)$ is the decision measure involving the objective and preference. If there are two or more alternatives with the same decision value, then a tie breaking procedure is conducted by taking the next highest decision value for each alternative and having their respected values compared. To evaluate equation (3-5) one would have to relate the objective, O_i, and the preference, b_i, and at the same time include negation. That exercise can be done by using the classical implication operator for a particular alternative a in the form

$$M(O_i, b_i) = b_i \rightarrow O_i(a) = \neg b_i \vee O_i(a). \quad (3\text{-}6)$$

Therefore (3-5) can be rewritten for a particular alternative as

$$D(a_i) = \min[\max\{\mu_{\neg b1}(a_i), \mu_1(a_i)\}, \ldots\ldots \max\{\mu_{\neg br}(a_i), \mu_{Or}(a_i)\}]. \quad (3\text{-}7)$$

To find the optimum decision we have

$$D^* = \max\{D(a_1), D(a_2), \ldots\ldots, D(a_r)\}. \quad (3\text{-}8)$$

If there exists a tie between alternatives($D(a_i) = D(a)$), then a tie breaking procedure(see Ross) is conducted to find the optimum decision. Here we have $C_k(a_i) = D(a_i)$ and $C_g(a_j) = D(a_j)$. Then we proceed to find the optimum decision by

$$\wedge D(a_i) = \min[C_i(a_i)] \text{ and } \wedge D(a_j) = \min[C_i(a_j)] \quad (3\text{-}9)$$

where i # k and j #g. If $\wedge D(a_i) > \wedge D(a_j)$ or vice versa, then the procedure is finished. If it is not finished, then the procedure continues.

4.0 MULTI-OBJECTIVE DECISION MAKING AND MOTION CONTROL

When the multi-objective decision making process for navigation ends, a path(for the robot to travel) has been selected and a steering direction, in degrees, has been chosen. The path that is chosen for the robot to move, will be between two obstacles. The distances from the obstacles to the robot will be used to help determine the appropriate speed for the vehicle to move. Algorithms have been created to where there are membership functions that cover different distances. In other words one algorithm uses membership functions for distances greater than it does in the other. However, normalized distances will be used when creating and using the membership functions. Here the x-axis on the membership function plot will be dimensionalless. With the distances normalized, one will be able to use the membership functions in all cases when regarding distance.

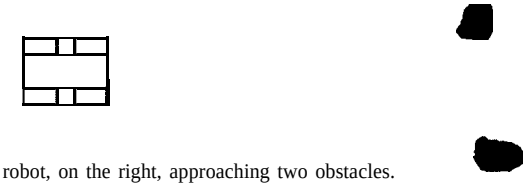


Figure **one.** Mobile robot, on the right, approaching two obstacles.

The above figure shows an example of when the multi-objective decision process for motion control will begin. Once the distances to the obstacles have been determined, the robot will be able to begin to select an appropriate speed.

The objectives fuzzy sets, for motion control, have pertained mainly to an individual speed. For example if an objective happens to be for the fastest speed, then the alternative for the fastest speed will be set to a degree of

one. The alternative for the slowest speed, in this fuzzy set, will have the lowest degree. Various distributions, rectangular, triangular, etc., will be used to find the appropriate values for the objective fuzzy sets. Throughout the design making process, the values in the objective fuzzy sets have remained constant.

While the objectives have remained constant, the preferences have varied. The preference will vary according to which obstacle is closest to the robot. Each element in the preference matrix will be an equation, and that equation will vary depending on the desired speed The variable in each equation will be the distance from the robot to the obstacle. This is an example where the mobile robot will be more involved in finding the speed while it is in operation. Also, the human operator will become less involved.

5.0 FUTURE WORK

So far the focus has been on ultrasonic sensors. However, other types of sensors(see Everett[3] and Jones and Flynn[8]) will be examined in order to find their affect on the mobile robot's speed and their overall usefulness in this work.

Within the multi-objective decision process, the number of alternatives and objectives will be varied. This will be done to see what will be the appropriate number of alternatives and objectives to run the decision making process. The results from using multi-objective decision making will be compared to using fuzzy IF-THEN rules to find the speed.

Another concern will be the acceleration of the vehicle. For instance when the obstacle is at a vary far distance away from the obstacles, the desired speed will be very fast. But when the vehicle comes in close proximity to the obstacles, the acceleration issue will come into effect. Also the angle deviation and the width of the path are of a concern. The sharper the angle and the narrower the path could be a determining factor in the speed and acceleration of the vehicle. By being concerned about these issues, the mobile robot will be made more efficient when it is maneuvering around the site.

6.0 REFERENCES:

1. Tunstel, Eddie, "Autonomous Robot Mapping and Behavior-based Navigation", **JPL Technical Document** D-10458, December, 1992.
2. Tunstel, E., Lippincott, T., & Jamshidi, M., "Introduction to Fuzzy Logic With Application to Mobile Robotics", *Proc. of the First National Students Conference of the National Alliance of NASA University Research Centers, NC* A&T State Univ., Greensboro, NC, March 1996.
3. Everett, H. R., *Sensors for Mobile Robots: Theory and Application,* A.K. Peters. Ltd., Wellesley, Mass. 1995.
4. Ross, Timothy J., *Fuzzy Logic with Engineering Applications,* McGraw-Hill, Inc. 1995.
5. Ruspini, E.H., Saffiotti, A., & Konolige, K., "Progress in Research on Autonomous Vehicle Motion", Industrial Application and Fuzzy Logic and Intelligent Systems, IEEE Press, 1995.
6. Wong, C., Kortenkamp, D., & Speich, M., "A Mobile Robot That Recognizes People", 1995 *IEEE International Conference on Tools with Artificial Intelligence, 1995.*
7. Yen, J., & Langari, R., *Fuzzy Logic and Control Theory and Practice,* Prentice-Hall, 1997.
8. Jones, Joseph L. & Flynn, Anita M., *Mobile Robots: Inspiration to Implementation,* A.K. Peters, Wellesley, MA, 1993.

Title: Multi-Objective Decision Making and the Navigation of an Autonomous Mobile Robot

Earl B. Smith' and Reza Langari, Ph.D.
Department of Mechanical Engineering, Texas A&M University

1.0 ABSTRACT

Mobile Robots have been used in working in unstructured environments like hazardous waste sites and lunar surfaces. Advantages of using robots are that humans will not be exposed to the harsh conditions and the use of them is cheaper. Teleoperated robots have been used where the operator controls every movement of the robot. However, the method can be fatiguing to a person. Therefore, researchers have been working on building autonomous mobile robots, so a person will have as little interaction with the robot as possible. In order for a robot to move through a site, a premium must be put on it avoiding obstacles. In fuzzy logic membership functions and the linguistic terms have been used in making a robot autonomous by use of IF-THEN rules. Fuzzy decision making can give the robot another tool for it to be autonomous. For this project, multi-objective decision making is being explored for the optimum path selection. Multi-objective decision making has several decisions or alternatives where each alternative will accomplish a particular objective to a certain degree, and each objective has a preference. At the end of the process, the alternative with the optimum value will be chosen.

2.0 INTRODUCTION

An unstructured or non-engineered environment is an environment in which humans have little to no apriori knowledge. These environments can include hazardous waste sites and lunar surfaces. Hazardous waste sites can include(or have formed) leaky drums, radioactive substances and concrete like substances as explained in reports by Pope[5], Kiernan[6], and Davidson[7]. Lunar surfaces include surfaces like the ones on the moon or other planets in the solar system.

At times humans have entered radioactive environments(while wearing protective gear). Yet, it can be very risky for them to enter and working in such an environment can be very fatiguing. This does not take in consideration the extreme temperature on the moon or other planets.

With the extreme conditions that exist, agencies, companies and universities have research on the use of autonomous mobile robots in the unstructured environments as explained by Tunstel[1], Pope[13], and Anderson[16]. The advantages of using these types of robots are that they can go into environments that are hazardous or extreme, they will not become fatigued like humans, and the cost of the operation would be less.

In order for a mobile robot to effectively work at a site, it must avoid the obstacles that exist. By doing so the robot will be able to complete its work more efficiently, possibly spend less time at the site, and reduce the chance of it, the robot, being damaged.

In this paper the emphasis will be on a mobile robot navigation its way through the site. Multi-Objective decision making, a technique out of fuzzy logic, will be examined on how to aid in the navigation of a mobile robot through such a site.

3.0 MULTI-OBJECTIVE DECISION MAKING

A simple decision can be made to satisfy a single objective. One decision could be made to satisfy objectives such as saving money, decreasing time, or conserving energy. However, a single decision can effect more than one objective where each objective is satisfied to a certain degree. For example when a car is manufactured, the objectives could be to increase safety and reliability, and reduce cost. In this case let the increased safety and reliability be the most important objective. Then in the multi-objective decision making process(for instance see Ross[4]), functions will be set up to where the decision or alternative would be chosen to satisfy a certain objective to a certain degree within an interval of [0, 1]. The objectives would be assigned weights of importance or preferences.

During the multi-objective decision making process, there will be a set of alternatives, $A=\{a_1, a_2, \ldots . a_r\}$ or decisions to make. Also, there exist a set of objectives, $O=\{O_1, O_2, \ldots . 0,\}$, that the decision making process must consider. Each alternative will satisfy each objective to a certain degree, and the objective is a fuzzy set in the form

[1] NASA ACE Fellow

$$O_i = \{ \frac{\mu_1}{a_1} + \frac{\mu_2}{a_2} + \ldots\ldots + \frac{\mu_n}{a_n} \}, \quad (3\text{-}1)$$

where a_i represents the alternative and μ_i is the degree to which that alternative satisfies that objective. The decision maker establishes a weight of importance or preference, P = {b,, b_2, ... b,}, to each objective. Note that there is the same number of preferences and objectives. Preferences can be linguistic values like many, low, or perfect or intervals such as [0, 1], [-1,1] or [1, 10].

To get a decision function that satisfies all of the objectives one can use the decision function for each alternative

$$D = O_1 \cap O_2 \cap \ldots\ldots \cap O_r \quad (3\text{-}2)$$

or

$$\mu_D(a) = \min[\mu_{O1}(a), \mu_{O2}(a), \ldots\ldots \mu_{Or}(a)] \quad (3\text{-}3)$$

To get the optimum decision, a*, we use the following function

$$\mu_D(a^*) = \max(\mu_D(a)). \quad (3\text{-}4)$$

When the objective is computed with its respected preference we have

$$D = M(O_1, b_1) \cap M(O_2, b_2) \cap \ldots\ldots \cap M(O_r, b_r) \quad (3\text{-}5)$$

where $M(O_i, b_i)$ is the decision measure involving the objective and preference. If there are two or more alternatives with the same decision value, then a tie breaking procedure is conducted by taking the next highest decision value for each alternative and having their respected values compared. To evaluate equation (3-5) one would have to relate the objective, O_i, and the preference, b_i, and at the same time include negation. That exercise can be done by using the classical implication operator for a particular alternative a in the form

$$M(O_i, b_i) = b_i \rightarrow O_i(a) = \neg b_i \vee O_i(a). \quad (3\text{-}6)$$

Therefore (3-5) can be rewritten for a particular alternative as

$$D(a_i) = \min[\max\{\mu_{\neg b1}(a_i), \mu_1(a_i)\}, \ldots\ldots \max\{\mu_{\neg br}(a_i), \mu_{Or}(a_i)\}]. \quad (3\text{-}7)$$

To find the optimum decision we have

$$D^* = \max\{D(a_1), D(a_2), \ldots\ldots, D(a_r)\}. \quad (3\text{-}8)$$

If there exists a tie between alternatives($D(a_i) = D(a_j)$), then a tie breaking procedure(see Ross) is conducted to find the optimum decision. Here we have $C_k(a_i) = D(a_i)$ and $C_g(a_j) = D(a_j)$. Then we proceed to find the optimum decision by

$$\wedge D(a_i) = \min[C_i(a_i)] \quad \text{and} \quad \wedge D(a_j) = \min[C_i(a_j)] \quad (3\text{-}9)$$

where i # k and j # g. If $\wedge D(a_i) > \wedge D(a_j)$ or vice versa, then the procedure is finished. If it is not finished, then the procedure continues.

4.0 MULTI-OBJECTIVE DECISION MAKING ON THE MOBILE ROBOT

The input to the entire system will come from the data generated by the sensors(see Tunstel[2] and Everett[3]). For now the data are the lengths from each sensor to its corresponding obstacle. The sensors are to be ultrasonic and placed on the front of the vehicle. Algorithms have been created to where they are three four, and five front sensors. Sensor one will be on the left and sensor three, four or five(depending on the number of sensors) will be on the left.

Once the signal is received and the distance to the obstacle calculated, the distance between each obstacle is calculated using the law of cosines

$$c^2 = a^2 + b^2 - 2ab\cos\theta. \quad (4\text{-}1)$$

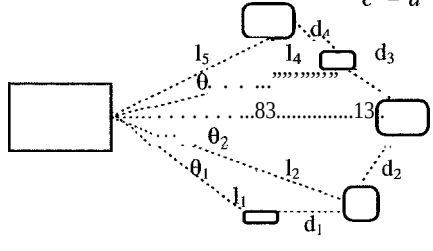


Figure One. Vehicle, on the left, approaching a group of obstacles.

Here *c* represents the distances(d_1, d_2, d_3, and d_4) between the obstacles or the widths of the paths. The distances between the robot and obstacles are represented by *a* and *b. The* angle θ is the angle between a pair of adjacent sensors and is represented by $\theta_1, \theta_2, \theta_3$, and θ_4 in the figure. The distances between the obstacle and

sensor, a and b, are represented by 1, and l_2 for θ_1, 1_2 and l_3 for θ_2, 1_3 and 1_4 for θ_3, and 1_4 and l_5 for 13_4. The angles between the sensors are to remain at 30° for this particular setup.

Note that the individual paths serve as the alternatives or decisions to be made. Each of them will accomplish an objective to a certain degree.

Once the path's widths are determined, there are ranked from widest to narrowest. When that process is complete, the objective fuzzy sets and preferences are set. Then the multi-objective decision making process for navigation will begin. Once the alternative with the optimum decision is found, the robot will move down the path corresponding to that particular alternative. The alternative in the algorithm is the steering direction angle. Therefore when a path is selected, the robot will rotate the selected number of degrees and then head into that direction.

Figure two shows the plot of the direction the mobile robot steers when a path is selected. The x-axis represents the instance the robot comes upon a group of obstacles. These results are for when the mobile robot has five front sensors. The values of 15° and -15° represent when the mobile robot moves along the straightest path.

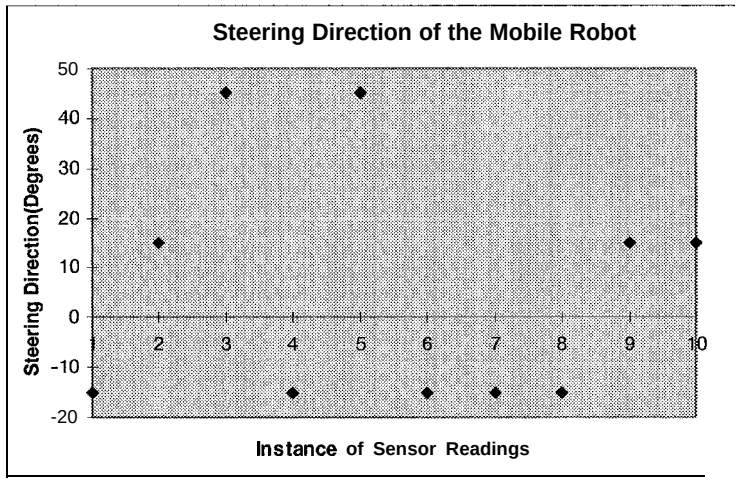


Figure Two. Plot of the Steering Direction when a path is selected.

5.0 FUTURE WORK

In the work until now the optimum path that was chosen was the widest one-regardless on whether it was the straightest one or not. Work is being focused on getting the decision making process will chose, for the most part, the straightest path(see figure two). The objective here is to chose the straightest path as much as possible or when it is wide enough. Here the robot will be moving through a group of multiple obstacles by using a minimum amount of distance. Therefore less time will be used to move pass the obstacles and arrive at the goal position.

In addition to being concerned about the location of the obstacles, one must be concerned about the variations in the terrain. This mainly pertains to the elevations, peaks and valley, and the vibrations that may be caused by the mobile robot moving through the site.

6.0 **REFERENCES**

1. Tunstel, Eddie, "Autonomous Robot Mapping and Behavior-based Navigation", JPL Technical Document D-10458, December, 1992.
2. Tunstel, E., Lippincott, T., & Jamshidi, M., "Introduction to Fuzzy Logic With Application to Mobile Robotics", Proc. of the First National Students Conference of the National Alliance of NASA University Research Centers, NC A&T State Univ., Greensboro, NC, March 1996.
3. Everett, H.R., ***Sensors for*** *Mobile Robots:* ***Theory and Application,*** A.K. Peters. Ltd., Wellesley, Mass. 1995.
4. Ross, Timothy J., ***Fuzzy Logic with Engineering Applications,*** McGraw-Hill, Inc. 1995.
5. Pope, Greg, "Shredding robots clean up a witches brew," ***Popular Mechanics,*** Vol. 11, p. 17, November, 1993.
6. Kiernan, Vincent, "Send in the Robots," New *Scientist*, Vol. 147, pp. 34-37, September 23, 1995.

7. Davidson, Clive, "Robots: the next generation," New *Scientist,* Vol. 145, pp. 32-34, January 14, 1995.
8. Anderson, Mary Rose, "Ecological Robots", Technology Review, Vol. 95, pp. 22-23, January, 1992.
9. Yen, J., & Langari, R., *Fuzzy Logic and Control Theory and Practice,* Prentice-Hall, 1997.

98 URC028

MODAL PARAMETER IDENTIFICATION OF A FLEXIBLE ARM SYSTEM

Jason Barrington, Jiann-Shiun Lew, Edward Korbieh, Montanez Wade,
and Richard Tantaris
NASA Center for Automated Space Science
Tennessee State University
Nashville, TN 37203-3401

ABSTRACT

In this paper an experiment is designed for the modal parameter identification of a flexible arm system. This experiment uses a function generator to provide input signal and an oscilloscope to save input and output response data. For each vibrational mode, many sets of sine-wave inputs with frequencies close to the natural frequency of the arm system are used to excite the vibration of this mode. Then a least-squares technique is used to analyze the experimental input/output data to obtain the identified parameters for this mode. The identified results are compared with the analytical model obtained by applying finite element analysis.

KEYWORDS: Modal parameter identification, Least-squares, Finite element analysis, Flexible arm experiment.

INTRODUCTION

The system identification of large space structures is an important and challenging task [1,2], since the structural dynamics of the real system maybe different from the analytical model, which can be obtained by applying finite element analysis. To maintain the performance and safe operation of the system, it is necessary to model the system accurately.

The accuracy of the identified parameters depends on the equipments, experimental setup, and the chosen input-signal. Various kinds of inputs, such as random input and sine-sweep input, can be used to excite the vibration of structures for system identification. In this paper, an experiment is designed for modal parameter identification of a flexible arm system. Here a function generator is used to generate many sets of sine-wave signals with frequencies close to the natural frequency of the system to excite the vibration of each mode. A least-squares method [3] is used to process these sets of input/output response data to obtain the identified parameters of each mode. The identified results are compared with the analytical model obtained by using finite element analysis [4].

EXPERIMENTAL PROCEDURE

A flexible arm system as shown in Figure 1 is the test apparatus for study. A motor mounted on the top of the system is used to generate the torque force to excite the vibration of the beam. This aluminum beam is pinned to a shaft. The length, height, and thickness of the beam are 0.504,

0.0381, and 0.00159 meters, respectively. An accelerometer at the tip of the beam is used to measure the vibration of the beam.

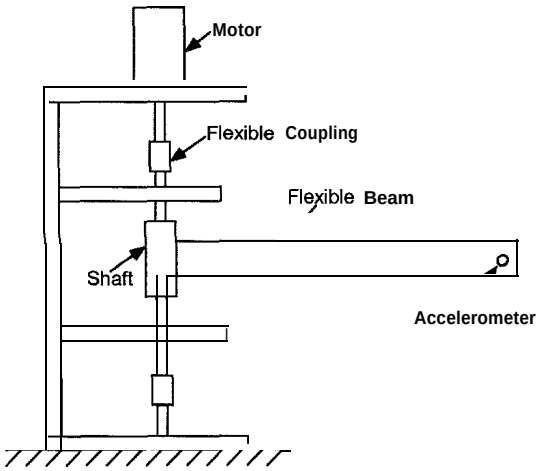


Figure 1: Flexible arm system.

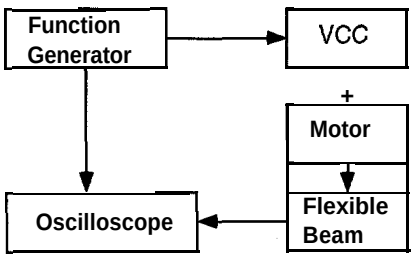


Figure 2: Experimental setup.

Figure 2 shows the experimental setup for the modal parameter identification. A function generator is used to provide sine-wave input signal. The input voltage signal is converted to the current signal through a VCC (Voltage-Current Converter). A motor transfers current to torque force to move the flexible beam. An accelerometer at the tip of the beam is used to measure the vibration of the beam. Both input signal and acceleration output signal are saved in an oscilloscope. Then the saved data are transferred to a Pentium computer for analysis.

LEAST-SQUARES APPROACH

This section presents a least-squares technique used to obtain the modal parameters. The differential equation of a single-degree spring-mass system is

$$m\ddot{x} + kx = u \qquad (1)$$

where *m is* mass, *k is* spring constant, *u is* input force, and x is displacement. The preceding equation can be written as

$$\ddot{x} + \omega^2 x = au \tag{2}$$

with

$$\omega = \sqrt{\frac{k}{m}}, \qquad a = \frac{1}{m} \tag{3}$$

where *a is* the amplitude corresponding to the natural frequency ω. The vibration of each mode of the flexible beam system corresponds to a differential equation as Eq. (2) with its natural frequency ω. In this experiment, many sine-wave inputs with frequencies close to the natural frequency are used to excite each mode for the identification of parameters ω^2 and a of this mode. For the given sine-wave input $u(= A_i \sin(\omega_i t))$ with amplitude A_i and frequency ω_i, the displacement x in Eq. (2) is $B_i \sin(\omega_i t)$. Substituting $u = A_i \sin(\omega_i t)$ and x = $B_i \sin(\omega_i t)$ into Eq. (2) yields

$$(\omega^2 - \omega_i^2) B_i = aA_i \tag{4}$$

$$\Rightarrow \frac{B_i}{A_i} = \frac{a}{\omega^2 - \omega_i^2} \tag{5}$$

In this experiment, the measurement is acceleration

$$\ddot{x} = -\omega_i^2 B_i \sin(\omega_i t) = C_i \sin(\omega_i t) \tag{6}$$

with amplitude

$$C_i = -\omega_i^2 B_i \tag{7}$$

The amplitude ratio between acceleration output and torque input is

$$D_i = \frac{C_i}{A_i} = \frac{-\omega_i^2 B_i}{A_i} = \frac{-a\omega_i^2}{\omega^2 - \omega_i^2} \tag{8}$$

From the experimental data, we can obtain the amplitude ratio D_i for each frequency ω_i. Then we can form the cost function for a linear least-squares problem with the unknown variables a and ω^2 as

$$J = \sum_{i=1}^{n} (D_i \omega^2 + \omega_i^2 a - D_i \omega_i^2)^2 \tag{9}$$

The least-squares problem is to find the values of a and ω^2 to minimize the cost function J. *Since J* is a quadratic function of a and ω^2, a unique solution can be obtained by solving [3]

$$\frac{\partial J}{\partial a} = 0, \qquad \frac{\partial J}{\partial \omega^2} = 0 \tag{lo}$$

The transfer function between input u and displacement x can be obtained by taking the Laplace transform of Eq. (2) as

$$\frac{\mathrm{x(s)}}{\mathrm{u(s)}} = \frac{a}{s^2 + \omega^2} \tag{11}$$

The transfer function for the acceleration measurement is

$$\frac{\ddot{x}(s)}{u(s)} = \frac{as^2}{S^* + \omega^2} \tag{12}$$

RESULTS

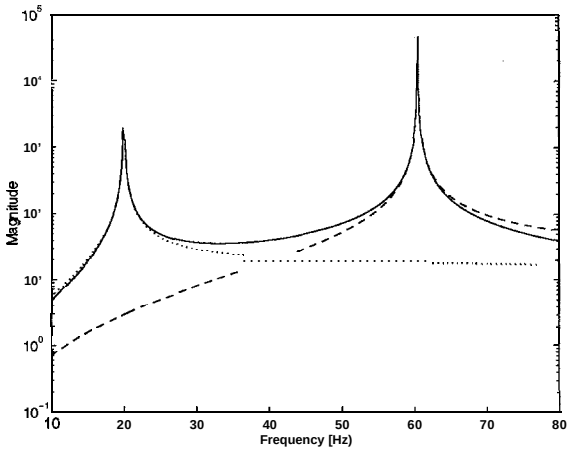


Figure 3: Transfer function of finite element model.

In this experiment, we present the results of the first two flexible modes. Figure 3 shows the magnitude of the transfer functions of the first two modes of the finite element model with torque input and acceleration output. The solid, dotted, and dashed lines are the transfer functions of the two modes, the frost mode and the second mode, separately. Mode 2 has little effect on mode 1 when the frequency is around the first natural frequency 20.08 Hz, and mode 1 has little effect on mode 2 when the frequency is around the second natural frequency 60.58 Hz. This implies that we can identify each mode separately. Table 1 shows the identified parameters, which are obtained by applying the preceding least-squares technique to analyze the experimental data, and the parameters of finite element model.

Table 1: Identified parameters and parameters of finite element model.

	1st Mode	1st Mode	2nd Mode	2nd Mode
		ω (Hz)	a	ω (Hz)
FEM	3.1:8	20.08	-4.769	60.58
ID	2.915	19.88	-4.344	59.82
|ID-FEM%|	7.70	1.04	8.91	1.26

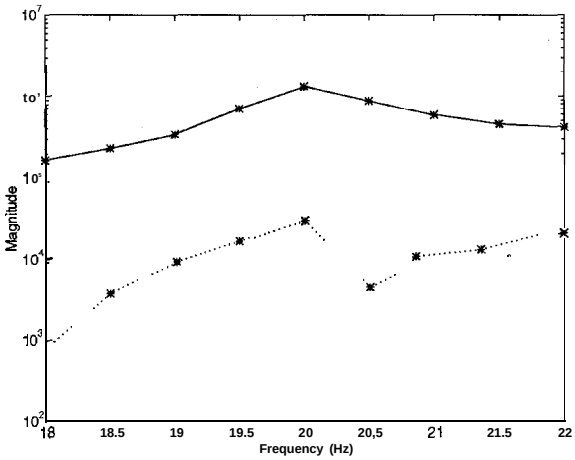


Figure 4: Results of the identified first mode.

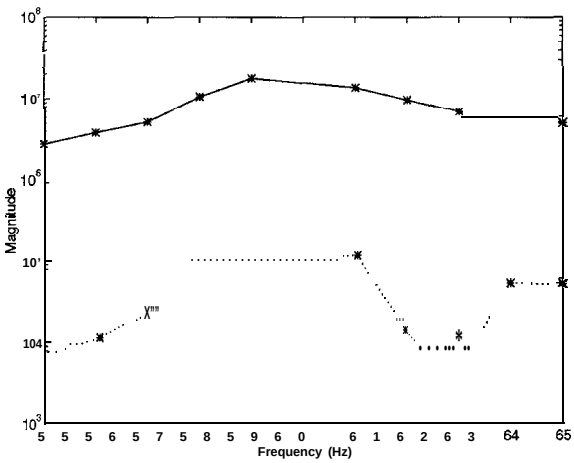


Figure 5: Results of the identified second mode.

In Table 1 'FEM' represents the results of finite element model, 'ID' represents the results of the identified parameters, and 'ID-FEM%', which equals ID-FEM/FEM x 100%, represents the difference between the identified parameters and the parameters of finite element model. The results show that the identified parameters are close to the results of finite element model. Figures 4 and 5 show the absolute values of the experimental data (solid line) D_i and the error (dotted line) $D_i\omega^2 + \omega_i^2 a - D_i\omega_i^2$ of the identified model for the first mode and second mode, separately. The error is about 100 times smaller than the experimental data at each frequency ω_i for both modes. This implies (1)the experiment is accurate, and (2)the identified linear model well represents the flexible arm system. From the results in this paper, the following observations are noted:

1. Table 1 shows that the identified natural frequencies ω and amplitudes a for both modes are around 1% and 8% lower than the ones of finite element model, respectively. The finite element

model well represents the real system. Possible errors are from (i)friction of motor and shaft part, (ii)the weight of the wire for accelerometer, (iii)the model error of the flexible arm pinned to the shaft.

2. The identified models of the first two modes precisely represent the real system. This experiment is well designed for the modal identification of this flexible arm system.

3. When the frequency of the sine-wave input is very close to the natural frequency, it is difficult to get reliable output data because of the resonance of the flexible arm system. The experimental data for the second mode with input at frequency 60 Hz are not used.

CONCLUSIONS

The modal parameter identification of a flexible arm system is presented in this paper. The identified results, which are obtained by applying a least-squares technique to the sine-wave response data, show that the identified linear model precisely represents the real system and the experiment setup is well designed for the system identification of this flexible arm system. The identified results are close to the model obtained by applying finite element analysis. This shows that the finite element model well represents the real system. When the frequent y of the sine-wave input is very close to the natural frequency of the system, it has difficulty to get reliable data because of the system resonance.

ACKNOWLEDGMENTS

This research is supported in part by NASA Grant NCC5-228. This support is greatly appreciated.

REFERENCES

1. Lew, J.-S., Juang J.-N. and Longman R. W., " Comparison of Several System Identification Methods for Flexible Structures," *Journal of Sound and Vibration,* Vol. 167, No. 3, 1993, pp. 461-480.

2. Pappa, R. S., Schenk, A., and Noll, C., "ERA Model Identification Experiences with Mini-Mast," *Proceedings of the 2nd USAF/NASA Workshop on System Identification and Health Monitoring of Precision Space Structures,* Pasadena, CA, March 1990.

3. Hildebrand, F. B., *Introduction to Numerical Analysis,* McGraw-Hill, New York, 1974.

4. Craig, R. R., *Structural Dynamics: An Introduction to Computer Methods,* Wiley, New York, 1981.

98 URC029

Autonomous Mobile Platform for Research in Cooperative Robotics

Ali Daemi, Edward Peña, Paul Ferguson

NASA Center for Autonomous Control Engineering
Electrical and Computer Engineering Builiding
The University of New Mexico
Albuquerque, NM 87131-1341, USA
daemi@slider.unm.edu

Abstract

This paper describes the design and development of a platform for research in cooperative mobile robotics. The structure and mechanics of the vehicles are based on R/C cars. The vehicle is rendered mobile by a DC motor and servo motor. The perception of the robot's environment is achieved using IR sensors and a central vision system. A laptop computer processes images from a CCD camera located above the testing area to determine the position of objects in sight. This information is sent to each robot via RF modem. Each robot is operated by a Motorola 68HC11E micro-controller, and all actions of the robots are realized through the connections of IR sensors, modem, and motors. The intelligent behavior of each robot is based on a hierarchical fuzzy-rule based approach.

Keywords: mobile robotics, autonomous robots, autonomous agents, cooperative robotics

1 Introduction

Robotic systems have been employed in many space applications to replace humans in dangerous environments. NASA is currently investigating the use of multiple robots as part of its initiative for subsequent Mars Pathfinder missions. Using multiple cooperative robots in applications such as planetary exploration can result in a more cost-effective system. A team of mobile robots will reduce the overall operation time for the mission, and by assigning fewer tasks to each robot, the design of the robot becomes simpler and less expensive. Also, using multiple robots increases the reliability and robustness of the system through redundancy.

Recently, a group at the NASA Autonomous Control Engineering Center was formed to investigate the use of soft computing techniques in cooperative robotic applications. One of the objectives was to design and construct a test bed for various cooperative robotic applications including terrain exploration. This paper describes the general architecture of this platform including the structure and mechanics, environment perception, computational hardware and intelligence.

2 Development of Test bed

Our platform is similar to the robot soccer test bed which consists of a team of mini-robots, central vision system, and a laptop computer. This platform can be used for various applications in cooperative robotic research within a laboratory environment. The following subsections will describe the platform architecture.

2.1 Mechanics of Autonomous Robot

The robots structure and mechanics were based on 1/18 and 1/10 scale radio controlled cars. These cars employ a four wheel design: two front wheels and two rear wheels on a rectangular chassis. The rear wheels are driven by a single DC motor. The DC motor is powered by a 7.2V NiCd battery pack through a H-Bridge power transistor circuit. Ackerman steering is implemented using a servo motor to control the position of the front wheels via a four bar linkage. The advantage of using a servo steering system is the simplicity in implementation and control. The disadvantage is that an optimal turning radius is compromised as compared to an alternative steering system such as differential drive where the turning radius may be completely eliminated. Power is transmitted directly to the rear wheels through a ball differential gear fixed to the rear axle. Making use of a differential gear reduces the turning radius and improves maneuverability of the vehicle.

2.2 Environment Perception and Communication

The robots sensory perception is realized through a remote vision system and onboard infra-red sensors. A Cognachrome 2000 Color Vision System is used as the central vision system for the robots. A CCD camera is positioned above the testing area and transmits a color NTSC signal at 60Hz to the video processing board which is connected to a laptop computer through the serial port. The vision system has the capability of tracking multiple objects in the camera's field of view. Objects are recognized as regions of color in the image of each frame, and the position of each object is determined by calculating the center of the identified color region. The laptop functions as an image acquisition/recognition system sending position information of all objects on the testing area to each robot.

Communication between laptop and robots is achieved by way of a spread spectrum 19.2 Kbps RF modem and is connected to the micro-controller via serial connection. Each robot will be identified by a unique call number. This call number will be used throughout the test duration by all the robots to perform certain functions dictated by the task situation.

Four infra-red sensors were positioned around the rectangular perimeter of the robot. Each sensor has a range of approximately 12 inches and provides proximity information of nearby objects. This information is to be used in a collision avoidance routine as part of the robot's intelligent behavior. The mechanics of the environment perception and communication are illustrated in Figure 1.

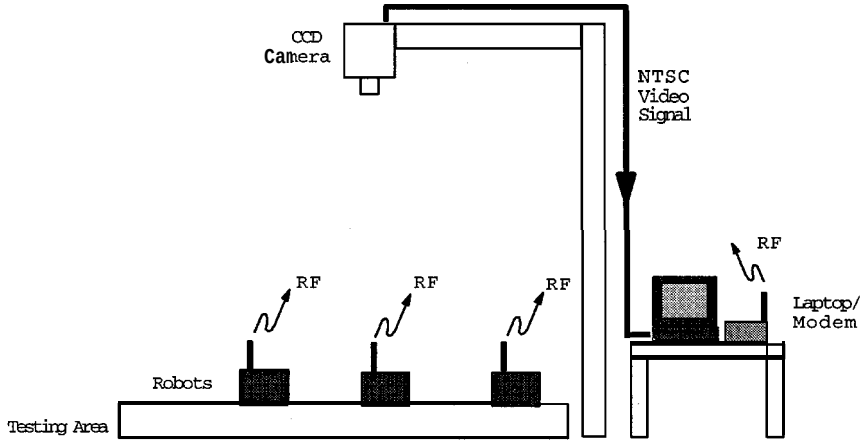


Figure 1: Illustration of Environment Perception and Communication

2.3 Computational Hardware and Intelligence

The robot's autonomous behavior (action and perception) is executed by an onboard 8 bit Motorola 68HC11 micro-controller board. This board contains 7 analog inputs, 9 digital inputs, 4 motor outputs, an expansion bus, infra-red transceivers, and 32KB of RAM. The board is powered by a 9.6V NiCd battery pack. The sensory perception, communication, and mobility is accomplished through the interconnections of the Motorola board and peripheral components (IR sensors, modem, and motors).

The intelligent behavior is based on an adaptive hierarchical fuzzy control scheme [2, 3, 4] which is executed locally on the micro-controller. An example is shown in Figure 2.

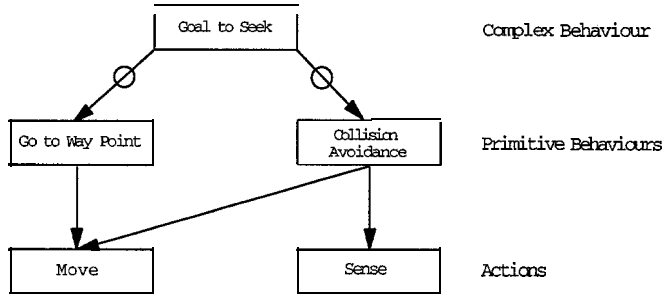


Figure 2: Behavioral Hierarchy for Finding Rocks

The hierarchical structure contains several layers of behaviors and the number of behaviors are determined based on the task. Complex behaviors are located at the top of the hierarchy structure. For example, if our objective is to seek a particular rock, a complex behavior could be goal *to seek*. This behavior describes a collision-free navigation to a rock of interest. Under complex behaviors, there are primitive behaviors such as go *to way point* and *avoid collisions*. These primitive behaviors will direct the robots to travel a straight

line trajectory to the rock of interest and avoid collisions with other objects on the testing area. Basic actions such as move and sense are located at the bottom of the hierarchy, and implement a direct connection between software and the hardware.

3 Conclusions

One of the objectives of this group was to design and construct a platform for various applications in cooperative robotics. A test bed that can be used for applications such as planetary exploration was presented. One limit ation with this setup is that the camera used in the vision system is stationary which restricts testing to a small area. Future considerations for this platform will include decentralizing the vision system by implementing vision onboard each robot and modifying the structure and mechanics of the robots to accommodate outdoor environments. This will increase the complexity of our system, but the result is that each robot will be more self-sufficient and have the capability of exploring larger areas and managing unstructured terrain.

Acknowledgements

This work was supported in part by NASA under contract # NCCW-0087.

References

[1] Jones, Joseph, Flynn, Anita. Mobile Robots. AK Peters Publishing Co., Wellesley Massachusetts, 1993.

[2] Tunstel, E., "Mobile Robot Autonomy via Hierarchical Fuzzy Behavior Control", Proceedings of 6th Intl. Symp. on Robotics and Manuf., 2nd World Automation Congress, Montpelier France, May 1996, ASME Press, pp 837-842.

[3] Mirabal, S., Padilla, D. and Oliveira, M. A., "A Hierarchical Fuzzy Controller Based Autonomous Soccer Player: A hierarchical fuzzy controller for cooperative multiagent autonomous robotics", submitted to WAC'98.

[4] Delgado, S., et. al. ," Soft Computing Based Approach to Cooperative Autonomous Robotics", submitted to URC'98.

[5] Daemi, A., Pena, E., Ferguson, P., "Autonomous Robot for Playing Soccer", submitted to WAC'98.

Comparison Of Human Modelling Tools For Efficiency Of Prediction Of EVA Tasks

H. Charles Dischinger, Jr. and Tomas E. Loughead
EO66, Marshall Space Flight Center, AL 35812 and Sigmatech, Inc., Huntsville, AL

Introduction

Construction of the International Space Station (ISS) will require extensive extravehicular activity (EVA, spacewalks), and estimates of the actual time needed continue to rise. As recently as September, 1996, the amount of time to be spent in EVA was believed to be about 400 hours, excluding spacewalks on the Russian segment (Dischinger, Evans, and Loughead, 1996). This estimate has recently risen to over 1100 hours, and it could go higher before assembly begins in the summer of 1998. These activities are extremely expensive and hazardous, so any design tools which help assure mission success and improve the efficiency of the astronaut in task completion can pay off in reduced design and EVA costs and increased astronaut safety.

The tasks which astronauts can accomplish in EVA are limited by spacesuit mobility. They are therefore relatively simple, from an ergonomic standpoint, requiring gross movements rather than tine motor skills. The actual tasks include driving bolts, mating and demating electric and fluid connectors, and actuating levers; the important characteristics to be considered in design improvement include the ability of the astronaut to see and reach the item to be manipulated and the clearance required to accomplish the manipulation. This makes the tasks amenable to simulation in a Computer-Assisted Design (CAD) environment. For EVA, the spacesuited astronaut must have his or her feet attached on a work platform called a foot restraint to obtain a purchase against which work forces may be actuated. An important component of the design is therefore the proper placement of foot restraints (Loughead and Dischinger, 1997). Prior to the development of CAD and human models which operate in CAD, the human factors engineers responsible for EVA design relied heavily on two-dimensional drawings, experienced guessing, and repeated testing to properly place foot restraints. This iterative process is effective, but it is expensive, primarily the tests are costly. The practical result of this was that tests were often deferred until late in the design process in an attempt to save costs. However, at a late phase of design, drawings of the structural support hardware were nearing completion. Redesign of these components typically became warranted only if the EVA interfaces were found to be completely unworkable. With the advent of simulated humans in CAD environments, many of these gross design drawbacks could be detected while the hardware was in the early phases of design.

Human modelling software. One of the first efforts to model a human that was three-dimensional, had appropriate (human-like) motion capabilities, and was anthropometrically scaleable was conducted by the University of Pennsylvania, partially with NASA funding (*cf.* Estakov *et al.*, 1989). The resulting software, called Jack, allows the user to import CAD files of hardware for the simulated human to interact with. The human factors designer then develops an appropriately-scaled human model and examines the human/machine interfaces. The model is moved to a position from which the task might be performed, and reach, visibility, and clearance are evaluated.

Software evaluation. The ultimate test of space hardware comes on-orbit, when astronauts actually use the system. Prior to that, however, most hardware systems with human interfaces are examined underwater, where both mockups of the hardware to be manipulated and the test subject can be made neutrally buoyant. Since the hardware neither sinks nor floats, it handles as if it were weightless. The simulated weightlessness causes the test subject to require attachment to solid structure in order to accomplish work, as on orbit. This type of simulation thus provides a good earth-based approximation of how the tasks will be accomplished in space. Thus, evaluation of task feasibility in a neutral buoyancy simulator is an excellent method of assessing the effectiveness of a design that has been developed using human modelling software.

The reliability of the software in predicting a usable design is clearly the most important criterion for its utility. However, since many human factors engineers may not have expertise in CAD and computer programming, the simplicity of use of the package is also an issue. Most human factors engineers use human modelling as one of their professional tools, rather than as the primary tool, as CAD might be for hardware engineers. Usability is thus especially important to this type of user. The interface should be so configured that, after an hiatus in use of a week or two, the human engineer can at least recall how to locate a common functionality without resorting to using the documentation. The package should be robust enough that it crashes only under extremely demanding circumstances. It should be capable of importing CAD files from a variety of sources (i. e., different formats), and the resulting models should be usable with minimal manipulation and repair.

ISS Assembly Flight 6A. The Space Station Remote Manipulator System, SSRMS, is the Space Station version of the Canadarm robot, the Shuttle Remote Manipulator System, which currently flies on most Space Shuttle missions. The SSRMS will be delivered to Station by the Shuttle on the sixth American construction mission (Assembly Flight 6A). Marshall Space Flight Center is responsible for packaging the robot in the Shuttle and supporting its deployment once the Orbiter arrives at the Station. As part of this effort, the authors have designed the EVA work environment; this includes placement of foot restraints such that the spacesuited astronauts can see and reach the components to be manipulated, with proper clearance to accomplish the work.

Methods

Human Modelling "Validation." In order to assess the reliability of human modelling as a human factors tool, a series of computer simulations was conducted, and the results were compared with the evaluation of the design in underwater testing (*cf.* Etter, *et al.*, 1996; Loughead and Dischinger, 1997). These simulations were done using Jack with CAD models supplied by the hardware developers (Spar Aerospace, The Boeing Company, McDonnell Douglas Aerospace, and Marshall Space Flight Center); the spacesuit model was obtained from NASA Ames Research Center. The human models were configured such that anthropometric sizes of 5th, 50*, and 95th percentile humans were simulated. The neutral buoyancy tests were conducted at the Marshall Space Flight Center Neutral Buoyancy Simulator. They consisted of spacesuited engineers and astronauts working in pairs. The test subjects attempted the tasks and then rated their feasibility based on whether the object to be manipulated could be seen and reached and whether there was sufficient clearance to accomplish the task. The ratings were used to determine whether the computer simulations had correctly predicted task feasibility. In addition,

photographs of test subjects attempting tasks were compared to the images generated by the computer simulations.

Usability. Informal evaluations of the software's usability were conducted throughout the simulation process.

Discussion

Human Model "Validation." Neutral buoyancy evaluation of the design was considered the standard with which to compare the modelling simulation results. This is based on NASA's experience, which has shown that tasks which can be performed in neutral buoyancy can be accomplished on-orbit. Thus, if a task was rated acceptable in the water, the modelling simulation was considered successful. Three important parameters which were assessed with respect to acceptability by the spacesuited test subject were work clearance (the envelope in which the subject operated to access the task worksite), visibility (the subject's ability to see the object to be manipulated at the worksite), and reach (the subject's ability to place a hand or a tool on the work object). These were assessed from the foot restraint as it was placed on the support structure. Failure of the design to support this last component has significant impact, because movement of the restraint attach point requires further hardware structural analysis and possibly redesign. The results of the evaluations for these parameters are given in Table 1.

Parameter	Percent tasks acceptable as designed
Work clearance	91
Visibility	100
Reach and manipulate	100

Table 1. Results of neutral buoyancy task evaluations of design derived from computer simulation.

For one placement of the foot restraint, work clearance was insufficient, and the restraint had to be moved. Since visibility and reach/manipulate were acceptable for all tasks, it is possible that inaccurate spacesuit or hardware models, rather than the human simulation, gave rise to this error. That is, in all cases, the human model correctly assessed those components of the task that were driven by anthropometry; the work clearance in EVA is defined by interference between the spacesuit and the hardware.

The images derived from the modelling exercises reliably predicted the ways that the test subjects performed the tasks. They were used to brief the test subjects before the test, so that they would understand the nature of the tasks. In some cases, they highlighted weaknesses in the design. In Figure 1a, the modelling simulation depicts one astronaut passing a long bolt to the other. These bolts must be stored in the tubes below the horizontal subject. The simulation showed that subjects shorter than 50th percentile for stature would not be able to reach the hand-off; the bolts are 107 cm long, and the reach distance was determined to be 127cm. In the briefing, astronauts suggested they could accomplish the task by having the recipient get out of his/her restraint to receive the bolt, as shown in the photograph (Figure lb). This solution was shown in the test to be acceptable, and redesign was avoided.

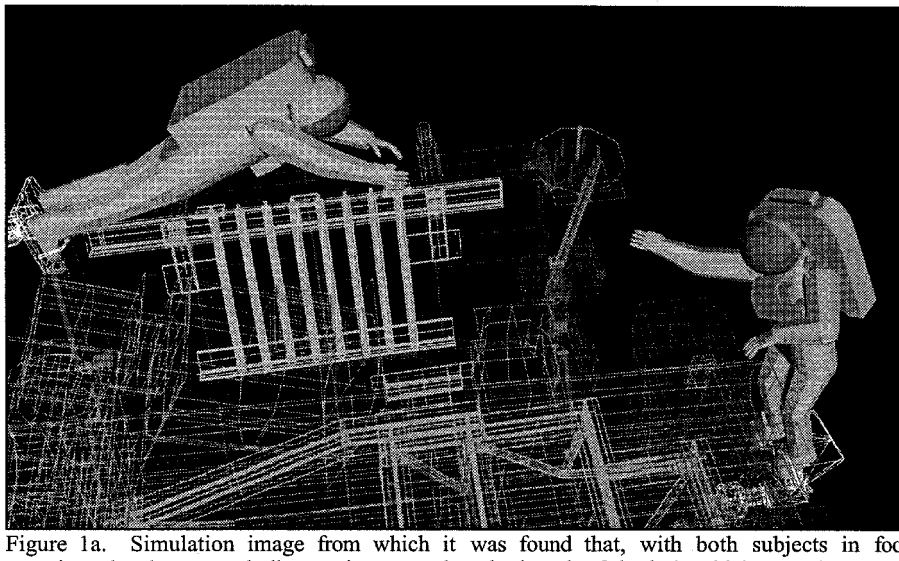


Figure 1a. Simulation image from which it was found that, with both subjects in foot restraints, the closest reach distance is greater than the length of the bolt which must be passed between them.

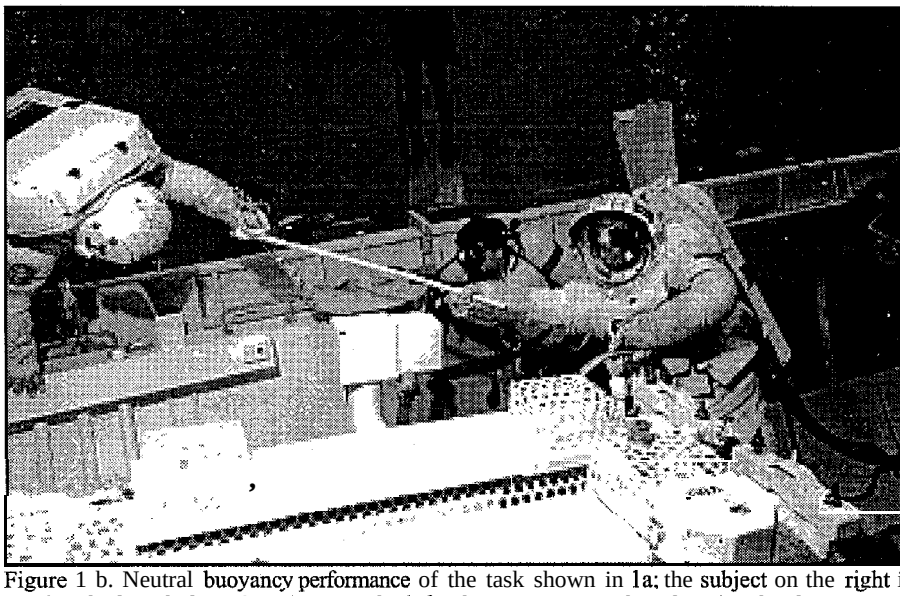


Figure 1 b. Neutral buoyancy performance of the task shown in 1a; the subject on the right is passing the long bolt to the subject on the left. The recipient is tethered to the Shuttle to prevent drifting away but is out of the foot restraint. The bolt will be stowed in the tubes below the recipient.

Usability. As discussed above, usability is a concern for human factors professionals who are not CAD experts or computer scientists. Jack presented some challenges for us. Probably the most frustrating interactions with Jack were the occasional system crashes. They often resulted in unrecoverable loss of work. The user interface lacks consistency; some functionalities are easily located in pull-downs, while others of seemingly similar level of utility and complexity

might only be executable by an arcane command at the command line. File conversion into the Jack format frequently resulted in models which were extremely large, lost features such as surfaces, had inappropriate surface extensions , or were uninterpretable by Jack. The conversion utilities in Jack will handle a fairly wide variety of common CAD formats including IGES and dxf, but file was cleanup was often required.

A comparison study. Thus, while Jack contains an excellent simulated human and gave very reliable results in predictions of EVA work tasks, our tool needs include software which has been well human-factored. We have initiated a study to determine whether a similar package might have Jack's capabilities without Jack's usability drawbacks. We have obtained ERGO, a human model developed by Deneb Robotics, Inc. and are initiating a systematic comparison of its capabilities and strengths with those found in Jack. The approach is to evaluate the reliability of ERGO as we have done for Jack, comparing task simulation with actual neutral buoyancy results. Following this, a set of usability metrics will be applied to both packages to determine whether one has greater utility for the non-expert user that human factors engineers are likely to be. Finally, trade studies will be performed to help identify the characteristics of the two packages which are most important in decisions about the circumstances under which one or the other might be used.

New Jack. Shortly after this study was proposed, Jack licensing was transferred by the University of Pennsylvania to Transom Technologies, Inc. Transom is commercializing the product for application as a human factors tool. The company has stated that usability is a key concern in the development of new releases. Thus, in order to derive a valid comparison, the usability metrics will be applied to commercially-available versions of each product.

References

Dischinger, H. C., Jr. and T.E. Loughead. 1997. The application of human modelling software in the development and verification of eva interfaces on an international space station launch assembly successes and limitations. Proc. Technology 2007.22-24 Sept., 1997. Boston.

Dischinger H. C., Jr., A. M. Evans, and T. E. Loughead. 1996. Use of human factors computer modeling to predict locations and settings for crew aids used in extravehicular activity on an International Space Station assembly mission. In Proc. Hum. Fact. Ergon. Soc. 40th Ann. Mtg. 2-6 Sept., 1996. Philadelphia. 1254.

Esakov, J., N.I. Badler and M. Jung. 1989. Human task animation from performance models and natural language input. Graphics Technology in Space Applications, NASA JSC Conference Publication 3045.

Etter B.D., H. C. Dischinger, Jr., and T. E. Loughead. 1996. Evaluation of an anthropometric human body model for simulated EVA task assessment. Proc. Southeastern Simulation Conference '96. 7,8 Oct., 1996. Huntsville. AL.

Loughead, T.E. and H.C. Dischinger, Jr. 1997. Using a virtual environment in ergonomic evaluation of the placement of crew aids for an International Space Station assembly flight. In High Performance Computing 1997. A. Tentner, ed. Proc. 1997 Simulation MultiConf. 6-10 April, 1997. Atlanta.

Finite Element Modeling of Crack Propagation-A Practical View

Robert Valerio[1] and Walter Gerstle[2]
Department of Civil Engineering, University of New Mexico
Albuquerque, NM 87131
Phone: (505) 277-3458; FAX: (505) 277-1988
Email: gerstle@unm.edu; rvalerio@cochiti. arc.unm.edu

Abstract

The finite element method (FEM) modeling of discrete crack propagation is an important unsolved problem in the practical analysis of mechanical and structural systems. Although academic software has been developed over the past twenty years for the purpose of FEM modeling of crack propagation, the software has been extremely limited in its capability. Essentially, only two-dimensional elasto-static crack propagation is currently possible. The first elasto-static three-dimensional crack propagation modeling software is just now becoming available. Commercial-grade software for this type of analysis is not readily available and is very expensive to maintain and operate.

To perform practical linear elastic FEM analysis, three fundamental capabilities are required. The first capability is to represent a crack in a 2D or 3D body. The second capability is to automatically mesh the cracked geometry. The third capability is to compute stress-intensity factors along the crack front, and from these to predict a new geometry with a propagated crack (i.e. determine the direction and length of crack propagation). With these three capabilities in place, it should be possible to model crack propagation as a sequence of repeated steps.

Some commercial software is now available with sophisticated attributes such as fully automatic meshing and command languages that allow the analyst to write scripts with sophisticated FEM analysis "recipes". The two primary goals of this paper are (1) to identify whether it is currently practical to use a commercial FEM code to perform crack propagation analysis; and (2) to identify the necessary attributes a FEM code must have to accomplish crack propagation analysis.

One of the most advanced FEM codes, in terms of its geometric modeling and automatic meshing capabilities is SDRC-IDEAS. This code is used here to perform crack propagation analysis. We report on the successes and problems encountered.

A FEM code called CoMeT, currently under development at the University of New Mexico, is designed to address many of the issues involved in automatic crack propagation analysis. Using CoMeT, a strategy is being developed for performing crack propagation analyses. By using an example of elasto-static FEM analysis, au objective comparison can be made of the level of effort required to generate useful results obtained from both SDRC-IDEAS and from CoMeT.

Introduction

Investigation of linear elastic fracture mechanics (LEFM) behavior often requires use of the FEM. FEM programs (such as the software SDRC-IDEAS and Parametric Technologies Mechanica) have become increasingly user-friendly and powerful in recent years. In this paper,

[1] Graduate Student, Civil Engineering, University of New Mexico
[2] Associate Professor, Civil Engineering, University of New Mexico

we investigate the use of such commercially available FEM programs for performing three-dimensional LEFM analyses. As will be seen, even with the sophistication offered by these modem applications, they still fall somewhat short of providing the functionality necessary for practical LEFM analyses. In this paper, we determine the additional functionality that these FEM programs must provide to accomplish practical three-dimensional LEFM crack propagation simulations.

Energy-based automatic mixed-mode crack-propagation modeling has been described in (Xie, Gerstle, and Rahulkumar 1995) where, based upon an energy principle and a virtual crack extension technique, a practical FEM modeling approach for automatic LEFM crack-propagation analysis is described. By decomposing the displacement field obtained from a conventional FEM analysis into symmetric and anti-symmetric displacement fields with respect to the crack, Mode-I and Mode-II energy release rates can be determined using a generalization of the stiffness derivative method (Parks 1974). Using this virtual crack extension technique, the corresponding stress-intensity factors can then be calculated. The load at which a crack propagates and the propagation direction can be determined using one of the well-established mixed-mode crack-propagation criteria, as for example is described in (Hussain, Pu, and Underwood, 1974). Unlike the previously developed approaches to crack propagation, the energy-based approach does not require the use of a symmetric crack-tip mesh, nor crack-tip singular elements, which greatly simplifies the subsequent re-meshing process to allow for crack propagation. The approach is so simple that it would appear that currently available commercial general-purpose FEM programs might be capable of modeling LEFM crack propagation in three-dimensional solids.

In this paper, for simplicity, we limit our attention to Mode I 3D problems with known analytical solutions. For such problems, to determine the Mode I stress-intensity factor, K_I, it is necessary to determine the potential energy, Π_1, of the problem, then move the crack front locally be a small amount, Δa, resulting in a change in crack area of AA, and then compute the potential energy, Π_2. The energy release rate, G_I, is then approximated by $G_I = (\Pi_2 - \Pi_1)/\Delta A$, and the stress-intensity factor at the location where the crack area is changed has an average value given by $K_I = (EG_I)^{1/2}$, where E is Young's Modulus.

Analysis Results

We choose two different test problems to see how the software meshes the solid, handles modifications to the solid model, handles changes to the nodes and elements, and compares to a bench mark LEFM hand calculation. The two different problems are those pictured in Figure 1. The cylinder has a uniform unit tensile traction acting normal to the penny-shaped crack at the center. The rectangular medium also has a crack at its center, but this crack goes fully through the medium. The elastic media are composed of the same material with a Young's modulus of 1 psi and a Poisson's ratio of .3. The following figure is the solid model from which the FEM models were constructed.

The FEM model featured two separate element types for comparison. The first type of element was a four-noded tetrahedron. The second element type featured was that of a ten-noded tetrahedron. A freer mesh around the crack tip was also imposed on the model with the 10-noded tetrahedral elements. Within the cylindrical model, the number of elements along the crack front is six for the model with the four-noded tetrahedral elements while the model with the ten-noded tetrahedral elements had twenty elements along the crack front.

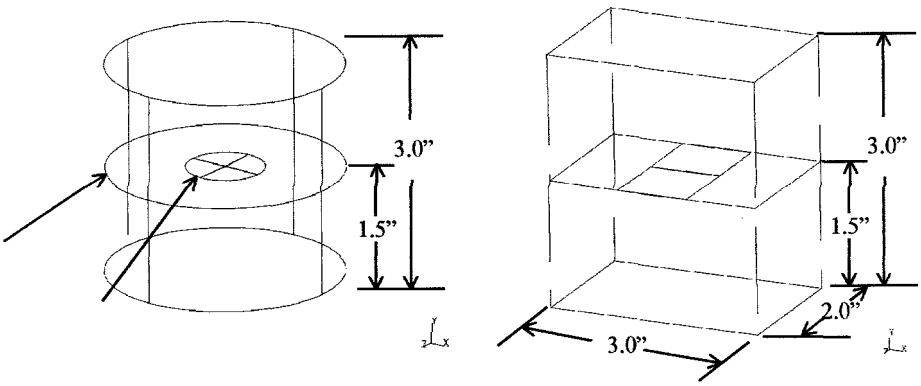


Figure 1: (a) Solid Model of Cylinder with Penny Shaped Crack; (b) Solid Model of Rectangular Medium with Through Crack.

The FEM model used nodal constraints to restrict the system from rigid body motion by applying Pin connections (fixed in X, Y, and Z translation axis), roller connections (fixed in the Y and Z translation axis) and a slider connection (freed in the Z translation axis). These connections are pointed out in Figure 2.

The analysis was run using the "Linear Statics" solver in I-Deas. The output requested for each run was the displacement of the system, the maximum principal stress and strain of the system, the strain energy of the system, and finally the strain energy density of the system. AU the outputs correlated well with hand calculations except for the main result we were looking for and that was the total strain energy of the system. The total strain energy value was off quite a bit from our hand calculations. Sometimes it was greater than our calculated value and sometimes it was less then our calculated value but there was never a strong correlation. It was decided to investigate this problem further and a unit cube was modeled and analyzed. We knew the material type, the load on the system, the number of elements to be used and what the total strain energy of the system should be. Again, the results from the code were off by a factor of around one hundred percent. However, it was noticed that the ratio of nodes to elements was also around that same factor. The investigation resulted in the determination that the output given by the code is the total strain energy at the nodal level and not the elemental or system level. At that point a ratio was calculated between nodes and elements of the system and the total strain energy output was divided by that ratio. Once this was done the hand calculations and system results coincided. We notified SDRC of the bug in their program; hopefully it will be fixed soon.

Now there was a different problem encountered with the results given by I-Deas. Finite element models were constructed from the different geometrical mediums and nodes and elements were moved and reoriented to simulate propagation of a crack. The resulting total strain energy of the system in each case did not provide the significant digits necessary to compute an accurate K_I value for the system. In some cases, the change in total strain energy of the system ended up being zero because or the significant digit issue.

Figures *3* and *4* show the mesh and results from one run of the cylinder and rectangle with a fine mesh. Table I compares the results of the analysis of the coarse, free, and theoretical values generated.

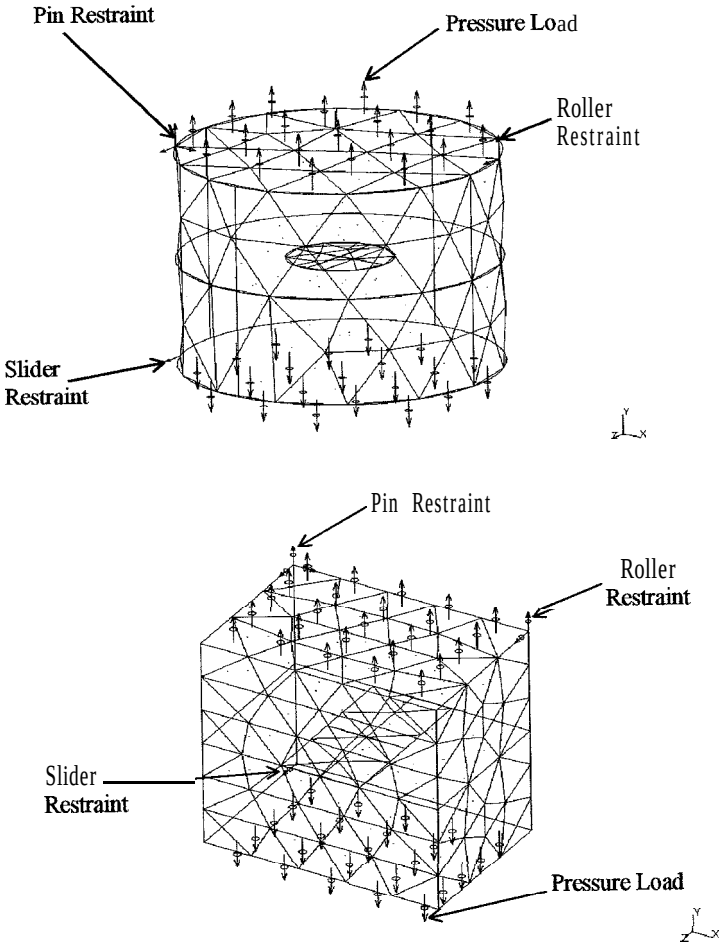


Figure *2.* a) Coarse Mesh and Restraint of the Cylindrical Model; (b) Coarse Mesh and Restraint of the Rectangular Model.

Conclusions

In conclusion, SDRC I-deas would be capable of computing LEFM stress-intensity factors if it were not for several shortcomings:

(1) there appears to be an error in computing total strain energy;

(2) total strain energy is reported to only four significant digits: not enough to perform accurate finite difference calculations;

(3) the lack of a scripting language makes it extremely tediuos to perform stress-intensity factor calculations.

The first two points could be easily addressed be the vendors of the software. The third point, the lack of a scripting language, is far more difficult to address, and basically would require a total re-write of the code. It appears that ProEngineer supports such a scripting language, but we have not been able to try ProEngineer for the study of fracture problems.

We are currently engaged in the development of a computational mechanics toolkit, CoMeT, at the University of New Mexico, designed using the command language called Scheme. With its current design, CoMeT is capable of displaying quantities to a user-specified degree of precision. Therefore, we belive that in the near future *LEFM* modeling in three dimensions will be easily performed using gneral purpose finite element codes.

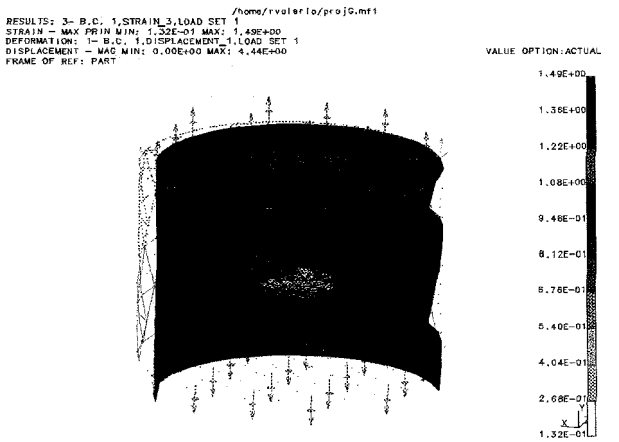


Figure 3. Results of Fine mesh of Cylindrical Medium.

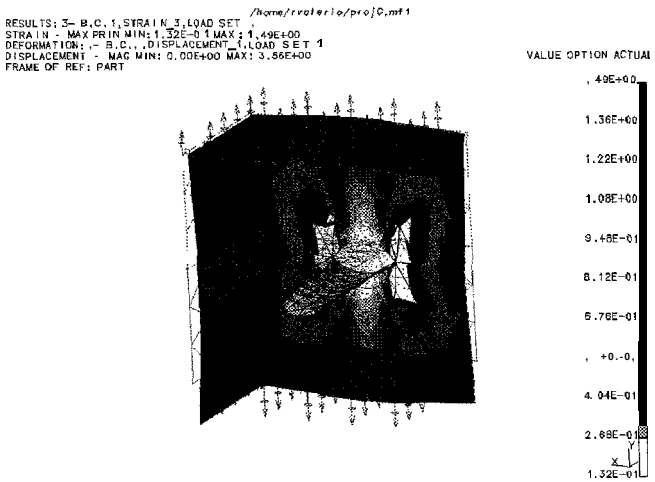


Figure **4.** Results of Fine mesh of Rectangular Medium.

Table I

	Π_1 (lb-in)	Π_2 (lb-in)	Δl (in)	ΔA (in^2)	G_I (lb-in)	K_I $Psi\text{-}in^{1/2}$	$K_{I(exact)}$ $Psi\text{-}in^{1/2}$
Cylinder - Coarse Mesh							
Corner Node	58.44	58.44	.006000	.001800	O.	0.	1.128
Crack Front	58.44	58.44	.006000	.01900	o.	o.	1.128
Cylinder - Fine Mesh							
Corner Node	65.08	65.08	.003000	.0004500	0.	0.	1.128
Mid-side Node	65.08	65.08	.003000	.0004500	0.	0.	1.128
Crack Front	65.08	65.08	.003000	.0094500	0.	0.	1.128
Block - Coarse Mesh							
Corner Node	9.182	9.185	.006000	.001800	2.	1.	1.487
Crack Front	9.182	9.185	.006000	.02400	O.	0.	1.487
Block - Fine Mesh							
Corner Node	10.98	10.98	.00200(-)	.000200	0.	0.	1.487
Mid-side Node	10.98	10.98	.002000	.000200	o.	0.	1.487
Crack Front	10.98	1099	.002000	.008000	1.	1.	1.487

Acknowledgments

The financial support provided by the Center for Autonomous Control in Engineering (ACE) and the Waste Management Education and Research Consortium (WERC) is gratefully acknowledged. Also, the help of Lary Lenke is appreciated.

References

Hussain, M.A., Pu, S.L., and Underwood, J.H. (1974). "Strain energy release rate for a crack under combined mode I and mode II." Fracture Analysis, STP 560, ASTM, Philadelphia, Pa.

Parks, D. M., (1974). "A stiffness derivative finite element technique for determination of crack tip stress intensity factors." Int. J. Fracture, 10(4), 487-502.

Xie, M., Gerstle, W. H., and Rahulkumar, P. (1995). "Energy-Based Automatic Mixed-Mode Crack-Propagation Modeling." J. Engrg. Mechanics, ASCE, Vol. 121, No. 8,914-923.

[1]Influence of Material Distribution on Impact Resistance of Hybrid Composites

Ayo Abatan and Hurang Hu
Department of Engineering, Clark Atlanta University, Atlanta, GA 30314

INTRODUCTION

Impact events occur in a wide variety of circumstances. A typical example is a bullet impacting a target made of composite material. These impact events produce time-varying loads on a structure that can result in damage. As a first step to understanding the damage resistance issue in composite laminates, an accurate prediction of the transient response during an impact event is necessary. The analysis of dynamic loadings on laminated composite plates has undergone considerable development recently. Cairns and Lagace (1989), and Qian and Swanson (1990) used Rayleigh-Ritz energy method to determine the impact response of laminated plates. Christoforou and Swanson (1991), Kook et al. (1992), and Prasad et al. (1993) analyzed the impact response of composite plates using shear deformation plate theory. In recent work, Abatan et al. (1997) have obtained a closed-form solution for a rectangular plate with four edges simply supported subjected to a center impact load using classical plate theory. Abatan, Hu and Olowokere (1997) further investigated the problem and compared the analysis results of both classical plate theory and shear deformation theory, and found that classical plate theory predicts very accurate results for the range of small deformations considered.

In this study, the influence of cross sectional material distribution on the comparative impact responses of hybrid metal laminates subjected to low and medium velocity impacts is investigated. A simple linear model to evaluate the magnitude of the impact load is proposed first, and it establishes a relation between the impact velocity and the impact force. Then a closed-form solution for impact problem is presented. The results were compared with the finite element analysis results. For an 11 layer-hybrid laminate, the impact response as a function of material distribution in cross-section is presented. With equal areal weight, the effect of the number of laminate layers on the impact resistance is also investigated. Finally, the significance of the presented results is discussed.

A LINEAR MODEL FOR EVALUATING IMPACT FORCE

In literature, most of the investigators determine the impact force by using Hertz's law of contact (e.g. Cairns and Lagace, 1987; Kook at al. 1992). Since Hertz's law is nonlinear in nature, it is almost impossible to get the closed-form solutions for the problem. In this paper, a simple linear model is developed to evaluate the impact force.

Applying Newton's second law to an impactor, the impact force F(t) can be written as

$$F(t) = m\frac{dv}{dt} \tag{1}$$

It is assumed that the velocity of the impactor is zero compared to its initial velocity V_0 after it contacts the target. Integration of equation (1) gives

$$\int_0^{t_0} F(t)dt = mV_0 \tag{2}$$

‘ This research is supported through NASA-LeRC Contract NAS3-27785 and NASA-HiPPAC Contract NCC3-552

where t_0 is the impact duration. From the observation of experimental data, F(t) may be approximated as a half sine wave, i.e.

$$F(t) = A\sin\frac{t\pi}{t_0} \tag{3}$$

where A is the magnitude of the impact force. Substituting equation (3) into (2) yields

$$A = \frac{mV_0\pi}{2t_0} \tag{4}$$

The impact force F(t) is then obtained by substituting (4) back into (3) as

$$F(t) = \frac{mV_0\pi}{2t_0}\sin\frac{t\pi}{t_0} \tag{5}$$

This result is consistent with that obtained by Christoforou and Swanson (1991) using a Laplace transform technique for a linearized contact law.

The significance of equation (5) is that it establishes a linear relation between impact velocity (or impact momentum) and impact force. Since the impact energy is given by

$$E = \frac{1}{2}mV_0^2 \tag{6}$$

equation (5) also connects the impact energy with the impact force,

ANALYTICAL SOLUTION FOR A IMPACT PROBLEM

Assume an impact force P(x,y,t) is exerted over the top surface of a rectangular laminated plate with four edges simply supported. Assume the plate is a specially orthotropic laminate such that $B_{ij} = O$, and $A_{16} = A_{26} = D_{16} = D_{26} = O$, where A_{ij}, B_{ij}, D_{ij} *are* the stiffnesses of the laminate. The governing differential equations in terms of the midplane displacements are as follows (Reddy and Miravete, 1995):

$$A_{11}\frac{\partial^2 u}{\partial x^2} + (A_{12} + A_{66})\frac{\partial^2 v}{\partial x\partial y} + A_{66}\frac{\partial^2 u}{\partial y^2} = I_0\frac{\partial^2 u}{\partial t^2} \tag{7a}$$

$$A_{66}\frac{\partial^2 v}{\partial x^2} + (A_{12} + A_{66})\frac{\partial^2 u}{\partial x\partial y} + A_{22}\frac{\partial^2 v}{\partial y^2} = I_0\frac{\partial^2 v}{\partial t^2} \tag{7b}$$

$$D_{11}\frac{\partial^4 w}{\partial x^4} + 2(D_{12} + 2D_{66})\frac{\partial^4 w}{\partial x^2\partial y^2} + D_{22}\frac{\partial^4 w}{\partial y^4} = P - I_0\frac{\partial^2 w}{\partial t^2} + I_2\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right) \tag{7c}$$

where I_0 and I_2 are the mass inertias defined as

$$(I_0, I_2) = \int_{-h/2}^{h/2}(1, z^2)\rho dz \tag{8}$$

and p is the material density of the solids. The simply supported boundary conditions are satisfied by assuming the following displacement fields:

$$u(x,y,t) = \sum_{m,n=1}^{\infty} U_{mn}(t)\cos\alpha x\sin\beta y \tag{9a}$$

$$v(x,y,t) = \sum_{m,n=1}^{\infty} V_{mn}(t)\sin\alpha x\cos\beta y \tag{9b}$$

$$w(x,y,t) = \sum_{m,n=1}^{\infty} W_{mn}(t)\sin\alpha x\sin\beta y \tag{9c}$$

where $\alpha = m\pi / a$, $\beta = n\pi\ I\ b$. The problem is therefore reduced to the determination of the functions $U_{mn}(t)$, $V_{mn}(t)$ and $W..(t)$. With initial displacements and initial velocities assumed to be zero, it is found that $U_{mn} = V_{mn} = 0$, and

$$W_{mn}(t) = \frac{1}{\Omega_{mn} m_{33}} \int_0^t P_{mn}(\tau) \sin\Omega_{mn}(t-\tau) d\tau \qquad (lo)$$

where $P_{mn}(\tau)$ *is* the coefficient of impact load expanded by double sine series, Ω_{mn} is the natural vibration frequency of the plate given by

$$\Omega_{mn} = \left(c_{33} / m_{33}\right)^{1/2} \qquad (11)$$

where

$$c_{33} = D_{11}\alpha^4 + 2(D_{12} + 2D_{66})\alpha^2\beta^2 + D_{22}\beta^4$$

$$m_{33} = I_0 + I_2(\alpha^2 + \beta^2)$$

On substituting $W\},.(t)$ back in equation (9c), deflection w(x,y,t) can be determined. The strains and stresses in the laminate can be computed by using the strain-displacement relation and the constitutive relation, respectively.

With the same magnitude of the maximum force, the pulses of a half sine wave and a rectangular wave produce similar patterns for dynamic response (Abatan, Hu and Olowokere, 1997). Therefore, only rectangular waves will be considered hereafter. For a rectangular wave, the displacement w(x,y,t) is found from equation (1 O) as

$$w(x,y,t) = \sum_{m,n=1,3,}^{\infty} \frac{4f}{ab\Omega_{mn}^2 m_{33}} (\cos\Omega_{mn} t - 1) \sin\alpha x \sin\beta y \qquad (12)$$

where f is the magnitude of the impact force. It is noted that the maximum deflection will occur in the center of the plate, i.e. at x = a/2 and y = b/2.

In order to ensure the reliability of analytical solution, we performed finite element analysis, and compared the finite element results with the analytical analysis results. For a three layer laminate and a metallic plate, the relative difference of the maximum deflections predicted by equation (12) and finite element method was less than 1.5%. Therefore, the analytical results are reliable and accurate.

ANALYSIS RESULTS

In the following, the effect of the number of layers on dynamic response of laminates will be examined first. This is followed with results of the effect of material distribution on dynamic response for an 11 layer laminate. For all the laminates examined, the dimensions in the x-y plane are 7“ x 7“ (O. 1778m x O.1778m), and areal weight is 3 psf.

Effect of the Number of Layers on Dynamic Response

Four laminates were analyzed. These laminates were composed of metal and polymer. The material properties of the metal and polymer are listed in Table 1. The stacking sequence for all the laminates are similar, i.e. [metal/polymer/metal/.. ./metal]. The details of these laminates are given in Table 2.

TABLE 1. MATERIAL PROPERTIES FOR METAL AND POLYMER

Property	E(msi)	v	$\rho(lb \cdot \sec^2 / in')$
Metal	28.0	0.3	**7.254E-4**
Polymer	1.0	0.2	**9.067E-5**

TABLE 2. LAYOUT OF FOUR LAMINATES (UNIT: in)

No.of layers	metal/ layer(t_m)	polymer /layer(t_p)	metal/ total(h_m)	polymer/ total(h_p)	total thickness(h)	h_m/h (%)
3	0.0312	0.1000	0.0625	0.1000	0.1625	38
5	0.0210	0.0456	0.0630	0.0912	0.1542	41
9	0.0100	0.0488	0.0500	0.1952	0.2452	20
21	0.0060	0.0060	0.0660	0.0600	0.1260	"521

The deflections at the center of the plate for these four laminates as functions of time are shown in Fig. 1. From the figure, it can be seen that the 21 layer laminate has the higher deflections than other laminates, and the 9 layer laminate has the lowest deflections.

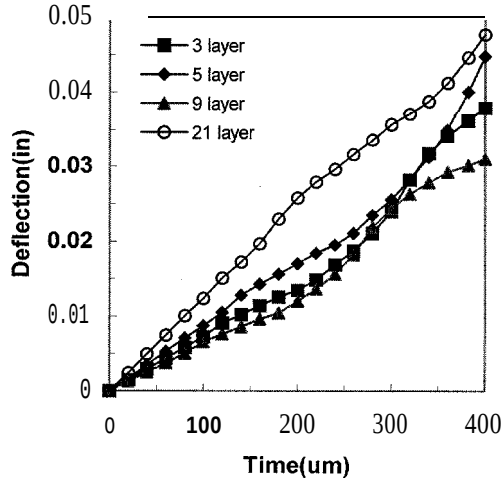


FIG. 1. DEFLECTION VERSUS TIME FOR FOUR LAMINATES.

For numerical comparison, Table 3 lists the maximum deflections and maximum stress component σ_{xx} at the center of the bottom surface of the plate for four laminates.

TABLE 3. A COMPARISON OF ANALYSIS RESULTS FOR FOUR LAMINATES

No. of layers	De flection(in)	Stress σ_{xx} (ksi)	h_m/h (%)
9	0.0311	69.45	20
3	0.0380	60.57	38
5	0.0448	83.60	41
21	0.0477	99.66	52

From the analysis results, it can be seen that for hybrid laminates with the same areal weight, the number of layers of laminate does not have much influence on impact response, but the overall percent of the metal (or polymer) has strong influence.

Effect of Material Distribution on Dynamic Response For 11 Layer Laminates

Seven 11 layer laminates with equal areal weight of 3 psf were analyzed to investigate the effect of material distribution on dynamic response. These laminates were composed of monolithic titanium (Ti-6Al-4V) sheets and Carbon Fiber Reinforced Plastic (CFRP) sheets. The material properties for Ti-6Al-4V and CFRP are listed in Table 4.

TABLE 4. MATERIAL PROPERTIES FOR Ti-6Al-4V AND CFRP

Property	E(msi)	E(GPa)	v	$\rho(lb \cdot \sec^2/in')$	$\rho(kg \cdot \sec^2/m')$
Ti-6Al-4V	17.2	118.6	**0.33**	4.114e-4	448.34
CFRP	19.7	135.7	*0.28*	1.449e-4	157.91

The layout of these laminates is listed in Table 5. Panels 1 and 7 are two special cases. They are homogeneous plates of pure CFRP and Ti-6Al-4V, respectively.

TABLE 5. LAYOUT FOR SEVEN LAMINATES OF EVEVEN LAYERS (UNIT: in)

panel	Ti-6Al-4V/ layer(t_m)	CFRP/ layer(t_p)	Ti-6Al-4V/ total(h_m)	CFRP/ total(h_p)	total thickness(h)	h_m/h (%)
1				0.3720	0.3720	0
2	0.0080	0.0472	0.0480	0.2360	0.2840	17
3	0.0120	0.0335	0.0720	0.1675	0.2395	30
4	0.0160	0.0199	0.0960	0.0995	0.1955	49
5	0.0186	0.0110	0.1116	0.0550	0.1666	67
6	0.0205	0.0047	0.1230	0.0235	0.1465	84
7			0.1311		0.1311	100

A cross sectional material distribution may be measured by h_m/h, where h_m is the total thickness of Ti-6Al-4V, and h is the thickness of the whole laminate. The maximum deflections at plate center versus cross sectional material distribution parameter h_m/h for all seven laminates are shown in Fig. 1. The maximum stress component σ_{xx} at the center of bottom surface of the plate for seven laminates versus h_m/h is shown in Fig. 2.

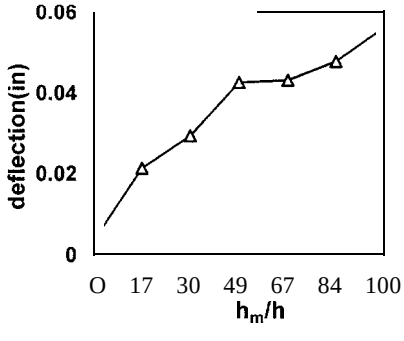


FIG. 1. MAXIMUM DEFLECTIONS OF SEVEN PLATES VERSUS h_m/h.

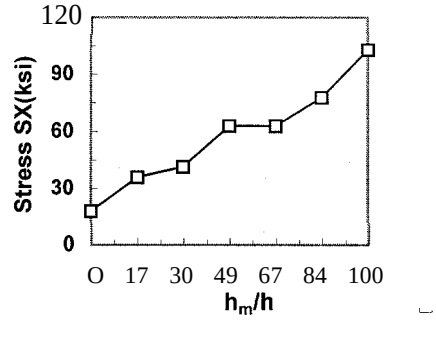


FIG.2. MAXIMUM STRESSES σ_{xx} OF SEVEN PLATES VERSUS h_m/h.

It is observed that for all seven laminates, the maximum deflections and maximum stresses σ_{xx} are increased as the cross sectional material distribution parameter h_m/h increases. For h_m/h=0, which is a pure CFRP plate, both the deflection and stress are minimum; For h_m/h=1 00%, which is a pure Ti-6Al-4V plate, both the deflection and stress are maximum. The results for all other laminates fall in between. It appears that the relations between the maximum deflection and h_m/h and between the maximum stress σ_{xx} and h_m/h are approximately linear.

CONCLUSION

A linear analytical model is proposed to evaluate the magnitude of impact force. This model provides a tool to estimate the impact load from the impact energy. For hybrid metal laminates subjected to low and medium velocity impact where elastic deformation is assumed, the effect of cross sectional material distribution on impact response was investigated. With equal areal weight for composite plates considered, the number of layers in a laminate does not have significantly effect on its impact resistance, but its relative material ratio does. Similarly, with equal areal weight, the relative ply thickness in a laminate does not have much effect on its impact resistance, but the total relative material distribution does have a significantly effect on impact response.

REFERENCES

Abatan, A. O., Hu, H., Lin, M., Roberts, G., and Pereira, M. (1997): On the Fabrication and Impact Modeling of Hybrid Metal Laminates, Proceedings of the HITEMP Review, NASA Publication #10192, Vol. I, Paper 12, 10pp.

Abatan, A. O., Hu, H., and Olowokere, D. (1997): Impact Resistance Modeling of Hybrid Laminated Composites, Proceedings of 12th Annual Technical Conference of the American Society for Composites, Dearborn, Michigan, pp. 592-601.

Cairns, D.S. and Lagace, P.A. (1989): Transient Response of Graphite/Epoxy and Kevlar/ Epoxy Laminates Subjected to Impact, AIAA J., 27:1590-1596.

Cairns, D.S. and Lagace, P.A. (1987): Thick Composite Plates Subjected to Lateral Loading, J. Appl. Mech. **54**: 611-616.

Christoforou, A.P. and Swanson, S.R.(1991): Analysis of Impact Response in Composite Plates, Int. J. Solids Structures, 27:161-170.

Kook, J. S., Suzuki, M., Adachi, T., and Ujihashi, S. (1992): Load and Strain Histories for CFRP Laminates under Low-Velocity Impact, JSME Int. J., Series I, 35: 159- I 64.

Qian, Y. and Swanson, S.R. (1990): A Comparison of Solution Techniques for Impact Response of Composite Plates, Composite Struct., 14:177-192.

Prasad, C. B., Ambur, D.R. and Starnes, J.H. (1993): Response of Laminated Composite Plates to Low-Speed Impact by Airgun-Propelled and Dropped Weight Impactors, 34th AIAA Struct. Structural Dyn and Mat Conf, La Jolla CA, Part 2, pp. 887-900.

Reddy, J.N. and Miravete, A. (1995): Practical Analysis of Composite Laminates, CRC Press, New York.

Computational Analysis of a Laminar Jet in a Cross-Flow

Damon K. Jeffries, R. Krishnamurthy and S. Chandra
NASA CAR, College of Engineering
North Carolina A&T State University
Greensboro, NC 27411

Introduction

A jet in cross-flow (JIFC) consists of a jet exhausting at a large angle into a freestream flow. It is a flow field which is relevant to a wide variety of technologies and applications. Despite the nearly 65 years of JIFC research there are few results available for laminar hypersonic flows, a combination which will be encountered by re-entry and high altitude vehicles over some portion of their flight path [1]. This research consists of developing a numerical model to investigate the interaction of a normal sonic jet exhausting into a hypersonic cross-flow. The model was validated by comparing experimental measurements with corresponding numerical results generated by the model.

Experimental Description

The experimental problem being modeled examines the interaction between a normal sonic jet of gas (Nitrogen, Helium, and Argon) and a hypersonic cross-flow of nitrogen over a flat plate [2]. In these experiments, performed in the Southampton University Light Piston Isentropic Compression Tube, a free stream of nitrogen at Mach 6.69 flows over an Isothermal Flat with a sharp leading edge. The freestream interacts with a gas injected at room temperature via a normal slot jet located 0.0745 m from the leading edge of the plate (figure 1). The experimenters defined the flow as fully laminar based on flow visualization methods and calculation of the Reynolds number at separation point for the non-injection flow providing no evidence of turbulence. Oil flow visualization and thermographs were used to define the flowfield. And surface pressure data for the region upstream of the jet was recorded.

Numerical Setup

The problem was modeled and analyzed by the General Aerodynamic Simulation Program (GASP). The CFD code GASP solves the fill Reynolds-averaged, compressible form of the Navier-Stokes, energy and species conservation equations [3]. The code was configured to run in a UNIX environment and all of the numerical computations were performed on the Cray-T916 system at the North Carolina Supercomputing Center. In GASP input decks are used to define the problem and describe a solution process. The full Navier-Stokes equations were used to capture all the dynamics of the flowfield. The boundary conditions selected were appropriate for a hypersonic freestream with sonic injection. The grid with which the numerical simulation and solution were performed had to be created outside of GASP. The grid was generated by a FORTRAN program, written specifically for the problem, and then attached to the numerical model. the two-dimensional grid was created with a domain large enough to capture the jet interaction phenomenon. An exponential stretching scheme was used to produce the appropriate grid clustering at the inlet, injection region, and surface of the plate (figure 2). And the grid contains 241 points in the x-direction and 141 points in the y-direction (initial grid setup). After the numerical model was completely setup, it was verified and then validated.

Results and Discussion

The numerical simulations have been performed and results recorded. A velocity vector plot was generated to show that the numerical model captured flowfield characteristics such as jet penetration, flow re-circulation, and flow separation (figure 3). A comparison of the surface pressure distribution in the region upstream of the jet was conducted to validate the model. The numerical results show good agreement with the experimental measurements for all injectants at an injection pressure of 142 KPa (figure 4). The same comparison was also performed with a injection pressure of 70 KPa, these results show better agreement was obtained for the separation point and separation length. The primary objective of the experimental study was to determine if the jet interaction could generate a jet induced force, which has applications to reaction control. The force is obtained by integrating the surface pressure distribution generated by the flowfield, where the largest positive contribution to the force results from the pressure distribution in the region upstream of the j et. The pressure induced force was calculated for the experimental and numerical studies and a comparison of the results made (figure 6). The comparison demonstrated that the numerical model was nearly as accurate in predicting the jet induced force [4].

There was an investigation into improving numerical agreement with experimental results. The first part of that investigation lead to examining model components. The components (input decks and grid) were analyzed and selectively modified in an attempt to improve numerical results. Model parameters (such as boundary conditions and inviscid fluxes) were modified and the grid domain and point number was increased. A modification of the grid only, resulted in improved agreement (figure 6). This result generated interest in examining the effects of the grid on numerical results. Various model setups have been examined to study the effects of grid configuration on the results of simulation. The setups included increasing domain size, increasing number of grid points, relaxing grid clustering, and other combinations of the setups. The results of the investigation show that the finest grid, without increasing domain size, produced the best results (figure 7).

Concluding Remarks

The numerical simulation of a laminar jet into a cross-flow has been performed and results analyzed. Good agreement between the experimental data and the numerical results has been achieved. Accurate prediction of the jet induced force has been obtained. The agreement of the numerical results with the experimental data has been improved with finer grid configurations. A study of the effects of grid sizes showed that while a finer grid setup does improve results, a finer grid with a relaxed clustering scheme may improve results further and at the same time decrease overall computational time. A numerical study of this problem with a fine, relaxed grid setup is recommended. Once numerical agreement is improved, a three-dimensional simulation of this problem would provide more detail of the flowfield and its characteristics.

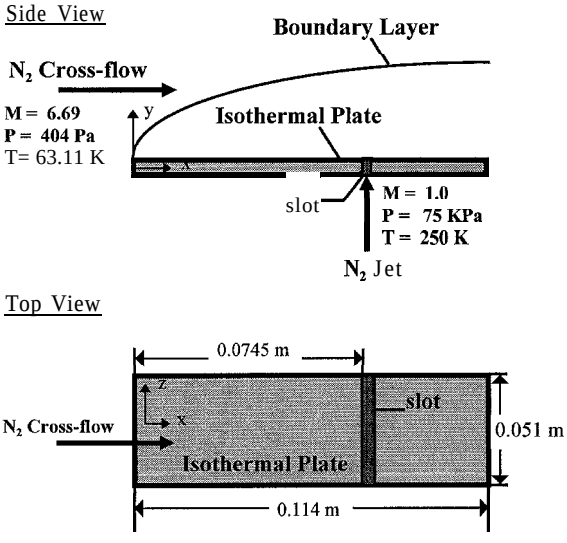


Figure 1. Schematic of experimental problem,

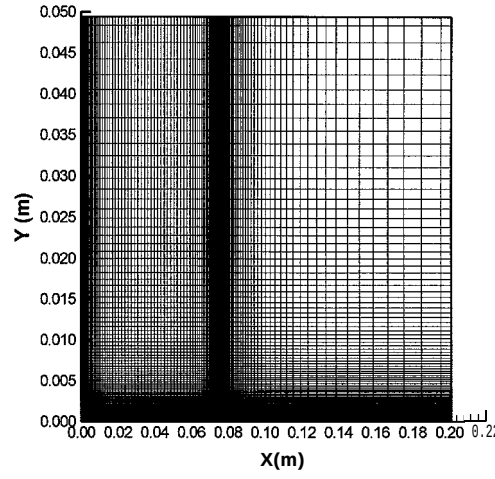


Figure 2. Computational Grid A.

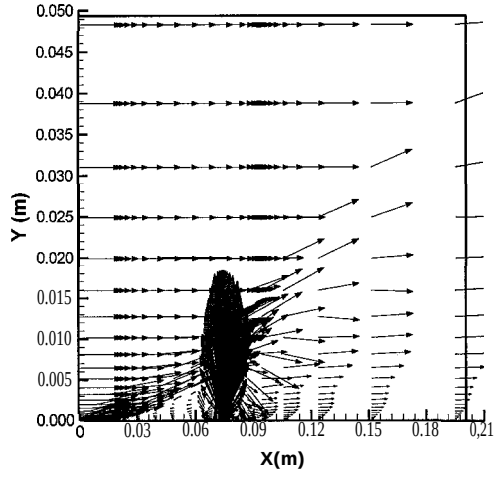


Figure 3. Injection case velocity vector plot over plate.

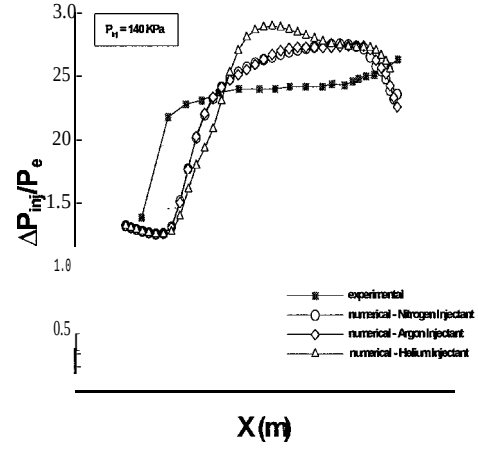


Figure 4. Surface pressure comparison ahead of slot.

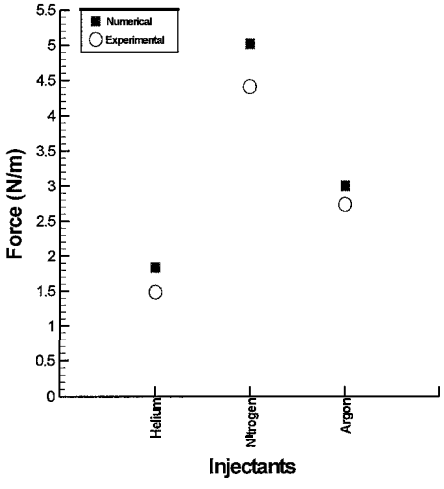


Figure 5. Plot of jet induce force for various injectants.

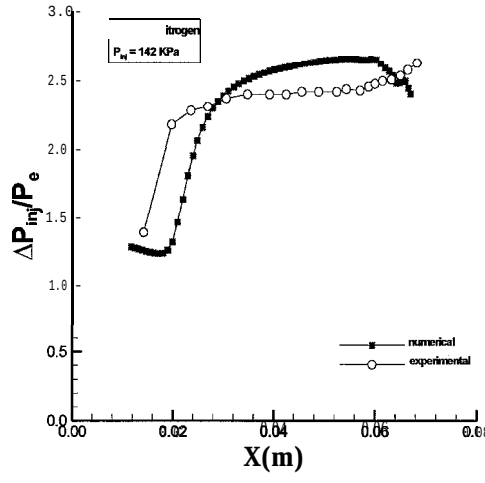


Figure 6. Surface pressure distribution with Fine Grid.

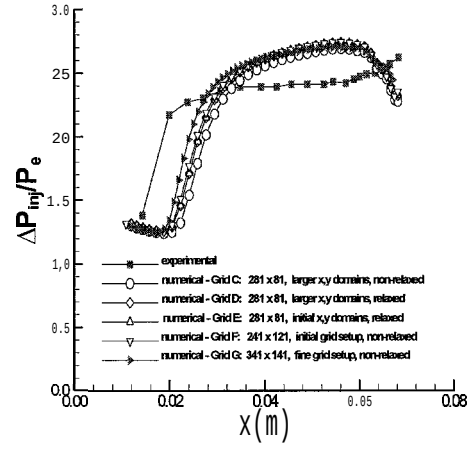


Figure 15. Effects of Grid Dependence on Numerical Results.

ACKNOWLEDGEMENTS

The authors of this paper gratefully acknowledge the support of the North Carolina Supercomputing Center for their award of computer time on the Cray-T90, Dr. Graham Ball for his assistance and the financial support provided by the grant NAGW-2924 from NASA Headquarters.

REFERENCES

[1] AGARD Conference Proceedings 534: Computational and Experimental Assessment of Jets in Cross Flow, North Atlantic Treaty Organization, November 1993.

[2] Powrie, H.E.G., Ball, G.J., East, R.A., Comparison of the Interactions of Two and Three Dimensional Transverse Jets with a Hypersonic Free Stream, AGARD Conference Proceedings 534, November 1993, pp. 20-1--20-8.

[3] McGrory, William D., Slack, David C., Applebaum, Micheal P., and Walters, Robert W., GASP version 2.2 User's Manual, Aerosoft, Inc., 1993.

[4] Jeffries, Damon K., Chandra, S., Krishnamurthy, R., A Numerical Simulation of a Normal Sonic Jet in a Hypersonic Cross-Flow, Thesis, August 1997.

98 URC034

SoftLab®: a Soft-Computing Software for Experimental Research with Commercialization Aspects'

M.-R. Akbarzadeh-T, S. Shaikh, J. Ren, Rob Hubbell, K. K. Kumbla, and M. Jamshidi
Autonomous Control Engineering
EECE Building
University of New Mexico,
Albuquerque, NM 87131, USA
E-mail: akbazar@eece.unm.edu

Abstract

SoftLab® is a software environment for research and development in intelligent modeling/control using soft-computing paradigms such as fuzzy logic, neural networks, genetic algorithms, and genetic programs. SoftLab® addresses the inadequacies of the existing soft-computing software by supporting comprehensive multidisciplinary functionalities from management tools to engineering systems. Furthermore, the built-in features help the user process/analyze information more efficiently by a friendly yet powerful interface, and will allow the user to specify user-specific processing modules, hence adding to the standard configuration of the software environment.

I. Introduction

Computers and computer software play an increasingly significant role in our modem world. This is particularly evident in research and innovation where complex systems are analyzed and controlled. Intelligent control/modeling is one of the research areas which has received a great deal of attention and interest among researchers in science and engineering in recent years due to its ability to incorporate human knowledge and reasoning ability in computing processes. The fundamental basis of such approaches are expert systems, fuzzy logic, neural networks, and evolutionary computations, collectively named the "soft computing" paradigms. Soft computing paradigms and their hybrid combinations with other approaches are commonly used to enhance artificial intelligence (AI) and incorporate human expert knowledge in computing processes such as in the design of intelligent autonomous systems/controllers and handling complex systems with unknown parameters such as in prediction of world economy, industrial process control, prediction of geological changes within earth ecosystem, and robot control in space for scientific research or on earth for handling hazardous waste.

Even though soft computing paradigms can potentially provide answers to a wide array of today's problems, there is not yet a comprehensive software environment which can address the multidisciplinary functionalities expected from soft computing paradigms. Currently, there are several software environments in the areas of fuzzy logic, neural networks, and evolutionary algorithms. These include *Togai Infra Logic [1], Fuzzy Inference Design Engine* (FIDE) [2], *Fuzzy Logic Development Kit* (FULDEK) [3], and O *'INCA [4].* These software environments, however, are each developed for only a limited capability and are application specific. None of the current software allow users to improve the functionality of the software by incorporating new user-defined modules. They do not utilize all of the soft-computing paradigms in a single software environment. And finally, none of them support add-on hardware modules/processors for real-time experimentation and improving computational efficiency. As a result, researchers, in the area of intelligent systems, often devote a significant amount of resources to develop their own application specific software and data acquisition hardware, making it time-consuming and costly to perform research and development in this area.

This paper addresses the inadequacies of the current software by proposing a comprehensive soft-computing based software environment which can address multidisciplinary functionalities from management tools to engineering systems. The software environment, SoftLab®, is written using Visual C++ for a user friendly environment. It is a general purpose development/analysis environment where the inadequacies of the existing soft-computing software is addressed. Furthermore, the built-in flexibility in SoftLab® help the user process/analyze information more efficiently by a friendly yet powerful graphical user interface (GUI), and allow the user to specify

[1]This work was supported in part by NASA contract # NCCW-0087 and by Waste-management Education and Research Consortium (WERC) contract # WERC/NMSU/DOE Amd 35.

user-specific processing modules, hence adding to the standard configuration of SoftLab®. It also allow applications in a wide range of directions thereby easily adding artificial intelligence to areas where AI has not yet been explored.

In the following section, the objectives of the software are outlined. Section III presents the β–version of the software which was released in July of 1997. In section IV, we shall discuss the next version of the software, expected to be released in May of 1998. Section V addresses the Universal Data Acquisition Interface (UDAI) for performing experimental research. Section VI illustrates the commercialization aspect of this software.

II. Objectives of SoftLab®

The proposed computationally intelligent environment revolves around five main objectives:

1. Integrating Soft Computing Techniques

Modeling and control of complex systems is often performed in a hierarchical decision making process. This hierarchy could either comprise of several layers of fuzzy rule-sets, or an interrelated hybrid combination of fuzzy logic, neural networks, genetic algorithms and genetic programs. In the literature, hybrid combinations of various soft computing paradigms is helping solve a wide array of problems.

SoftLab® provide a number of soft-computing tools (such as fuzzy logic rule-set, back propagation neural networks, and genetic algorithms). The user simply selects, drags, and drops the tool on the main window. After double clicking, a new window appears where the user may enter parameters of the tool through easy to use dialog boxes. For example, in case of a fuzzy rule set, parameters are membership functions and rule structure. In case of genetic algorithms, the parameters may be size of population, gene selection mechanism, rate of mutation, etc.

2. *A Modular System for Future Development*

Success of a research environment depends to a great extend on its ability to incorporate new ideas. Even though the proposed package is equipped with a variety of predefined tools, it cannot possibly contain the many different schemes in the literature or schemes which will be developed by other researchers in the future. As a result, the software environment is designed in a modular fashion such that new tools could to be attached to the software.

3. *Interface to Real-Time World and Computational Efficiency*

One of the functions of SoftLab® is the real-time interface to physical systems for the purpose of data acquisition and process control. This warrants a need for standalone processor and hardware equipment. The control and data acquisition can be performed by new digital signal processor which have proven to have performed such tasks in a cost efficient way. We have already interfaced Texas Instruments digital signal processors and data acquisition circuits for the purpose of process control. The hierarchical hybrid fuzzy control also has been applied to a small scale desalination plant using data acquisition boards.

4. *Interface to Knowledge Base*

Applying intelligent techniques to analyze data can provide valuable information which can be used in many ways. In the area of Science and Engineering intelligence tools such as fuzzy logic, neural networks and genetic algorithms have been applied successfully in a variety of applications such as image processing, process control modeling of dynamic characteristics of physical systems etc. But in the area of financial and industrial data base not many applications of using intelligence techniques are cited. The SoftLab® tools can provide the means for doing just this purpose. SoftLab® can be interfaced with data base management system (DBMS) tools such as Oracle, Microsoft Access, DB2 etc. SoftLab® can be used to apply fuzzy logic tools to a conventional data base by adding a fuzzy querying function compatible with human reasoning. SoftLab® can also be used to build a fuzzy database with imprecision and vagueness represented by fuzzy or possiblistic elements, with querying and updating.

To illustrate such an idea consider in a water quality context an user requires to know all points in a watershed where the environmental pollution is *severe.* It is however very difficult to adequately represent *severe* by a conventional query with the ANDing and ORing of attributes only, and an adequate representation of severe (pollution) when, e.g., *most of the important* pollution indicators *considerably* exceed some (possibly fuzzily specified) limits. Notice that a linguistic quantifier *most* cannot be accommodated in a conventional DBMS. Similarly a statement ``Amount in Stock is *Much Greater Than* Amount on Orders" exemplifies an imprecise relation which can be easily quantified using fuzzy operators.

SoftLab® provide an easy interface to DBMS software, and will have the capability of analyzing data using variety of tools ranging from fuzzy querying functions to neural network based pattern recognition to predict for example the performance of a company, trends in stock market, etc.

5. *Relation to Longer Term Goals*

A proper software environment which can cater to innovative ideas of a researcher is a valuable asset and can significantly help success of any project. The team of investigators intend to use the results of this project to solve several problem with which they are currently involved. These include control of a water desalination process, image processing for classifying satellite data, robot control in hazardous waste, intelligent navigation of mobile robots. The versatility of these projects demonstrates the extent of utility of the proposed software environment.

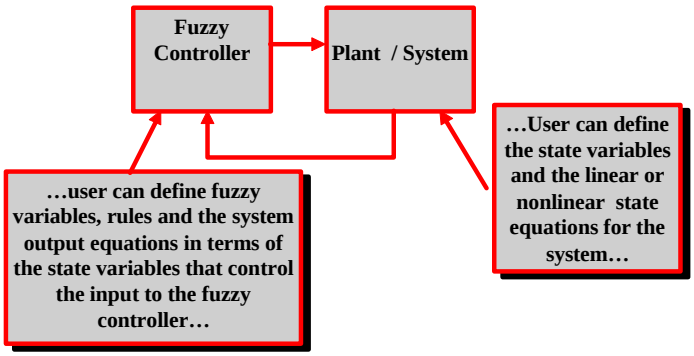


Figure 1. Simulation of a Control System

III. SoftLab® 1.β

SoftLab® 1.β simulates fuzzy logic control systems, as depicted in figure 1, of the following properties:

- The system:
  1. The system can be either *linear* or *nonlinear* and of any desirable dimension.
  2. The system equations can only contain mathematical functions defined in Visual C++ and must be presented in state space form.
- The controller:
  1. Controller should be a Fuzzy Logic Controller (FLC).
  2. FLC can have any number of rules and membership functions.
  3. Membership functions are triangular shaped.
  4. Antecedents are joined using AND and/or OR.
  5. Min/Max is used for rule evaluation.
  6. Center of Gravity is used for defuzzification of rules.

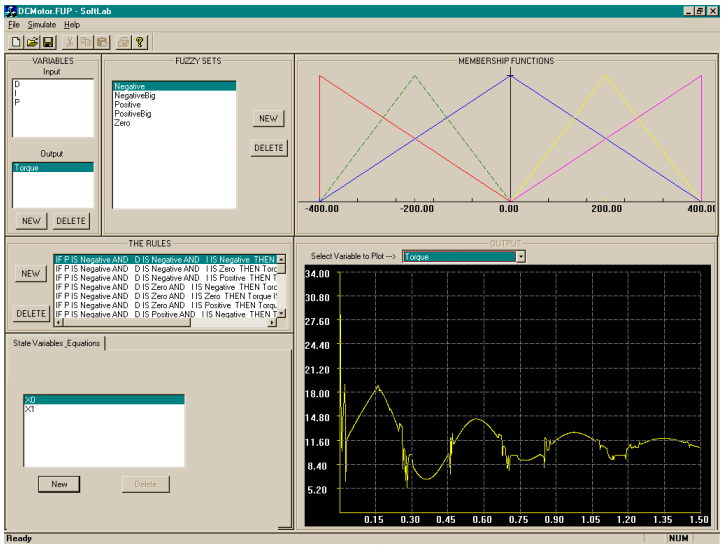


Figure 2. A snapshot of the current version of SoftLab®.

IV. Future Versions of SoftLab®

The future version of SoftLab® will be capable of simulating a system similar to the following block diagram.

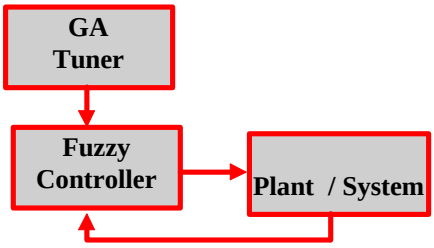


Figure 3. Block Diagram of a sample system which can be simulated with future version of SoftLab®.

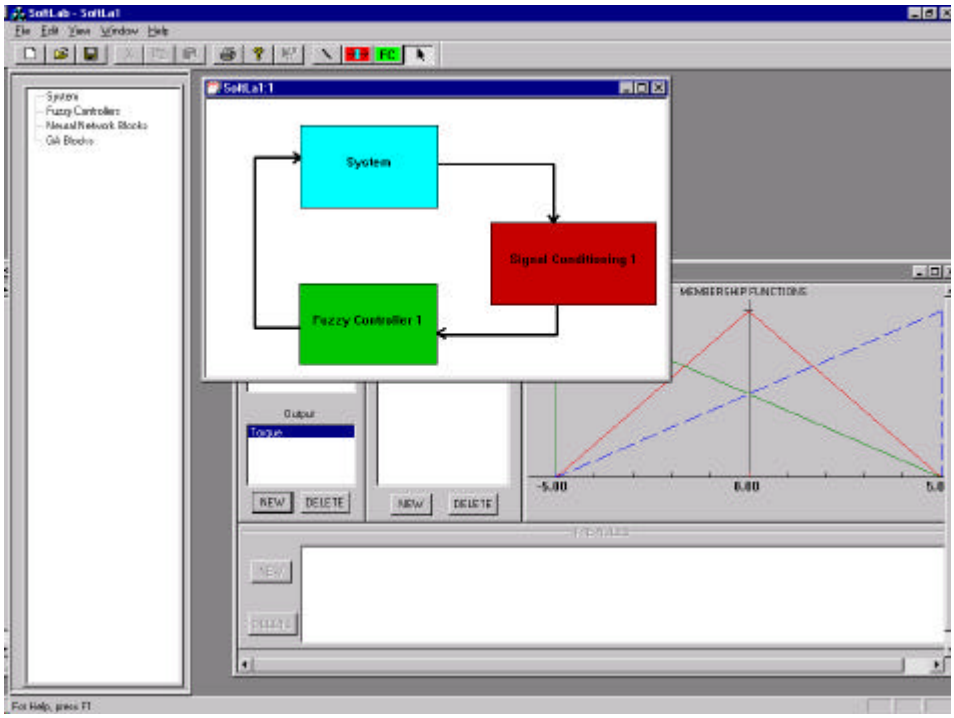


Figure 4. Snapshot of the future version of SoftLab® while under development

The following features will be added to the future versions of SoftLab®.

- Integrate fuzzy logic, neural network and evolutionary algorithms in one development environment where adaptation and machine learning can be performed.
- Develop a modular system where a third party can integrate an application specific module to the software environment. As a result, other methods of artificial intelligence can be incorporated in the future.
- Capability to add hardware modules such as Digital Signal Processors (DSP), parallel processors, data acquisition hardware, and remote processors through the Internet. This will address two essential purposes:
    - Addresses computational efficiency through multi and parallel processing environment.
    - Allows real-time experimentation and data gathering.
- Interface to knowledge based software and multi-media tools. The strength of a given software is increased significantly if it can take advantage of the computational power of other existing software. One of our objectives is to add soft-computing techniques to data-base management tools such as Oracle.
- Application areas such as in control systems, data-base management, dynamic simulations, and real-time experiments.

IV. The Universal Data Acquisition Interface for Experimental Research

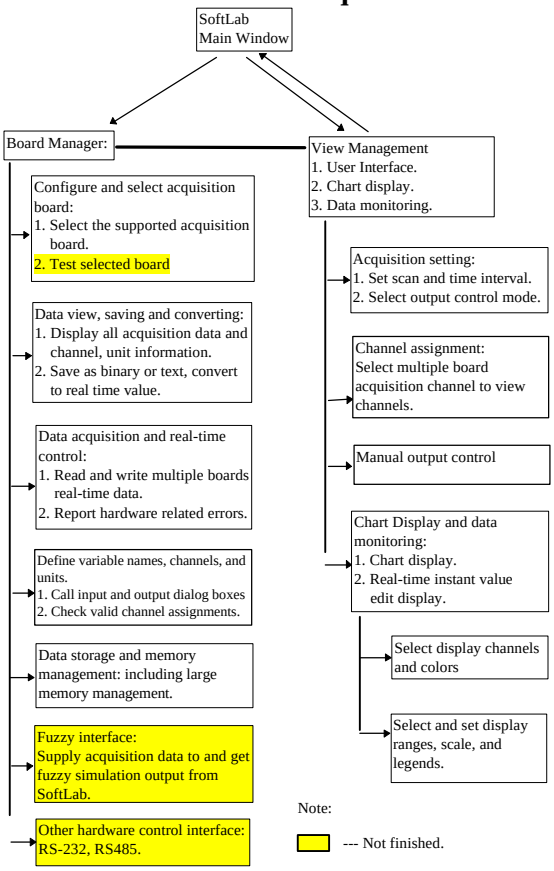


Figure 1. Experimental Controller Function Block Diagram

UDAI is mainly developed as the real time control part of the SoftLab® environment. The main functions are shown in the block diagram below:

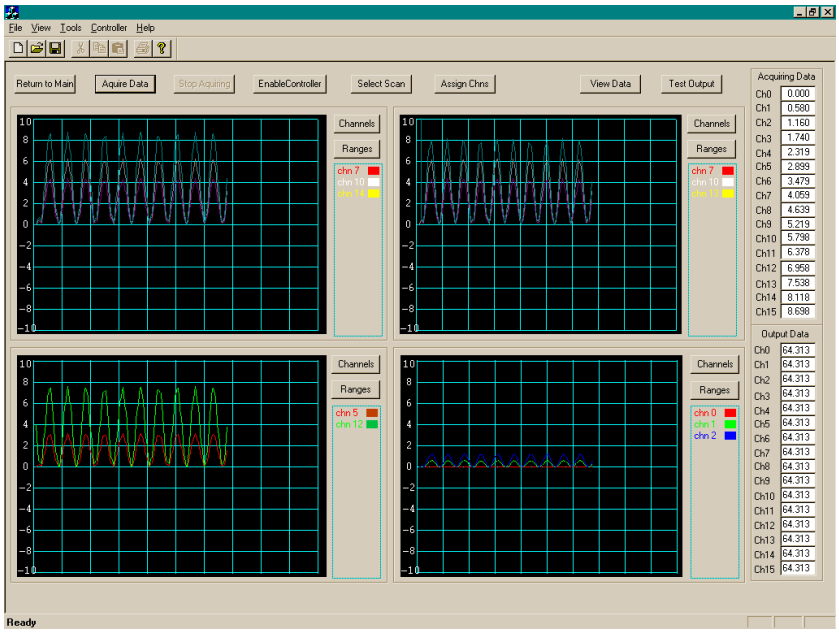
I. Board Manager:
This is an object class that links all the other components together. As shown in the diagram above, all the hardware controls are finished through this class while all the displays are managed by the view manager. When an acquisition board is selected or default board is used, choosing "Go real-time" in the SoftLab® main menu enables a control view window.

The functions through board manager are as follows:

1. Select acquisition board.
   This is an dialog box class that list all the hardware boards supported by the experimental controller. By default, two acquisition boards DAS08 and DAS16 are selected. Selecting more than two boards at a time is possible.
2. Data view, saving and converting
   This class is implemented to make the data more convenient for acquisition purpose. By choosing "ViewData" in the main control window, the data view dialog pops out. And the acquisition data for each channel is display according to its real time value (such as degree, Volt, etc.). The channel number and the unit set in the "DefineName" dialog box which come out here are associated with the corresponding channel. User can also select "SaveBinary" or "SaveAsASC" here. If "SaveBinary" is used, the original acquisition data is saved to the name entered or selected from the "FileSave As DialogBox". If "SaveAsASC" is used, the data is saved as ASCII text format along with the data index and channel information. Thus the file can be viewed by any word processor.
3. Data acquisition and output control
   The library functions supplied by Computer Board Corp. are called to do the acquisition and output control (according the control mode setting).
4. Define variable names, channels, and units
   This class is used to define variable associated with an input or output channel used for fuzzy control. At the same time, for the input channels, the units for real time value can also be enter here. And the setting points for the input channels are also set here. The units of real time variable are displayed in "ViewData" dialog box and the setting points are used to convert the data in fuzzy control, or data display. But the chart display the real voltage in the input channels. Thus give more information about real time acquisition and control.
5. Buffer and data storage management.
   Buffer is automatically managed according to the number of acquisition scans and the number of acquisition channels requested by user. Larger memory use is available (up to 4Mb at a time).
6. Fuzzy control interface
   Real time input is fed into SoftLab® fuzzy simulation packet to process and then the fuzzy output result is written out to the DA converter boards and the outside controllers.
7. Board configuration
   This class is used to configure different hardware parameters for the selected acquisition boards.

8. RS-485 and RS-232 port control (under development)

Since serial ports are different from memory boards, multiple threads are needed to execute these functions, which are under development.



II. View Management

This is the main controller view class. The main functions in this class include chart display, setting scans, setting control mode, manual output control, calling "ChartClass" to select channel color and displaying different channels. Also routes user interface to the "Board Manager".

1. Acquisition scan setting and control mode.

This dialog box class is used to set the scan rate and the total scan for one chart cycle. There are two modes for scan: *once* and *continuous*. If *once* is used, after one cycle, the acquisition stops. If *continuous* mode is chosen, the acquisition runs until the user pushes the "Stop" button. The *output control* mode is used to select different fuzzy control schemes or other control methodologies (PI, PD, PID etc.).

2. Assigning channel

If multiple acquisition boards are used at the same time, this dialog box class is used to select those channels are currently used. Sometimes this is useful in control. For example, by assigning channels, output or input can be routed to without redefining the control model.

3. Chart display

The functions in this class include setting the channel colors, selecting which channel to display, and setting the range of the display. As stated above, the display units are in voltages. The color can be selected which are available in the window system. Up to four charts are displayed in the main control view window.

V. Commercialization

This project would result in a powerful soft computing environment for experimental and theoretical research. The resulting software promises to become an invaluable asset for investigators, from a wide background, who wish to integrate soft computing technologies easily and efficiently. The environment will also become a standard medium for exchange of ideas and collaborations since new algorithms can be added. A national minority-owned company, head quartered in New Mexico, MediaTronix, Inc., in close collaboration with University of New Mexico's Science and Technology Corporation (STC), will be entrusted with the task of commercializing such potential product.

VI. Conclusion

SoftLab® is intended to respond to a growing demand for a user friendly, dynamic and powerful software environment for research in intelligent control. In this paper, various current and future aspects and features of SoftLab® are discussed. SoftLab® not only provides a good simulation environment, but also has several features which facilitates experimental research. At this time, only a limited test version is released for the purpose of testing and debugging. The next version of SoftLab® is expected to be ready for release to the public.

VII. References:

[1] Fuzzy C Development System, User's Manual, Togai Infra Logic Inc., 5 Vanderbilt, Irvine, CA, 92718, 1992.

[2] Fuzzy Inference Design Environment, Aptronix Inc., 2150 North First Street, Suite 300, San Jose, California 95131, 1992.

[3] Fuzzy Logic Development Kit, User's Manual, Bell Helicopters Textron Inc., P.O. Box 482, Fort Worth, Texas 76101, 1992.

[4] O'INCA Design Framework, User's Manual, Intelligent Machine Inc.

98 URC035

SOFT COMPUTING BASED APPROACH TO COOPERATIVE AUTONOMOUS ROBOTICS

Shannon M. Delgado, Denise D. Padilla, Marco A. A. de Oliveira, Rene Morris, Oracio Garcia, Olivier Pages
NASA Center for Autonomous Control Engineering
Department of Electrical and Computer Engineering
University of New Mexico, Albuquerque, NM 87131 USA
robosoccer@unm.edu

ABSTRACT

A hierarchical fuzzy system, optimized through evolutionary methods, is investigated as a controller for multiple cooperative autonomous robots. The use of multiple agents as opposed to one monolithic robot has the main advantages of lower production cost, robustness through redundancy, and reduced size and weight. Multiple robots offer a higher degree of operational robustness by ensuring the completion of all or most of the assigned tasks under unforeseen (adverse or otherwise) conditions. Such conditions may include robot hardware failure or loss of mission capability. The robot agents employ a hierarchical fuzzy system that consists of several layers of interconnected fuzzy logic controllers. This fuzzy system will consist of primitive behaviors, social behaviors, an advisory program, and a training program. The primitive behaviors consist of motion, obstacle avoidance, navigation, and telemetry. The social behaviors are generated by the interaction among robots and include identification, communication, and cooperative goal reaching. The advisory program is responsible for facilitating the execution of the task of the robot and the training program determines near-optimal or optimal social behaviors using genetic algorithms. The system is validated in a benchmark problem which incorporates many" of the real-world issues confronting autonomous mobile robots in dynamic and unstructured applications such as terrain exploration and hazardous site maintenance.

KEYWORDS: fuzzy controller, genetic algorithm

INTRODUCTION

The robots are to collectively exhibit real time reasoning and inter-team collaboration. Inherent in this is the need for the robots to autonomously identify situations, evaluate them, and react as a human would. Therefore, a fuzzy logic controller which mimics human reasoning by using rules and incorporating concepts of imprecision in knowledge representations, is ideal for this purpose [1].

The robot agents employ a hierarchical fuzzy system that consists of several layers of interconnected fuzzy logic controllers (Figure 1). Each controller represents a behavior and from their interaction emerges the overall robot behavior. The lowermost modules

correspond to the primitive behaviors displayed by the agents, such as obstacle avoidance, and motion. As one advances up the hierarchical network, the fuzzy controllers encode more complex behaviors, affecting the response of the modules localized below them. The more advanced functions are responsible for behaviors such as identification, communication, and cooperative goal reaching.

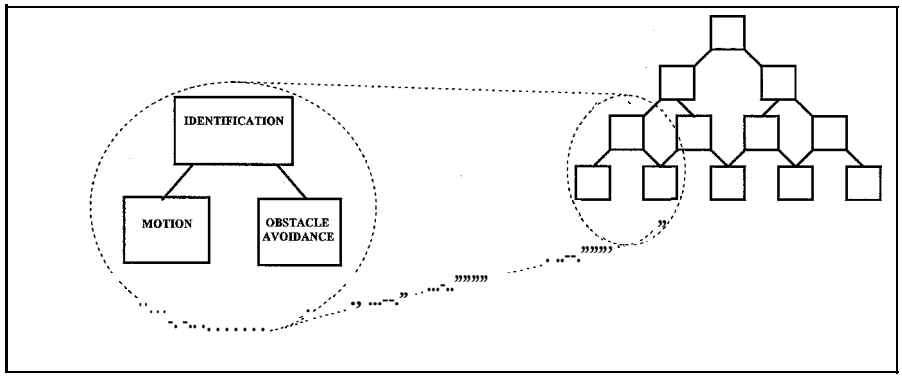


Figure 1: Hierarchical Fuzzy Controller

A genetic algorithm is an optimization method based on principles taken from theories of natural selection and evolution, and is used for determining optimal or near-optimal rule bases. An initial population is formed by randomly altering a common-sense set of fuzzy rule bases created by the researchers [2]. Each search point, an individual of the population (in our case a fuzzy rule), represents a candidate solution and is evaluated according to a fitness function. The fitness function assigns higher values to desired characteristics.

Data concerning the robots and their positions, as well as cooperative goal reaching information, is then relayed to the robots through a RF link. Cooperative goal reaching information, generated by the laptop computer, consists of directives that affect the robots higher level behavior modules. The analysis of the task and the emission of directives is realized by a different hierarchical fuzzy system running on the laptop. A second genetic algorithm will be employed to optimize the components of this tactical fuzzy system.

FUZZY BEHAVIOR

Autonomous navigation behavior can be decomposed into several interconnected special purpose behaviors. The interconnection of the behaviors is a hierarchical network. It is hierarchical in the sense that the overall system behavior is decomposed into a bottom-up hierarchy of increasing behavioral complexity. The behavioral activity at a given level is a function of the behavioral activities at the level(s) below. The lowermost modules, referred to as the primitive behaviors, are responsible for the simple, self-contained behaviors such as move and avoid. The intermediate behaviors, located in the mid-section of the hierarchical structure, interact with the primitive behaviors to perform

specific tasks, such as identification and communication. The primitive behaviors map inputs to control outputs (typically common actuators), while higher-level behaviors act as fuzzy decision systems which map goal and strategy information to dynamically adaptive weights associated with each primitive behavior [2].

Within each behavior module or fuzzy controller, is a fuzzy rule base which contains linguistic knowledge of the problem space in the form of IF-THEN rules. Let X and Y be

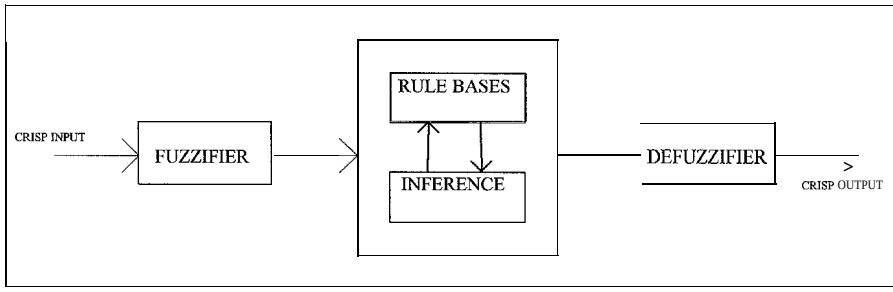


Figure 2: Process of a Fuzzy System

the input and output universes of discourse of a primitive behavior with a rule-base of size r. The generic fuzzy IF-THEN rule can be described as follows

$$\text{IF } x_1 \text{ is } A_1^k \text{ and } x_2 \text{ is } A_2^k \text{ THEN } y^k \text{ is } B^k \text{ for k= 1,2, . . ., r}$$

where A_1^k and A_2^k are the fuzzy sets representing the k^{th} antecedent pairs, and B^k are the fuzzy sets representing the k^{th} consequent [1]. This is analogous to a dual-input, single-output fuzzy system where each rule comprises two antecedents and one consequent and can be easily extended to fuzzy rule bases (or fuzzy systems) with any number of antecedents (inputs) and consequent (outputs). The process of a fuzzy system is illustrated in Figure 2. The crisp input set, sensory data or goal information, triggers all the rules in a parallel fashion for more efficient computing time. To conduct the inferencing of the set of disjunctive IF-THEN rules we have chosen the graphical Mamdani (max-min) inference method. The aggregated output for the r rules will be given by

$$\mu_B^k(y) = \max_k [\min [\mu_{A1}^k(\text{input(i)}), \mu_{A2}^k(\text{input(j)})]] \quad k = 1, 2, \ldots, \quad r$$

Figure 3 shows the graphical analysis of two rules, where A_{11} and A_{12} refer to the first and second fuzzy antecedents of the first rule, and B_1 refers to the fuzzy consequent of the first rule. Similarly, A_{21} and A_{22} refer to the first and second fuzzy antecedents of rule 2, and B_2 refers to the fuzzy consequent of the second rule [1]. The minimum function in the above equation arises because the antecedent pairs given in the rule structure for our system are connected by a logical “and” connective. The minimum membership value of the antecedent pair propagates to the consequent and truncates the membership function. The inference procedure is done for each rule and the truncated membership functions are aggregated using the graphical equivalent of the definition for disjunctive rules. The definition for rules that are connected by the logical connective “or” is the union of all

individual rule contributions. This results in an aggregated membership function comprised of the outer envelope of the individual truncated membership functions from each rule. The crisp value of the aggregated output can be found through a defuzzification technique known as the centroid method. This process is distributed throughout the entire hierarchical network of fuzzy controllers.

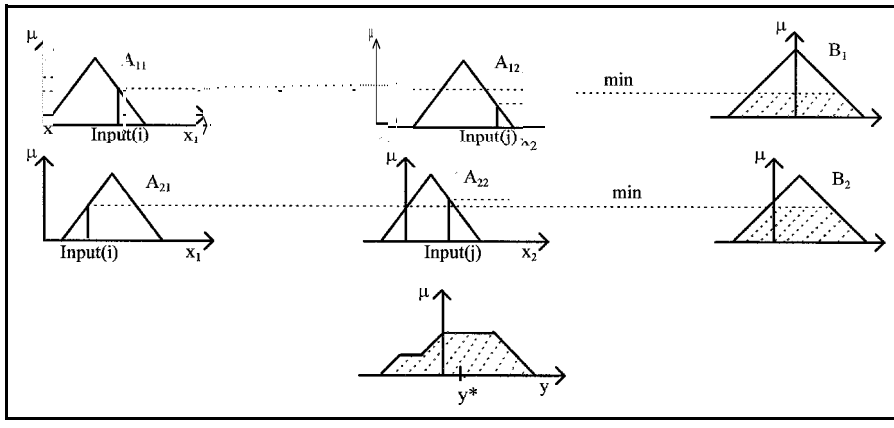


Figure 3: Graphical Mamdani (max-min) inference method with crisp inputs.

GENETIC ALGORITHMS

Genetic algorithms use the concept of Darwin's theory of evolution. The theory is based on the fact that the existence of all living things is based on the rule of "survival of the fittest" [1]. New classes of living things come into existence through reproduction, crossover, and mutation among existing organisms. First, an initial population of different possible solutions is created by randomly altering a common-sense set of fuzzy rule bases created by the researchers. Each search point, an individual of the population (in our case a fuzzy rule), is tested for its performance according to a fitness function. Among the possible solutions a fraction of the solutions is chosen, according to a cutoff value, to succeed into the next generation while the others are eliminated. This is the concept of survival of the fittest. The selected solutions undergo the processes of reproduction, crossover, and mutation to create a new generation of solutions that are expected to perform better than the previous generation. New generations are produced and evaluated repeatedly until there is convergence within a generation.

First, in our application, the parameter set of the problem that characterizes the solution is coded in the form of a finite string of bits. For example we create independent bit strings for the antecedent and the consequence of the fuzzy if-then rules, and then the strings of each parameter are concatenated to make one string that represents the whole parameter set. The length of the bit strings is based on the handling capacity of the computer being used. The strings are created randomly and then decoded, or mapped,

into a set of parameters they represent. This set of parameters is then passed through a numerical model of the problem that yields a solution based on the input set. The quality of the solution is evaluated according to a fitness function and a fitness value is assigned to the string. The fitness values of each string in the population are determined and an average is computed. The relative fitness of each string is determined by dividing the fitness value of each string by the average fitness value.

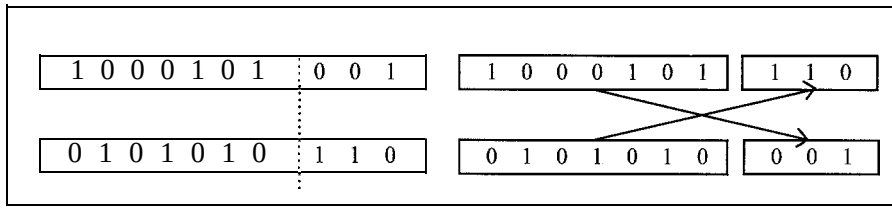


Figure 4: Crossover in strings.

Reproduction is the process by which strings with higher fitness values are duplicated in the next generation. The total number of strings must remain constant for computational efficiency, therefore strings with fitness values below the cutoff are eliminated. After reproduction, crossover occurs between two new randomly selected strings at a random location (Figure 4). The bits to the right of the location are interchanged as a way of information exchange and combining of desirable qualities. These new strings comprise the new generation and the process of decoding and evaluating is repeated.

Although most of the searching power is involved with reproduction and crossover, mutation is important when an optimum solution cannot be found. Mutation changes the value at a certain string location and is rare (takes place on the order of once in a thousand bit string locations). The nature of our problem space is such that there are an infinite number of possible solutions, and genetic algorithms search for a solution from a broad spectrum of possible solutions.

CONCLUSION

Genetic algorithm optimized hierarchical fuzzy controllers are employed as a behavioral structure for multiple autonomous cooperative robots. The use of small robots instead of a monolithic approach offers more flexibility, through easier reconfiguration, and robustness (distributed and cooperative processing, compensating for hardware faults). Neural networks, a computing method based on simplified models that simulate the working network of the neurons in the human brain, are being considered as interfaces between the internal fuzzy hierarchy and the environment. The neural networks would be used for encoding the fuzzy membership fictions and adapting them to changes in the environment.

REFERENCES

1. Ross, Timothy J., *Fuzzy Logic with Engineering Applications.* McGraw-Hill, Inc., New York, NY (1995);
2. E. Tunstel and M. Jamshidi, "On Genetic Programming of Fuzzy Rule-Based Systems for Intelligent Control", International Journal of Intelligent Automation Soft Computing, Vol.2, No.3;

98 URC036

THE DESIGN OF STABLE NONLINEAR CONTROLLERS FOR A CLASS OF NONLINEAR PLANTS BASED ON NEURAL NETWORKS

Olawale Adetona*, Ephrahim Garcia+
Sivapragasm Sathananthan*, and Lee H. Keel*

*NASA Center for Automated Space Science
Tennessee State University
Nashville, TN 37203-3401.
'Mechanical Engineering Department
Vanderbilt University
Nashville, TN 37235

ABSTRACT

Extensive empirical evidence has been published in the literature to demonstrate the enormous potential of multilayer neural network controllers. In spite of these results, practical implementation of these control schemes has been held back by the lack of an analytical proof of the stability of the controlled system. The objective of this paper is to present a nonlinear control scheme based on multilayer neural networks for the control of a class of nonlinear plants. The stability of the closed loop system will be rigorously established.

KEYWORDS: Neural Networks, Control, Stable, Popov Criterion.

INTRODUCTION

Most of the papers that discuss the use of artificial neural network controllers have been largely restricted to proposing new applications, new control schemes or new neural network architectures. Some researchers have attempted to analyze the stability of these systems by carrying out extensive empirical studies. The few analytical results available in the literature are usually restricted to the analysis of systems in which either the neural network weights or the control input appears linearly in the system equations. This includes the analysis of feedback linearizable nonlinear systems controlled by radial basis function neural networks. The objective of this paper is to present a method of designing stable control systems based on multilayer neural network (MNN) controllers for a class of nonlinear plants. The stability of the closed loop system can be rigorously established. The class of nonlinear dynamic systems that will be considered are a combination of SISO Hammerstein and SISO Wiener nonlinear systems whose linear plants are open loop stable.

The Popov Criterion

The Popov criterion is based on the Lur'e problem. The Lur'e problem is concerned with obtaining sufficient conditions for the stability of the class of nonlinear dynamic systems shown in Figure 1. This class of nonlinear systems have a linear dynamic system in the forward path and a nonlinear gain in the feedback path. Popov proposed a frequency domain criterion for absolute stability of this type of control system. The Popov criterion has a geometric interpretation which makes it easier to use than the direct method of Lyapunov. The absolute stability criterion due to Popov is stated below.

Theorem 1: If the closed loop system of Figure 1 consists of a stable linear part with a transfer function G(s), all poles of which are in the left half of the s plane and a single valued, time invariant, non-linear function ~(o), where $\frac{f(\sigma)}{\sigma} \in [0, k]$ for $\sigma \in (-\infty, +\infty)$. Given that $\alpha \equiv$ O, a sufficient condition for stability of the closed loop system is:

$$\mathrm{Re}\left[(1 + j\delta\omega) \cdot G(j\omega)\right] + \frac{1}{k} > 0, \text{ for } \omega \in (0, \infty) \qquad (1)$$

where δ is an arbitrary real number, $k <$ c-, and $\lim_{\omega \to \infty} G(j\omega) =$ O.

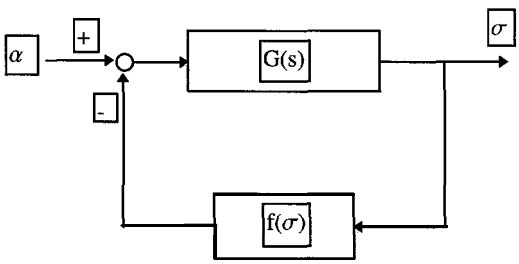


Figure 1: Lur'e System

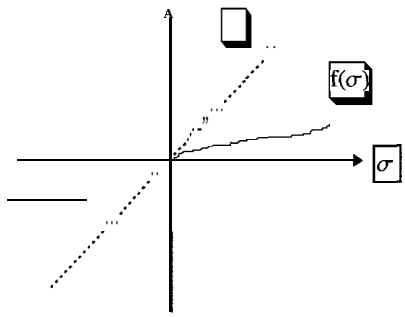


Figure 2: Sector Bounded Nonlinear Gain

Geometric Interpretation Of The Popov Criterion

Let $G(j\omega) \equiv u(\omega) + jv(\omega)$. **(2)**

Therefore, from Equation (1):

$$\mathrm{Re}\left[(1 + j\delta\omega) \cdot (u(\omega) + jv(\omega))\right] + \frac{1}{k} > 0 \qquad (3)$$

implies

$$u(\omega) - \delta\omega v(\omega) + \frac{1}{k} > 0. \tag{4}$$

Consider the following transformed frequency response G_t (co):

$$G_t(\omega) := u(\omega) + jv_t(\omega), \tag{5}$$
$$v_t(\omega) := \omega \cdot \mathbf{v(o).} \tag{6}$$

Also consider the straight line $r(\omega)$ in the Popov plane (u(o), $j\omega \cdot v(\omega)$ plane):

$$r(\omega) \equiv u(\omega) + jv_o(\omega), \tag{7}$$
$$v_o(\omega) := \frac{1}{\delta}\left[u(\omega) + \frac{1}{k}\right]. \tag{8}$$

Therefore, the Popov criterion requires that the Locus G,(o) must be to the right of the straight line $r(\omega)$. Note that $r(\omega)$ has an intercept of $-\frac{1}{k}$ on the $u(\omega)$ axis and a gradient of $\frac{1}{\delta}$. Finally, the absolute stability criterion can not be used directly if the linear part is neutral or unstable. If the linear part is unstable, the system will be unstable with any non-linear element.

CONTROL PROBLEM

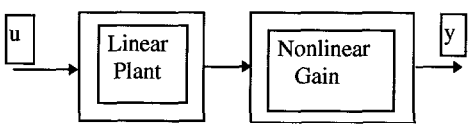


Figure 3: Wiener Plant

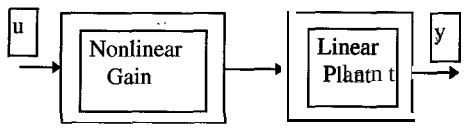


Figure 4: Hammerstein Plant

The nonlinear dynamic system is assumed to be a SISO Wiener or SISO Hammerstein nonlinear system. These are nonlinear dynamic systems made up of a linear dynamic system g(s) in series with a nonlinear gain $f_p(.)$, where g(s) is a SISO transfer function ands is the Laplace variable. The nonlinear gain $f_p(\cdot)$ is assumed to be: known, time invariant, memoryless and piecewise continuos with respect to time t. The objective of the control scheme is to control the plant so that the output of the plant *y(t),* will track a specified trajectory $r(t)$. In addition to accurately tracking the reference trajectory, it is also necessary to guarantee the stability of the closed loop system. Furthermore, a solution to the controller design problem must be known to exist for this problem to be

well posed. Therefore, the linear part of the unknown dynamic system g(s) is assumed to be known, observable, controllable and open loop stable. The nonlinear gain associated with the linear plant $f_p(\cdot)$ *is also* assumed to be a known sector bounded nonlinear gain.

The MNN controller is assumed to be implemented with a digital controller located in a feedback configuration with the nonlinear plant. The digital controller is assumed to be preceded by an analog to digital (AD) converter and followed by a digital to analog converter (DA). The sample time is assumed to be small enough to make the effect of the converters on the dynamics of the closed loop system negligible.

If the nonlinear plant is a Wiener system, the controller is made up of two parts. The first part is an integrator that is placed before the controlled plant. The second part is a nonlinear feedback gain $N(\cdot)$ that is implemented by a MNN. It is assumed that there exists a known sector condition *k* such that G(s) and *k* satisfy the Popov criterion, where G(s): = $\frac{1}{s}$ g(s) . If the nonlinear plant is a Hammerstein system, the controller will be made up of only $N(\cdot)$. However, g(s) will be required to contain an integrator term. For this system, it is assumed that there exists a known sector condition *k* such that G(s) and *k* satisfy the Popov criterion, where $G(s)$:= g(s).

The objective of the control scheme is to control the time invariant nonlinear plant using a nonlinear feedback gain implemented by a MNN. The stability of the closed loop system can be guaranteed if the combination of the nonlinear gain implemented by the neural network $N(\cdot)$ and the nonlinear gain $f_p(\cdot)$ satisfy sector conditions obtained from the Popov criterion. This means $N[f_p(g, u)]$ and $f_p[u - N(y)]$ must lie within the sector *k* obtained from the Popov criterion based on G(s) for the Wiener and Hammerstein nonlinear plant respectively, where, u is the input to the closed loop system and y is the output of the nonlinear plant. The usual sum of the squares error objective function used in the dynamic backpropagation algorithm (Narendra, 1997, pp. 80-98 ; Parthasarathy, 1990, pp. 75-91) should be replaced by an interior penalty function when the weights of the MNN are being computed. The constraint used in the interior penalty function should be the sector condition that must be satisfied by N(.). Therefore, the proposed constrained dynamic backpropagation algorithm is derived from the dynamic backpropagation algorithm by replacing the sum of squares error objective function of the dynamic backpropagation algorithm by the interior penalty function discussed above.

CONCLUDING REMARKS

A method of designing stable controllers for a class of nonlinear plants based on multilayer neural networks was presented. The proposed method is based on a proposed constrained dynamic backpropagation algorithm which was derived from the standard

dynamic backpropagation algorithm. The procedure presented in this paper can be extended to MIMO Wiener and Hammerstein nonlinear. However, this extension was not discussed because of space limitations. A lot of theoretical and practical issues were encountered while developing the proposed procedure. These issues ultimately led to the development of a stable adaptive control scheme for a relatively general class of unknown nonlinear dynamic systems (Adetona, 1997). This alternative scheme is not based on the approach presented in this paper and will be the subject of subsequent papers.

ACKNOWLEDGEMENTS

This research is supported in part by NASA Grant NCC5-228. This support is greatly appreciated.

REFERENCES

1) O. Adetona. Stable adaptive control of unknown nonlinear dynamic systems using neural networks. Ph. D. Dissertation at Vanderbilt University, December 1997.
2) Hassan K. Khalil. Nonlinear systems. 2nd Ed. Prentice Hall, 1996.
3) K.S. Narendra. Neural networks for intelligent control. American control conference, Albuquerque, New Mexico. Workshop Number 4, June 1997.
4) A. Netushill, Theory of Automatic Control, Mir Publishers, 1978.
5) K. Parthasarathy. Identification and control using neural networks. Ph.D Dissertation at Yale University, December 1991.
6) V.M. Popov, Hyperstability of Control Systems, Springer-Verlag, 1973.
7) M. Vidyasagar, Nonlinear Systems Analysis, Prentice-Hall, 1993.

Ultrasonic Velocity Variations with Soil Composition for Moisture Measurement

R. Metzl, J. Choi, M. D. Aggarwal
Department of Natural and Physical Sciences
A. Manu, Department of Plant and Soil Science
Alabama A&M University
Normal, AL 35762

Soil moisture content may be measured by many methods, but the presently available techniques all have drawbacks when used in ground truth measurements for remote sensing. Ultrasonic velocity varies with soil moisture content, and may be used as the basis of a new measurement technique. In order to characterize a sensor capable of field use, soil particle size distribution data are compared to ultrasonic velocity in a variety of soils over a wide moisture range..

INTRODUCTION

Soil moisture content may be measured by many methods, from the standard gravimetric technique to neutron flux counters, but the presently available techniques all have drawbacks. When used in ground truth measurements for remote sensing, accuracy, ease of use, reliability and cost are relevant factors.

While it is tempting to think of soil as a collection of individual particles with spaces between them, which can be filled with air or water, the actual case is more complicated. Not all the solid particles are the same size or shape, and not all of them are totally insoluble in water. As water is added to a soil, the particles can clump together to form a structure with bulk mechanical properties different from a discrete collection of particles. Individual particles exhibit only compressive strength, but a soil may also have tensile strength, as indicated by a tensile modulus, and this is dependent on the water content and particle size distribution as well as the composition of the individual particles.

Ultrasonic velocity varies with soil moisture content, since the velocity of sound is different in solid soil particles and in the air and water between the particles. While air has a sound velocity of about 315 m/s (depending on temperature and pressure), water supports a velocity of over 1550 m/s(l). Thus, velocity measurement may be used as the basis of a measurement technique. In order to characterize a sensor capable of field use, soil particle size distribution data are compared to ultrasonic velocity in a variety of soils over a wide moisture range.

DISCUSSION OF ALGORITHMS

The velocity of sound in soil, C_{TOTAL}, is defined as[1]

$$C_{TOTAL} = \sqrt{\frac{E\theta_m V_{TOTAL}}{M_w}\frac{1-\sigma}{(1+\sigma)(1-2\sigma)}} + \sqrt{\frac{G\theta_m V_{TOTAL}}{M_w}} \qquad (1)$$

where velocity is determined by the density, p, the modulus of elasticity, E, Poisson's ratio, σ, and the modulus of rigidity, G (also called the shear modulus or torsional modulus). The density of a volume of soil is described similarly to the density of a volume of solid material. The density of soil, ρ_B, is equal to the mass of the dry soil, M_S, divided by the volume occupied by the soil.

c_{TOTAL} can be further defined as $$C_{TOTAL} = \sqrt{\frac{E\theta_m}{M_w}}\alpha + \sqrt{\theta_m \beta} \qquad (1a)$$

where $$\alpha = V_{TOTAL}\frac{1-\sigma}{(1-0)(1-20)} \quad \text{(lb)}$$

and $$\beta = \frac{GV_{TOTAL}}{M_w} \quad \text{(1c)}$$

Since $$\theta v = \rho_B \theta m \quad (2)$$

Equation (2) can be substituted into Equation (1) to give

$$C_{TOTAL} = \sqrt{\frac{E\frac{\theta_v}{\rho_B}V_{TOTAL}}{M_w}\frac{1-\sigma}{(1+\sigma)(1-2\sigma)}} + \sqrt{\frac{G\frac{\theta_v}{\rho_B}V_{TOTAL}}{M_w}} \quad (3)$$

Equations (1) and (3) are the elementary algorithms (EI and EII) used for estimating mass and volumetric soil moisture content, respectively, from sound wave velocity through soils. The effects of porosity on C_{TOTAL} can be described using a modification of either of these algorithms as shown in Equations (4) and (5).

f JT [illegible] (4)

$$C_{TOTAL} = \sqrt{\frac{E\frac{\theta_v}{\rho_B}\frac{V_{SOLID}}{1-\phi}}{M_w}\frac{1-\sigma}{(1+\sigma)(1-2\sigma)}} + \sqrt{\frac{G\frac{\theta_v}{\rho_B}\frac{V_{SOLID}}{1-\phi}}{M_w}} \quad (5)$$

where V_{SOLID} is the volume of solid portion of the soil system. Organic matter and texture effects can also be evaluated through the measurement and evaluation of the volume of soil organic and mineral fractions within V_{SOLID}.

EXPERIMENTS

Soil samples were provided by the Plant and Soil Science Department of Alabama A&M University. Ultrasonic measurements were taken with a Hewlett Packard Digital Oscilloscope monitoring a Panametrics Pulser/Receiver unit consisting of an impulse type pulse generator and a receiving system. The pulser, receiver and sample holder are similar to equipment previously described by Choi, et. al[2]. Each sample was weighed and measured first with a low moisture content (<lo/o) and then water was added in increments of about 20% of saturation. After each addition, the sample was mixed to distribute the water evenly. Sample thickness and delay times were then measured. Velocity was calculated by subtracting the delay time of the setup without sample (O spacing) from the delay measured by the oscilloscope and dividing the sample thickness by this time.

Soil texture was measured using the standard hydrometer technique of Bouyoucos ‘3).

RESULTS

Figure 1. Velocity of ultrasound vs water content of soils.

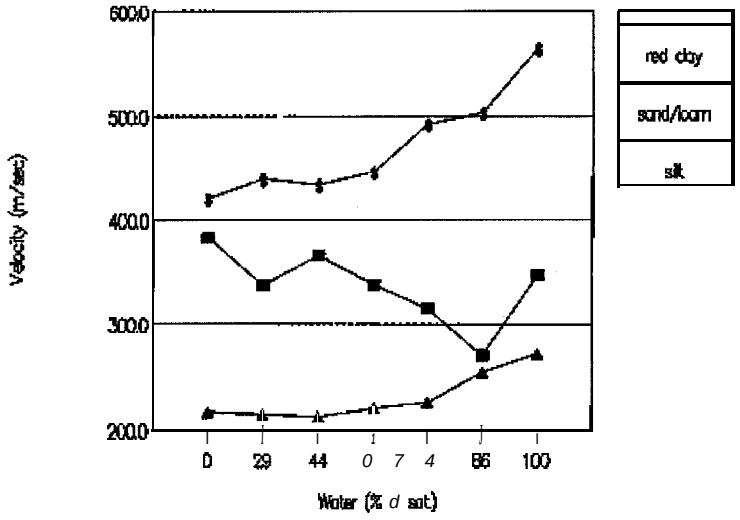


Table 1. Particle size distribution for soils.

1. Quincy Moisture content	0.79% clay	11% silt	12%	sand 77°/0
2. Bama	0.61	11	17	72
3. Decatur	1.69	33	54	13
4. Red Clay	?	37.6	0	62.4
5. sand/Loam	?	0	0	100
6. Silt	?	22.6	60	17.4

The velocity is seen to increase with moisture content in low clay content soils, but to decrease in high clay content soils except at high moisture levels.

CONCLUSIONS

While low clay content soils show a monotonic increase in ultrasonic velocity with soil moisture content, higher clay content produces a decrease in velocity due to a decrease in stiffness of the soil with increasing moisture content. At high water levels, the soil particles are separated by water, and the velocity then increases once more.

Ultrasonic sensors, like other soil moisture instruments, must be calibrated for each soil type. The use of the algorithms allows prediction of velocity based on soil type and moisture content, but a reference method (such as gravimetric moisture determination) must be used to assure correct calibration of the instrument for accurate results.

ACKNOWLEDGMENTS

Acknowledgment is extended to the Center for Hydrology, Soil Climatology and Remote Sensing (HSCaRS) support staff. Contributions from HSCaRS, the Department of Soil, Plant and Animal Science, and the Department of Natural and Physical Sciences, Alabama A&M University, Normal, AL 35762 are also acknowledged. Particle size distribution information was provided by W. Belisle and F. Archer. This work was supported by NASA URC grant 10-82104 (HSCaRS).

REFERENCES

(1) Belisle, W. R., R. Metzl, J. Choi, M. D. Aggarwal, and T. Coleman 1996 Ultrasound algorithm derivation for soil moisture content estimation, NASA University Research Centers Technical Conference on Education, Aeronautics, Space, Autonomy, Earth, and Environment Proceedings, Albuquerque, New Mexico, February 16-19, 1997.

(2) Choi, J., R. Metzl, M. D. Aggarwal, W. Belisle, and T. Coleman. 1996. Use of ultrasonic technology for soil moisture measurement, NASA University Research Centers Technical Conference on Education, Aeronautics, Space, Autonomy, Earth, and Environment Proceedings, Albuquerque, New Mexico, February 16-19, 1997.

(3) G. J. Bouyoucos, Science, 64362 (1926)

BIBLIOGRAPHY

T.E. Matikas, R.L, Crane, Mat. Res. Sot. Bull. 21, 18 (1996)
S.I. Rokhlin and T.E. Matikas, Mat. Res. Bull. 21,22 (1996)
E.P. Papadakis, in Physical Acoustics, Principles and Methods edited by W.P. Mason and R.N. Thurston Vol XII Academic Press, New York (1976)p.277.
M.C. Bhardwaj, Advanced Ceramic Materials 1,311 (1986)
N. C. Brady, 1990. The Nature and Properties of Soils, 10th Edition, Macmillan Publishing Company, New York.
G. J. Curtis, 1982. Ultrasonic Testing, Non-destructive testing of adhesively bonded structures with acoustic methods, John Wiley & Sons, New York, Chapter 13.
C. E. Dull,, H. C. Metcalfe, and J. E. Williams. 1964. Modern Physics, Holt, Rinehart, and Winston, Inc., New York, p. 298-313.
L. E. Kinser, A. R. Frey, A. B. Coppens, and J. V. Sanders. 1982. Fundamentals of Acoustics, 3rd Edition, John Wiley & Sons, New York, Appendix 10.
J. Szilard, 1982. Ultrasonic Testing, Non-conventional testing techniques, John Wiley& Sons, Chichester, p. 1-6.

98 URC038

Optical sensing of Soil moisture with optical fibers using infrared reflectance and Raman Scattering techniques

A. Jackson, B. R. Reddy*, T. L. Coleman and W. Belisle****
Alabama A&M University, Center for Hydrology Soil Climatology and Remote Sensing
*Department of Physics, **Department of Plant and Soil Science
Normal, AL 35762

Abstract

Raman and infrared reflectance techniques are used to measure soil moisture. A bifurcated fiber was used for launching light to the soil surface and collect reflected light. Raman scattering measurement revealed a systematic variation of the signal in a depth profile measurement of peat. Infrared reflectance at 1.4 and 1.9 μm was found to be suitable to detect soil moisture content. A low power Argon ion laser beam was used in the Raman scattering experiments. A tungsten-halogen lamp was used as light source in the infrared reflectance measurements.

1. Introduction

Soil moisture is an important parameter in the fields of agriculture, meteorology and civil engineering. There are several different techniques available to measure soil moisture[1]. The most common technique is the gravimetric analysis. In this technique, the sample is weighed in the laboratory. Later on it is dried in the oven over longer period of time and weighed again. The difference in weight is accounted for the moisture. In meteorology a knowledge of the depth profile is important to predict weather patterns. If one were to use gravimetric analysis then one had to dig out soil samples from different depths in order to estimate moisture content at different depths. This will damage the soil being investigated. This procedure is going to be tedious and time consuming. Hydrogen atoms are more effective in slowing down the neutrons. Hence, energetic neutrons slowed down by the soil moisture, is measured in nuclear techniques. Such thermal neutron density is proportional to the moisture content. This technique involves handling hazardous radioactive sources. Another technique that received wide attention is the microwave remote sensing. It is based on the fact that the dielectric constant of a wet soil is higher than that of a dry soil. Reflected or emitted microwave energy from a soil surface is measured in this technique. This technique measures moisture content to a depth of one tenth of the wavelength used. In time domain reflectometry (TDR), the transit time of an electric pulse is measured from which the dielectric constant and hence the water content can be estimated. This is based on the fact, that the speed of an electromagnetic wave in soil depends on the dielectric constant, which depends on the water content. The dielectric constant of water is 81 and that of dry soil is less than 5. The TDR probe is inserted into the soil to measure the moisture content. If moisture depth profile information is needed then one has to use several probes, which escalates the cost of the technique. TDR probe measures only moisture and does not give information about the presence of other chemicals. There are several other hydrometric and tensiometric techniques to measure soil moisture. In order

to estimate the moisture content one has to develop calibration standards. In some applications, knowledge of the moisture content to a depth of one meter is required. Existing techniques are not suitable for this purpose.

2. **Conceptual and theoretical background of the optical technique**

We are trying to develop a technique based on the optical methods. Optical techniques are superior to other methods because they are nondestructive techniques, do not require any sample preparation, facilitate in-situ measurements, fast in response time, and a simultaneous observation of depth profile is possible. The same probe detects soil moisture as well as any other chemicals or toxic elements present in the soil. Optical techniques are based on the spectral characteristics of a water molecule. Water molecule has an oxygen atom bonded to two hydrogen atoms. The oxygen and hydrogen atoms are connected by covalent bonds, and there are two such bonds in a water molecule. The atoms in a molecule may also rotate about their center of mass. Such rotations occur in the microwave region of the electromagnetic spectrum. In addition, the distance between the atoms can change due to the vibration of the molecule and such vibrations occur in the infrared region of the spectrum[2]. The prominent vibrational frequencies of the water molecule occur at 1,640 (6.1), 3,390 (2.95), 5,150 (1.94), 6,900 (1.45), 8,475 (1. 18), 10,310 (0.97), 13,160 cm^{-1} (0.76 pm), where the numbers within the parentheses are the corresponding wavelengths. Some of the vibrations are infrared active, and hence the water molecules absorb infrared radiation of frequency that matches with the vibrational frequency. A broad band source is used to expose the soil and the reflected light is detected. At the water absorption frequencies the reflectivity will be low[3]. As a result infrared reflectance measurements reveal the presence of water content in the soil. Some vibrations are not infrared active, which implies that infrared absorption is not efficient at those vibrational frequencies. Such transitions normally occur in the Raman spectrum. Raman scattering experiments can be performed using a visible laser. In this experiment a monochromatic laser irradiates the sample under investigation. An intense scattering occurs at the laser frequency, here after called Raleigh scattering. In addition to the Raleigh scattering, additional peaks occur on either side of the strong peak. These peaks are called Raman lines. The Stokes peaks occur on the low energy side whenever the incident beam transfers part of its energy to the molecule and the anti-Stokes peaks occur whenever the excited molecule transfers energy to the incident beam. However, the probability of the latter is very low. As a result the Stokes peaks are stronger than the anti-Stokes peaks. Energy shift of the Raman lines will remain the same even if the laser frequency is changed. Raman peak area is directly proportional to the moisture content.

3. **Experimental configuration**

The experiment includes a light source, a bifurcated optical fiber[4] and a spectrometer (Fig. 1). A tungsten halogen lamp or a globar is used as a light source in infrared reflectance measurements. Argon ion laser is used as a light source in Raman spectral measurements. The bifurcated fiber has three arms. Light is coupled into one arm (single fiber of core size 400 μm) using a lens. On the target side it has seven fibers.

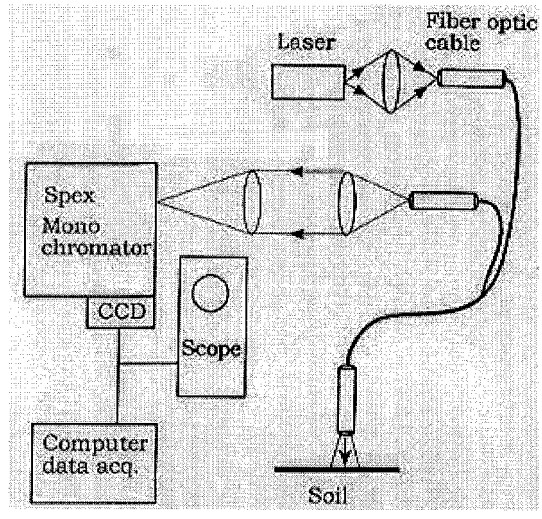


Figure 1. Experimental set-up

The central lighting fiber (400 μm) is surrounded by six collection fibers each of core size 200 pm. The fiber end is within a millimeter to the soil surface. The spot size is about a millimeter on the soil surface. The central fiber transmits light to the soil surface and the surrounding six fibers collect reflected light from the soil surface and transmit the signal to the spectrometer slit. On the spectrometer side the six fibers are arranged in a vertical line and installed close to the slit with an adapter. The main component of the spectrometer is a grating, which disperses the light collected by the fibers, into its individual colors. The spectrum is recorded by scanning the grating and is detected by a lead sulfide detector in case of infrared reflectance measurements and a photomultiplier or CCD camera in case of Raman spectral recordings. Infrared detectors respond to pulsed light input, only. Hence, the detector output is given to a lock-in amplifier for phase lock detection. The detector output is stored in a computer or plotted on a recorder depending on the experimental configuration used. The intensity of the signal collected by the fiber depends on the (1) fiber core size, (2) solid angle subtended at the fiber (3) radiance of the sample (4) power of the laser beam (or light intensity) and (5) transition probability for the molecular transition.

4. **Infrared reflectance measurements**

As mentioned earlier the water molecule shows absorption in the infrared region of the electromagnetic spectrum. The molecular vibrational frequencies at 5,150 (1.94) - 13,160 cm^{-1} (0.76 pm) are covered by the tungsten halogen lamp. Hence, in these experiments, a tungsten halogen lamp is used as a light source to irradiate the soil samples. A compact lamp source was coupled into one arm of the bifurcated fiber with a lens. We have recorded the reflectance spectrum of a wet soil (that was in a cup) which shows decreases in reflectivity at 1.4 and 1.9 μm (Fig. 2) as expected, whose strength increased on the addition of more water to the soil. To check the credibility of the

technique we recorded the reflectance of an aluminum block in the same wavelength region. Its spectral features are different from that of soil, as expected. This clearly

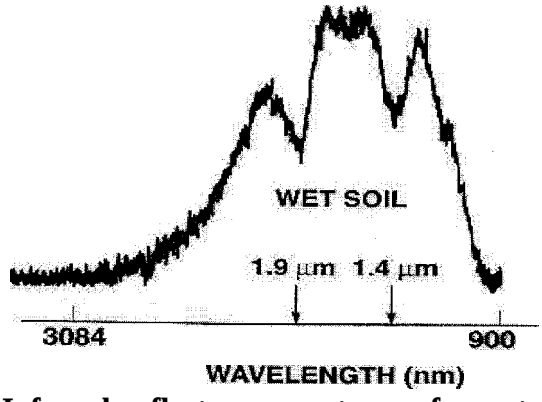


Figure 2. Infrared reflectance spectrum of a wet soil sample

shows that infrared reflectance is a viable technique for measuring soil moisture. Similar spectra were obtained even when we used two different fibers one for launching light and the other to collect reflected light at some angle.

5. **Raman Spectrum**

The vibrational frequency at 3,390 cm^{-1} is Raman active and hence appears in the Raman spectrum. In this technique 514.5 nm (19, 436 cm^{-1}) Ar^{+} laser was used to irradiate the soil and the scattered light was measured with a PMT/CCD. A large peak occured at the laser frequency, and additionally, weak scattering occured at a frequency that was shifted from the laser frequency by 3,390 cm-l. We used argon ion laser for these measurements. We recorded the Raman spectrum of pure water. As expected we have seen scattered peak at 624 nm (16, 025 cm^{-1}). Then we recorded the Raman spectrum of sand, dry peat, dry topsoil, dry and wet soil (obtained from Decatur, AL). Sand spectrum did not reveal any peak whereas peat revealed a Raman peak at 614 nm (Fig. 3). At the same wavelength we detected a peak in the Raman spectrum of topsoil. We were not surprised because the topsoil also has peat in it. Then we recorded the Raman spectrum of Decatur soil, which also revealed the Raman peak at the same wavelength. There is nothing unusual in it because, in general, soil collected from a field is bound to have some peat in it. Then we saturated the soil with water and recorded the spectrum again. The Raman signal appeared at 624 nm (shifted from the Raman peak wavelength obtained for a dry soil). This clearly identifies that the Raman spectroscopy is a powerful tool to identify the types of soils. The dip at 674 nm is an instrumental defect and is not because of the sample.

Another interest in the field of agriculture is the recording of depth profile. We collected soil samples from the Alabama A & M University agricultural research station from three different plots named cotton field, grass field, and alphalpha. At each site samples were collected as a function of depth at 1, 3, 5, 10 and 25 cm from the top

surface. These samples were dried and ground uniformly. A thick aluminum block was used as a sample holder. Five holes were made in that block each of half inch in diameter and one inch in depth. These holes were filled with soil samples to the top of

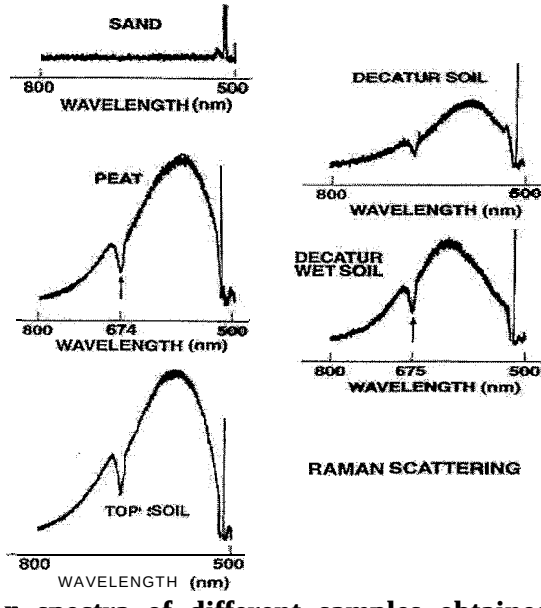


Figure 3. Raman spectra of different samples obtained under 514.5 nm laser excitation

the block, collected from different depths, at each site. The target end of the fiber probe was mounted on top of the block at 1 mm above the surface. The fiber end was fixed in position and the block was moved horizontally to expose different samples to the laser beam in a sequence, to facilitate spectral recordings. This also ensured equal distance for all the samples from the illuminating fiber. Raman spectra were recorded for all these samples as a function of depth. A set of spectra recorded for the cotton field is shown in Fig. 4. The peak area is directly proportional to the abundance of the peat. The peak area decreased as the sample depth was increased from 1 cm to 25 cm. This indicates that the peat content decreased as the depth increased, which was expected. Similar pattern was also observed for samples obtained from grass field.

To see the effect of water on the Raman spectrum we measured the Raman spectrum of a soil sample obtained from the cotton field at a depth of 25 cm. We recorded the Raman spectrum of the dry sample (Fig. 5) which peaked at 614 nm. Without disturbing any thing, we added 1 mL of water to the same powdered soil and the Raman spectrum was recorded again. The overall signal intensity reduced, but the signal amplitude was constant from 600 to 670 nm. There are two reasons for this effect: (1) When water was added to the powdered soil it settled down slightly, as a result the solid angle varied and hence the collected signal was less than that obtained for the dry sample.

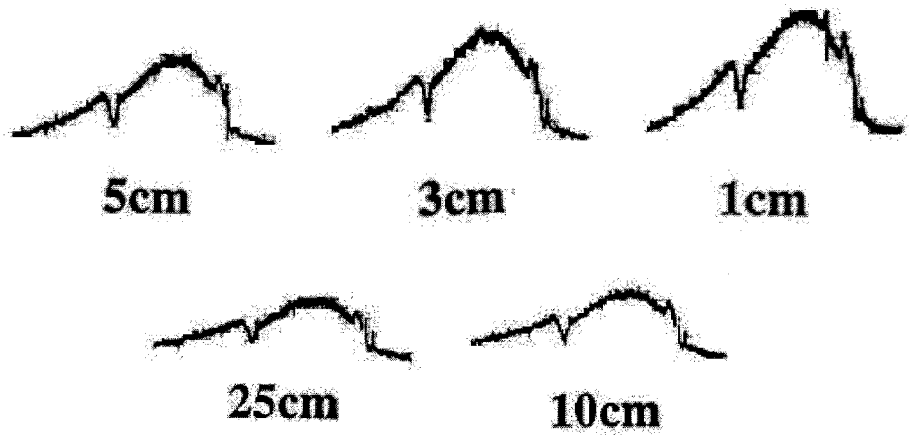


Figure 4. Raman spectra of soil samples obtained under 514.5 nm laser excitation, in the wavelength region 800-450 nm, as a function of depth. The samples were obtained from the cotton field.

Figure 5. Raman spectra of dry and wet soil samples obtained under 514.5 nm laser excitation, in the wavelength region 800-450 nm.

(2) The water signal occurs at 624 nm. As a result the signals due to peat and water superimposed resulting in a broader peak. To improve the signal to noise ratio there are a few approaches (1) modify the probe configuration (2) subtract the peat signal with a computer program (3) increase the incident light power (4) signal averaging and (5) using a high resolution instrument.

6. References

1. T. J. Schmugge, T. J. Jackson, L.M. McKim, Survey of methods for soil moisture determination, Water Resources Research 16,961-979 (1980).
2. G. E. walrafen, Raman Spectral studies of water structure, J. Chem. Phys. 40, 3249-3256 (1964).
3. K. Carl-Brion, Moisture sensors in process control, Elsevier (1986).
4. B. R. Reddy, T. L. Coleman, W. Belisle and M. Thomas, Proceedings of 13th Conference on Hydrology, 345-348 (1997).

98 URC039

X-RAY DIFFRACTION ANALYSIS OF NLO CRYSTALS: TRADITIONAL APPLICATIONS AND MORE NEW OPPORTUNITIES

Mikhail Yu. Antipin[1,2], Ronald D. Clark[1] and Vladimir N. Nesterov[2]

1 Department of Physical Sciences, New Mexico Highlands University, Las Vegas, NM, 87701, USA
[2] Institute of Organoelement Compounds, 28 Vavilov St., B-334, Moscow, 117813, RUSSIA

Introduction

Single crystal X-ray diffraction analysis is one of the more important methods for the molecular and crystal structure determination of matter and therefore it has a great importance in material science including design and engineering of different compounds with non-linear optical (NLO) properties. It was shown in our previous publications [1-3] that this method provides unique information about molecular structure of NLO compounds, their crystal symmetry and crystal packing arrays, molecular conformation and geometries and many other structural and electronic characteristics that are important for understanding the nature of NLO properties of solids.

There are a few important reasons for an application of the X-ray diffraction method for the studies of NLO materials. It is known in particular, that only acentric crystals are capable of second-harmonic generation (SHG) in the solid state under laser irradiation, and the X-ray diffraction technique is the fastest and easiest way to determine the crystal space group. In addition, X-ray data provide valuable information about "optimal" crystal structures of NLO compounds. It was demonstrated [4] that orientation of the dipole moment or charge transfer direction in a molecule with respect to the polar crystal axes is a crucial factor influencing the efficiency of the nonlinear molecular responses. For a given crystal symmetry there are some "optimal" orientation angles [5], and the closer the molecular dipole moment is to this orientation, the better NLO characteristics of the crystal might be. Therefore the presence of a very high molecular nonlinear response (say, calculated or measured in solution) cannot guarantee that even in an acentric crystal the molecule will have an optimal orientation. One such example is the crystal structure of *p*-nitro-dimethylaniline [6]. Therefore molecular modeling and analysis of factors influencing a given crystal structure is an important part in NLO materials design. This modeling includes as a rule energetic calculations of the molecular crystal packing arrays.

Finally, a very new application of the X-ray diffraction method is related to analysis of the electron density distribution p(r) in crystals and some of its characteristics (atomic and group charges, dipole and higher multipole moments, etc.), that may be obtained directly from the diffraction measurements [7]. Some of these characteristics are responsible for NLO properties of materials. It was shown for instance [8,9], that molecular quadruple and octopole moments are related to the components of the first- and second-order nonlinear polarizability tensors, that makes this approach very attractive for a prior estimation of the latter quantities.

All these applications of the modern X-ray diffraction technique together with molecular mechanics calculations of the optimal molecular associates in crystals, quantum chemical

calculations of the static nonlinear polarizabilities (β), and EFISH measurements of the β values in solutions were used recently in our detailed study of the structural and NLO properties in the series of substituted dicyanovinylaromatics [1 -3]. More than 10 compounds of this series were studied including dicyanovinylbenzene itself, its different mono-, di- and tri-substituted MeO-derivatives, *p*-dimethylamino-dicyanovinylbensene and some others. It was shown that there is a good correlation between the predicted static and experimental molecular values of β. X-Ray data (crystal packing analysis) and optimal molecular associates energetic calculations revealed the factors responsible for formation of centric/acentric crystal structures. Calculations of the possible molecular dimers of the compounds studied showed that stacking interactions in centrosymmetric dimers as well as in-plane electrostatic interactions like weak C-H.. .N hydrogen bonds (C-H bond belongs to the dicyanovinyl group) probably prevent crystal acentricity. Introduction of an *o*-methoxy substituent into the molecule does not increase the magnitude of β (in accord with calculations and experimental data) but inhibits the formation of the H-bonded dimers, giving the known acentric crystal structure of DIVA (*o*-methoxy-dicyanoninylbenzene). On the contrary, di- and trimethoxy-substituted derivatives in the series considered prefer the formation of centrosymmetric antiparallel staking-type dimers and therefore these compounds form centrosymmetric crystals.

These results for the given series of compounds have demonstrated clearly that the described approach for the molecular design of new NLO materials (that includes theoretical estimation of the molecular nonlinearities, molecular and crystal structure modeling, chemical synthesis, X-ray diffraction analysis, nonlinear optical measurements in solutions and solid state and "optimization" of the molecular and crystal structure, see [4,10]) is workable and therefore in the present paper we have used it for a new study of some potentially important NLO materials.

The list of the compounds studied includes new dicyanovinyl derivatives, namely *p*-dimethylamino-dicyanovinylbenzene 4-Me_2N-C_6H_4-CH=$C(CN)_2$ (1), *p*-diethylamino-dicyanovinylbenzene 4-Et_2N-C_6H_4-CH=$C(CN)_2$ (II) and 4-(4-methoxyphenyl)- 1,1 -dicyano-1,3-butadiene 4-MeO-C_6H_4-CH=CH-CH=$C(CN)_2$ (III), as well as two derivatives of 5-nitropyridine with bulky 2-cyclooctylamino (IV) and 2-adamantylamino (V) substituents. Compound IV is the well studied NLO material known also as COANP [11], and corresponding multitemperature diffraction data for IV were obtained for comparison with those for V and the study of the ring dynamic in the crystal. We should note that very short communications about the NLO properties in the solid state and X-ray structures for compounds I and V were reported earlier [12,13], but no important structural details including atomic coordinates were published in these papers. Finally, in the present work we will discuss our preliminary low temperature high-resolution X-ray data for the *m*-nitroaniline (mNA) single crystal (VI). This is one of the "classical" organic NLO materials and electron density distribution analysis in this simple compound has a great scientific interest.

Experimental part and details of calculations

All X-ray diffraction data were obtained using suitable single crystal samples obtained by slow crystallization from solution. The quality of the single crystals for compounds I and III was relatively low due to systematic twinning, therefore corresponding R-values (Table) are rather

Table. Structure Determination Summary for Compounds I-VI
(for compound IV the data are given for two temperatures 297 and 167K)

	I	II	III	IV	v	VI
Formula	$C_{12}H_{11}N_3$	$C_{14}H_{15}N_3$	$C_{12}H_{10}N_2O$	$C_{14}H_{19}N_3O_2$	$C_{15}H_{19}N_3O_2$	$C_6H_6N_2O_2$
Crystal system	monocl.	monocl.	monocl.	orthorh.	orthorh.	orthorh.
Space group	$P2_1$	Pc	$P2_1/n$	$Pca2_1$	$Pca2_1$	$Pca2_1$
z	2	2	4	4	4	4
Unit cell dimensions:						
a, Å	3.951(4)	9.378(5)	3.957(2)	26.285(5) 26.099(5)	28.158(6)	19.037(3)
b, Å	14.078(11)	9.564(5)	13.832(4)	6.647(1) 6.637(2)	6.584(2)	6.507(1)
c, Å	9.499(9)	14.664(8)	10.392(5)	7.631(2) 7.555(2)	7.395(1)	5.028(1)
β, deg.	99.88(8)	98.84(4)	94.11(3)	90.0 I 90.0	90.0	90.0
v, $Å^3$	520.5(8)	1299.5(12)	567.4(4)	1333.3(8) 1308.7(8)	1371.0(5)	622.9(5)
Temp. K	193	293	297	297 I 167	163	160
2θ range	2-44	4-58	2-44	2-60 2-60	3-80	2-80
Number of reflections and papameters:						
collected	810	3007	1176	1834 / 1798	4242	2154
indep. ohs.	656	2774	689	7751822	4186	1314
no. param.	136	214	143	163	257	115
Refinement method in least-squares (on F or F^2):	F	F^2	F	F / F	F^2	F
Final R indices and GOF:						
R	0.098	0.056	0.129	0.068 / 0.059	0.051	0.033
Rw	0.144	0.107	0.141	0.099 / 0.057	0.118	0.039
GOF	2.54	0.998	1.30	1.91 / 1.84	1.022	2.17

high. Diffraction data for all compounds were obtained using 4-circle automatic diffractometers "Siemens P3/PC" and "Syntex $P2_1$" at room **(II, III, IV)** and low temperatures **(I, IV, V, VI) with** MoKα radiation (λ=0.7 107 Å), graphite monochromator and θ/2θ-scan technique. Important crystal data, experimental details and results of refinements are summarized in the Table. Details of the energetic calculations of the optimal molecular dimers and quantum-chemical calculations of the first-order hyperpolarizabilities β are described in our previous papers [2,3,14].

Results and Discussion

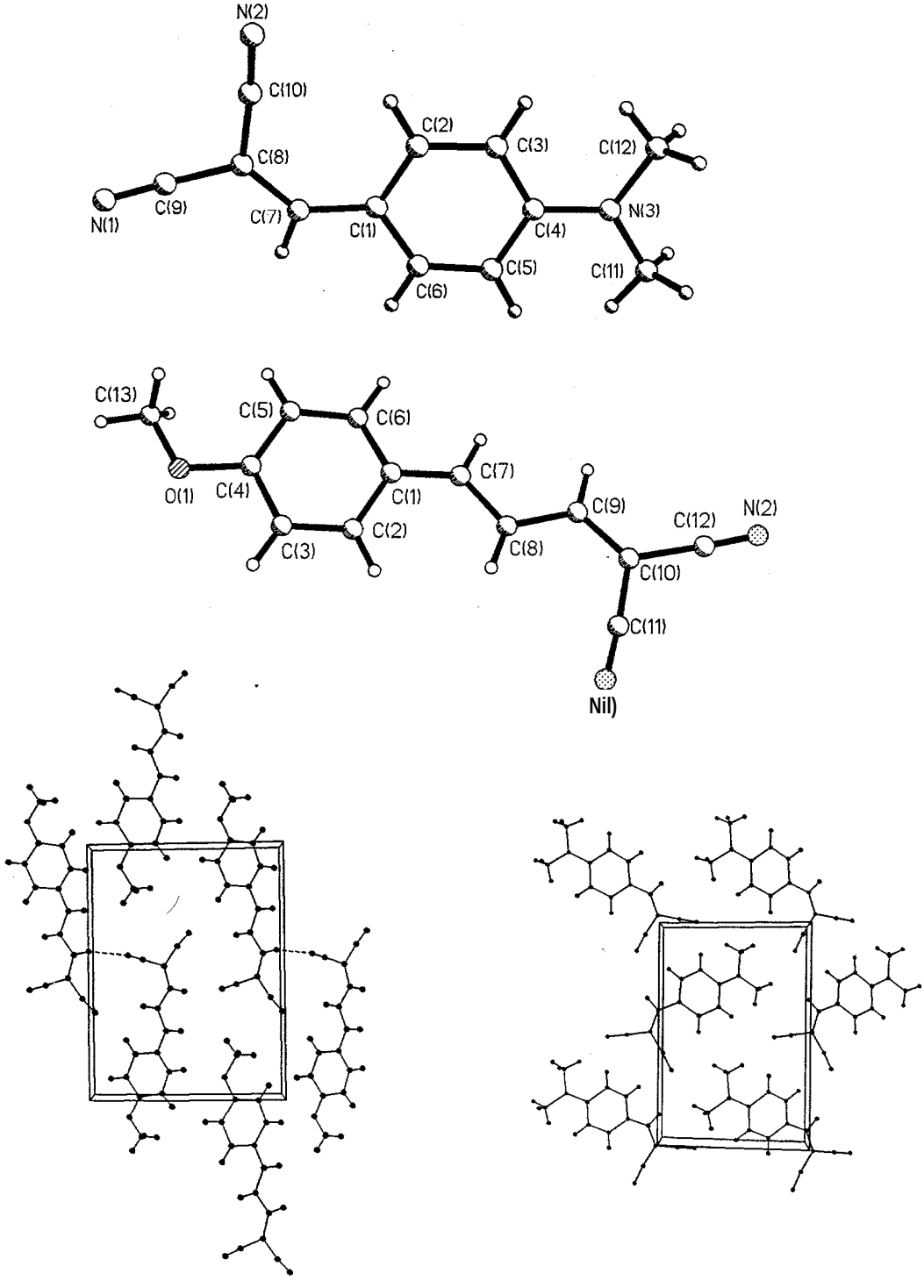
Let's consider first compounds I-III containing the dicyanovinyl fragment $-CH=C(CN)_2$ and p-substituted aromatic rings. Calculated and measured in solution using EFISH technique values of β for these compounds are relatively high and therefore they might be considered as prospective NLO materials. For example, measured β values for I and III are both equal to 32.10-30 esu in accord with [15] ($CHCl_3$ solution), and 39.2.10-30 (I) and 24.5.10-30 (III) esu in accord with our data [14] (in 1,4-dioxane, λ=1 064 nm, after correction for absorption and solute dilution). Two of these compounds, namely I and III, form acentric crystals and therefore we may expect a SHG signal for them in the solid state as well. Molecular structures of I and III and crystal packing diagrams are presented in Figure 1. The molecules are almost planar and their molecular geometry parameters have the expected values.

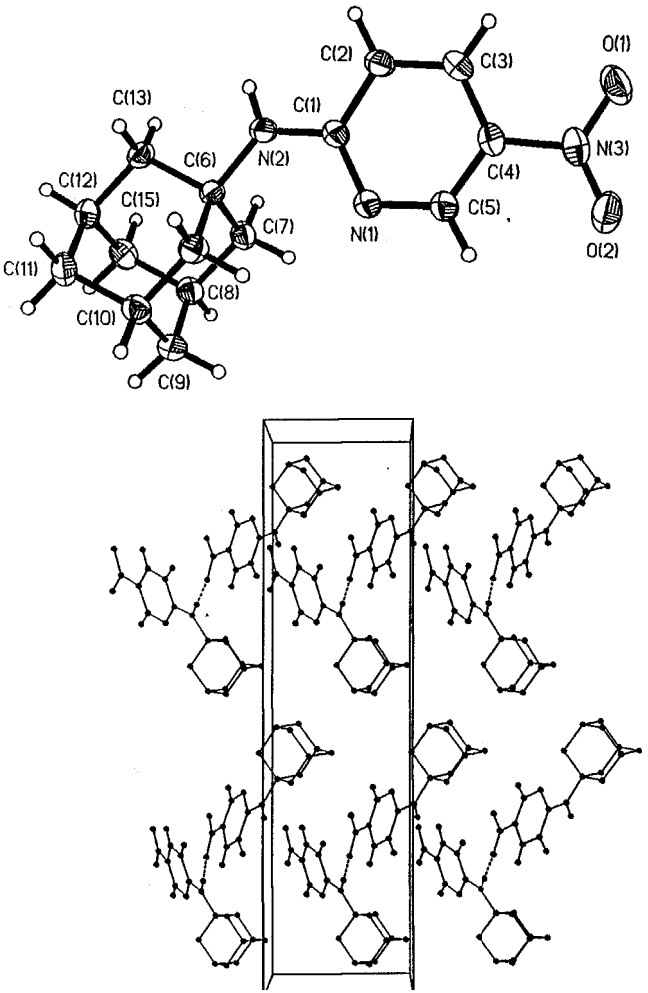
Crystal packing analysis shows that in the structure of I, orientation of the long molecular axis with respect to the polar crystal axis (Y, in the space group $P2_1$) is rather close to the optimal value (54.7°) for this space group: the angle between the crystal Y-axis and N(3)-C(8) line that might be a possible charge transfer direction in the molecule is equal to 59.3°. So, we could expect that this compound should be a rather effective SHG material in the solid state. However, the SHG signal by the powder sample of I using a Nd:YAG laser was so weak that it was barely visible at high power. Probably, we may relate this fact with the strong absorption in I near the second harmonic spectral line (λ_{max}=431nm, see [14]).

For compound III the orientation of the molecular charge transfer axis in the crystal is also close to the optimal one in the space group Pc, and its powder sample gives a strong SHG signal at the wavelength 532 nm that is detectable easily with naked eye. The λ_{max} band in III has a hypsochromic shift to the value of 392 nm [15], therefore this compound may be an interesting material for further NLO studies in the solid state.

The series of parent compounds I-III gives an attractive opportunity to check the workability of the simple theoretical approach for a feasible prediction of the centric/acentric crystal structure for a given compound based only on the calculation of the optimal molecular dimer configurations using molecular mechanics approximation. Details of this approach are described in [2,3]. It was found for molecules I and III that the energy difference between C_i and C_2 dimers is almost negligible [14], therefore one might not exclude for these compounds a formation either acentric or centric crystal structures. Present X-ray data showed that these molecules form acentric crystals that was not excluded in the calculations.

Figure 1. Molecular structures of the compounds I and III (top) and crystal packing diagrams down the crystals a-axes (bottom)





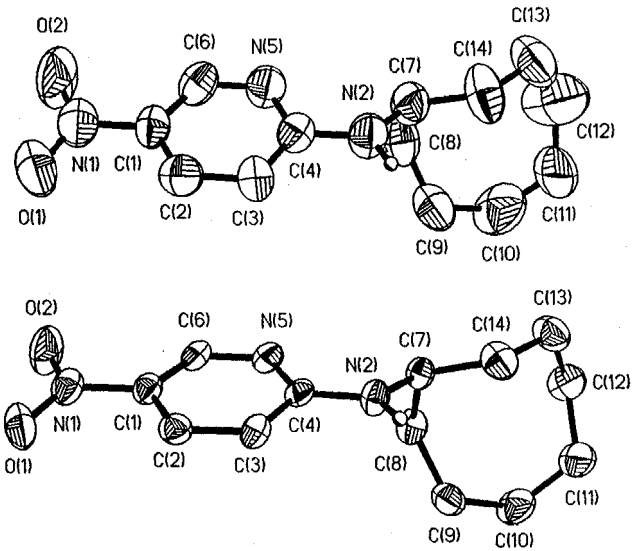


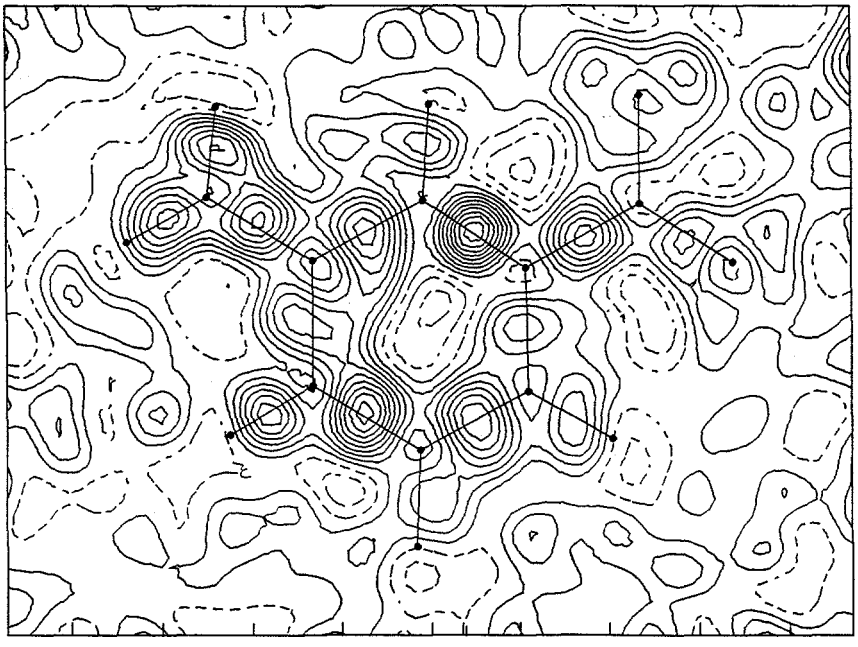
Figure 2. **Molecular structure of CONP (IV) in crystal** (top, right) **at 297 and 167 K. Smaller thermal ellipsoids at 167K indicate dynamic ring disorder.**

Figure 3. **Molecular structure of V in crystal** (top, left) **and crystal packing diagram** (bottom, left) **down the *b*-crystal axis.**

Similar optimal dimer calculations for compounds IV (COANP) and V (adamantyl analog of COANP) revealed that C_2 dimers are more stable than centrosymmetric ones (the difference is about 0.5 kcal/mol) and probably therefore both of them form acentric crystals. It was shown in [11,13] that these crystal have very high and useful NLO characteristics, and our rather accurate low-temperature X-ray diffraction data allow one to compare some structural features of these two compounds. Crystal data for IV and V demonstrate a close similarity of their crystal structures. On the other hand, it is known [16] that COANP forms a glassy state on cooling the melt that might be related with some kind of disorder in the flexible 8-member ring. Our diffraction data for IV at 297 and 167K (Figure 2) demonstrate that this disorder probably has a dynamic nature. As for IV, orientation of the charge transfer axis in V with respect to the polar crystal axis is close to the optimal for the space group $Pca2_1$. The molecular structure of V and crystal packing diagram down the b-axis are shown in Figure 3.

In the last part of this paper we will consider shortly some preliminary results of the charge density analysis in the crystal of *m*-nitroaniline (mNA, VI). The main goal of this work is to calculate important electronic characteristics of this compound (charges, dipole and higher multipole moments) responsible for its non-linear optical characteristics. Similar calculations have been done earlier for some organic NLO crystals, including DIVA [1]. A deformation electron density map in the molecular plane of **VI is** presented in Figure 4. This map contains definite maxima at the covalent chemical bonds and at the oxygen atom lone-pairs. Further analysis of the diffraction data for VI including multipole refinement and calculation of the first-order hyperpolarizabilities directly from the diffraction data (see [1 ,9]) is in progress.

Figure 4. Deformation electron density map in the molecular plane of mNA



Acknowledgments

This work was supported by NASA-IRA grant NAG5-5121. The authors acknowledge Dr. Robert L.S. Amai and Leonard Romero for synthesis of the compounds studied.

References

1. M. Yu. Antipin, R. D. Clark. *Technical Reports, Alliance for Nonlinear Optics:* Huntsville, Alabama, 1996; University of Puerto Rico, 1997; Alabama A&M University, 1997.
2. M. Yu. Antipin, T. A. Barr, B. H. Cardelino, R. D. Clark, C. E. Moore, T. Myers, B. Penn, M. Romero, M. Sanghadasa, T. V. Timofeeva. J. *Phys. Chem. B, 1997,101,2770-2781.*
3. M. Yu. Antipin, R. D. Clark, V.N. Nesterov, M. Sanghadasa, T. V. Timofeeva, K. A. Lyssenko. Mol. *Cryst. Liq. Cryst. 1998,* in press.
4. *Nonlinear Optical Properties of Organic Molecules and Crystals.* Ed. by D.S. Shelma and J. Zyss. Academic Press, Orlando, FL, 1987, Vol. 1.
5. J. Zyss, J. L. Oudar, *Phys. Rev. A, 1982,26,2028.*
6. T.C.W. Mak, J. Trotter. *Acts Crystallogr. 1965,18,68.*
7. *The Application of Charge Density Research to Chemistry and Drug Design.* Ed. by J. A. Jeffrey and J.F. Piniela. NATO ASI Series, Vol. 250, Plenum, N.Y. 1990
8. F.N.H. Robinson. *Bell System Technical Journal, 1967,913.*
9. A. Fkyerat, A. Guelzim, F. Baert, J. Zyss. *Phys. Rev. B.* 1996,53, 1636.
10. *Materials for Nonlinear Optics: Chemical Perspectives.* Eds. by S.R. Marder, J.E. Sohn, G.D. Stucky. ACS Symposium, v. 455. ACS: Washington, DC, 1991.
11. Ch. Bosshard, K. Sutter, P. Gunter, G. Chapuis. *J. Opt. Sot.* Am. 1989,6,721-725.
12. M. Matsuoka, M. Takao, T. Kitao, T. Fujiwara, K. Nakatsu. *Mol. Cryst. Liq. Cryst. 1990, 182A, 71-79.*
13. S. Tomaru, S. Matsumoto, T. Kurihara, H. Suzuki, N. Ooba, T. Kaino. *Appl. Phys. Lett. 1991,58,2583-2585.*
14. M. Yu. Antipin, T.V. Timofeeva, R.D. Clark, V.N. Nesterov, M. Sahghadasa. *J. Phys. Chem. 1997,* submitted for publication.
15. L.-T. Cheng, W. Tarn, S.R. Marder, A.E. Stiegman, G. Rikken, C. W. Spangler. *J. Phys. Chem.* 1991,95, 10643-10652.
16. G. Lahajnar, I. Zupancic, R. Blinc, A. Zidansek, R. Kind, M. Ehrensperger. *Z. Phys. B, 1994, B95, 243-247.*

98 URC040

Constitutive Modeling of the Mechanical Properties of Optical Fibers

L. Moeti[1], S. Moghazy[1],D. Veazie[1], E. Cuddihy[2]

'Clark Atlanta University, Atlanta, GA 30314
'Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109

ABSTRACT

Micromechanical modeling of the composite mechanical properties of optical fibers was conducted. Good agreement was obtained between the values of Young's modulus obtained by micromechanics modeling and those determined experimentally for a single mode optical fiber where the wave guide and the jacket are physically coupled. The modeling was also attempted on a polarization-maintaining optical fiber (PANDA) where the wave guide and the jacket are physically decoupled, and found not to applicable since the modeling required perfect bonding at the interface. The modeling utilized constituent physical properties such as the Young's modulus, Poisson's ratio, and shear modulus to establish bounds on the macroscopic behavior of the fiber.

INTRODUCTION

Polarization-maintaining (PM) optical fibers are increasingly being considered for new applications in, for example, advanced space craft inertial reference units as fiberoptic rotation sensors. Applications where volume is a consideration have used PM fibers wound onto small spools'. These new PM fibers, often referred to as PANDA fibers, are significantly different from non-PM optical fibers in two major ways. First, typical non-PM optical fibers generally consist of a wave guide surrounded by an acrylic jacket. In the PANDA fiber, there exists a silicone layer between the wave guide and the jacket. Secondly, the PANDA fiber contains two beryllium oxide stress rods around the wave guide core that produces a stressed wave guide core specific for polarized light.

Technical literature exists on the physical and mechanical properties of non-PM optical fibers which can be used for engineering design. Development of theories to predict fiber lifetime for non-PM optical fibers using fiber parameters have been established'. Literature reports on the temperature-dependent behavior of highly birefringent optical fibers are available"'. For the case of PANDA fibers the temperature dependence of fibers wound onto different diameter cylinders has been measured'. The modeling and prediction of mechanical properties of optical fibers based on material properties is not available in the open literature. The ability to predict mechanical properties of optical fibers for design applications where testing may not be practical would be of great use.

APPROACH

In this paper, the modeling of the mechanical properties of optical fibers based on micromechanics is presented. The properties predicted by micromechanical modeling are compared to experimental values determined earlier[8].

Micromechanical modeling is often used to quantitatively characterize the effects of constituent properties on the macroscopic behavior of composite materials. The effects of the arrangement of the constituents, as well as the distribution of stresses and deformations among the constituents, can be studied using micromechanical modeling'. In the micromechanical self-consistent field methods, approximations of phase geometries are made and a simple representation of the response field is obtained. The phase geometry is represented by one single fiber embedded in a matrix cylinder. This outer cylinder is embedded in an unbounded homogeneous material whose properties are taken to be equivalent to those of the average composite properties. The matrix under a uniform load at infinity introduces a uniform strain field in the fiber. Elastic constants are obtained from this strain field. The results obtained are independent of fiber arrangements in the matrix and are reliable at low fiber volume fractions V_f, reasonable at intermediate V_f, and unreliable at high V_f[10].

RESULTS AND DISCUSSION

In this research, micromechanical analyses was used to estimate the composite properties of a single mode optical fiber. As in the context of continuum mechanics, it is assumed that the interface between the glass fiber (wave guide) and the polymeric matrix (jacket) is perfect. Therefore throughout the entire loading application, the interface is a mathematical surface across which material properties change discontinuously, while the interracial traction and displacements are continuous across the interface. When perfect bonding is assumed, the macromechanical properties and strength of the composite (wave guide plus jacket) are determined solely by the properties of the constituent materials. For the control fiber as described earlier, the interface between the wave guide and the jacket can be regarded as being perfect, i.e., physically coupled. In the case of the PANDA fiber, the wave guide and the jacket are physically decoupled.

A simple model for estimating the modulus of a composite structure of two different materials assumes that in uniaxial load each discrete material is independent, equally strained, and therefore contributes to the total load measured in proportion to only their individual modulus and cross sectional area. The simple equation is

$$MC = (A_g M_g)/A_T + (A_p M_p)/A_T$$

where MC is the composite modulus based on total load divided by the total cross-sectional area A_T

A_g is the cross-sectional area of the glass wave guide

A_p is the cross-sectional area of the plastic wave guide

A_T is the total cross-sectional area equal to $A_g + A_p$

M_g is the modulus of the glass wave guide

and M_p is the modulus of the plastic jacket

As reported in experimental work done earlier* for the control fiber, the modulus M_g of the glass wave guide is 7.10 x 10' Psi, and that of the whole fiber (the composite

material) is M_c = 2.10 x 10^6 Psi. The ratio A_g/A_T is 1/3, and the ratio A_p/A_T is 3/4. With this information, the modulus M_p of the plastic jacket can be estimated as 430,000 Psi, which is in close agreement with published values for polymethylmethacrylate (450,000 Psi). For the micromechanical modeling presented herein the modulus of the jacket was assumed to be that of polymethylmethacrylate.

The micromechanical method used here focuses on the upper and lower bounds on the elastic constants. This method does not predict the properties directly, however, if the upper and lower bounds coincide, then the property is determined exactly. Frequently, the upper and lower bounds are well separated. When these bounds are close enough they can be used as indicators of the material behavior. This is the case for the longitudinal properties of a unidirectional system as considered here. Hill[11] derived rigorous bounds on the longitudinal Young's modulus, E_{11}, in terms of the bulk modulus in plane strain (k_p), Poisson's ratio (ν), and the shear modulus (G) of the two phases. For the glass wave guide of the optical fiber system, E_f, G_f and ν_f were 1.04 x 10^7 Psi, 6.07 x 10^6, and 0.16, respectively which are known published values for glass. The polymer jacket used on the single mode optical fiber is an acrylic polymer, believed to be similar to polymethyl methacrylate (PMMA). The published value for the modulus E_m of PMMA is given as 4.79 x 10^5 Psi, which was used for micromechanical modeling. The compressibility of PMMA, ξ_m = 0.0355 Psi, was used to derive the shear modulus for the jacket, G_m, as

$$G_m = \frac{3(1-2\nu_m)}{2\xi_m(1+\nu_m)}. \qquad (1)$$

The volume fractions for the wave guide and the jacket, V_f and V_m, were 0.25 and 0.75, respectively. Calculations to determine E_{11} of the optical fiber system were made by varying the Poisson's ratio (ν_m) and Young's modulus (E_m) of the jacket, separately. No restrictions were made on the core form or packing geometry. The term k_p is the modulus for lateral dilatation with zero longitudinal strain and. is given by

$$k, = \frac{E}{2(1-2\nu)(1+\nu)} \qquad (2)$$

The bounds on the longitudinal modulus, E_{11}, are

$$\frac{4V_f V_m(\nu_f-\nu_m)^2}{(V_f/k_{pm})+(V_m/k_{pf})+1/G_m} \leq \text{“1 ‘f ‘f} - E_m V_m \leq \frac{4V_f V_m(\nu_f - \text{v.})'}{(V_f/k_{pm})+(V_m/k_{pf})+ \text{I/G,}} \quad (3)$$

Figure 1 shows the bounds for the longitudinal modulus of the composite optical fiber, E_{11}, for a Young's modulus (E_m) of the jacket equal to 4.79 x 10' Psi, and a range of values of the Poisson's ratio (ν_m) of the jacket. A composite modulus, E_{11}, in the order of 2.96 x 10^6 Psi is estimated from Figure 1, for the single mode optical fiber used

in this study. In Figure 2, the bounds for El_1 *is* given for Poisson's ratio v_m = 0.37 and a range of values for E_m. These results show good agreement with the measured values.

A value of 2.87 x 10' Psi for the modulus was obtained from static loading' which is consistent with the value predicted from micromechanical modeling (2.96 x 10' Psi). Good agreement is observed in the values of modulus from micromechanical modeling and static loading, for the single mode optical fiber where the glass wave guide and the polymeric jacket are physically coupled. For the case of the PANDA fiber where the glass wave guide and the polymeric jacket are decoupled, the micromechanical modeling was not applicable.

It should be noted that this type of micromechanical analysis is limited by the assumption of a perfect interface between the wave guide and the jacket. However, the assumption of perfect bonding is sometimes inadequate, as in the case of the PANDA fiber where the wave guide and the jacket are physically decoupled thus providing a distinct interracial zone (interphase) between the core and the jacket. This weak, compliant interphase has lower strength and stiffness than those of the wave guide and the jacket. Recently, Veazie and Qu[9] obtained micromechanical solutions for the nonlinear inelastic behavior of composites with weak, compliant interphases using the Mori-Tanaka method as a building block. New micromechanical analyses such as these may find utility in predicting the macromechanical properties of advanced optical fiber systems such as PANDA fibers.

CONCLUSIONS

Micromechanical modeling was used to predict the composite modulus of a single mode optical fiber. The micromechanical analysis was limited to the case of a perfect interface between the glass wave guide and the polymeric jacket of the optical fibers. The modeling worked well in the case of the single mode optical fiber where the glass wave guide and the polymeric jacket are physically coupled. The micromechanical modeling did not work for the polarization-maintaining PANDA fiber where the glass wave guide and the polymeric jacket are physically decoupled. Good agreement for the single mode optical fiber was obtained for the modulus predicted by micromechanical modeling and values measured experimentally.

REFERENCES

1. M.S. Perlmutter, "*A tactical fiber optic gyro with all-digital signal processing,*" Proc. 16th Biennial Guidance Test Symposium, AFDTC-TR-93-06, pp. 337-350, 46th Guidance Test Squadron/CC (AFMC), Holloman AFB, NM (1993).

2. F.P. Kapron and H.H. Yuce, "*Theory and measurement for predicting stressed fiber lifetime*," Optical Engineering 30(6), pp. 700-708 (1991).

3. V. Ramaswamy, R.H. Stolen, M.D. Divine, and W. Pleibel, "*Birefringence in elliptically clad borosilicate single-mode fibers,*" Appl. Opt., vol. 18, pp. 4080-4084 (1979).

4. A. Ourmazd, M.P. Varnham, R.D. Birch, and D.N. Payne, *"Thermal properties of highly birefringent opticsl fibers and preforms,"* Appl. Opt., vol. 22, pp. 2374-2379 (1983).

5. S.C. Rashleigh and M.J. Marrone, *"Temperature dependence of stress birefrigence in an elliptically clad fiber,"* Opt. Lett., vol 8, pp. 127-129 (1983).

6. A.J. Barlow, "Optical *fiber birefrigence measurement using a photoelastic modulator,"* J. Lightwave Technol., vol LT-3, pp1 35-145 (1985).

7. P.B. Ruffin, J. Sawyer, and C.C. Sung, "*Effect of spool diameter on the temperature-dependent extinction* ratios of *polarization-maintaining fibers*," Optical Engineering 34 (l), 56-57 (1995).

8. L. Moeti, S. Moghazy, A. Ally, S. Barnes, A. Abhiraman, P. Desai, and E. Cuddihy, "Mechanical and Physical Testing of a Polarization-Maintaining Optical Fiber," J. of Lightwave Technology, Paper submitted.

9. D.R. Veazie, and Qu, J., Effects of Interphases on the Transverse Stress-Strain Behavior in Unidirectional Fiber Reinforced Metal Matrix Composites, *Composites Engineering, 5, No. 6,597-610, (1995).*

10. K.K. Chawla, *Composite Materials,* Springer-Verlag, New York, (1987).

11. R.A. Hill, A Self Consistent Mechanics of Composite Materials, *Journal of Mechanics and Physics of Solids, 13, 189, (1965).*

ACKNOWLEDGEMENTS:

Support provided by NASA-JPL's Minority Science and Engineering Initiatives Office (MSEIO) and the Clark Atlanta University NASA supported High Performance Polymers and Ceramics (HiPPAC) Center is gratefully acknowledged.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

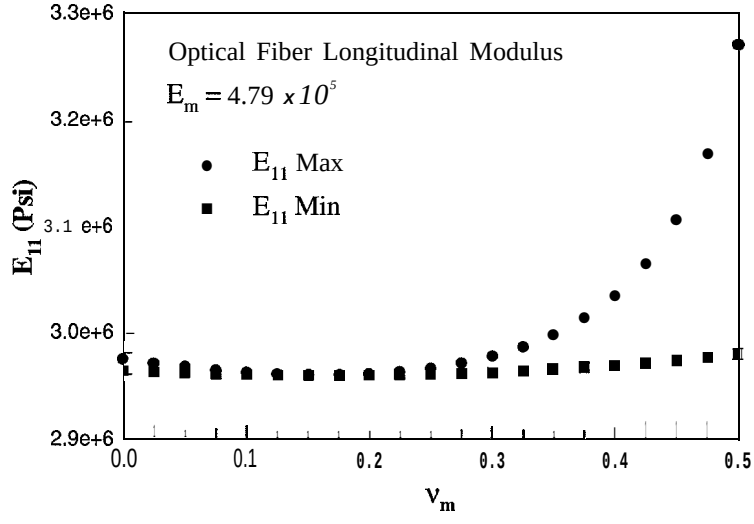


Figure 1. Bounds for the longitudinal modulus, E_{II}, for a young's modulus (E_m) of the jacket equal to **4.79** $\times 10^5$ psi, and a range of values of the Poisson's ratio (ν_m) of the jacket.

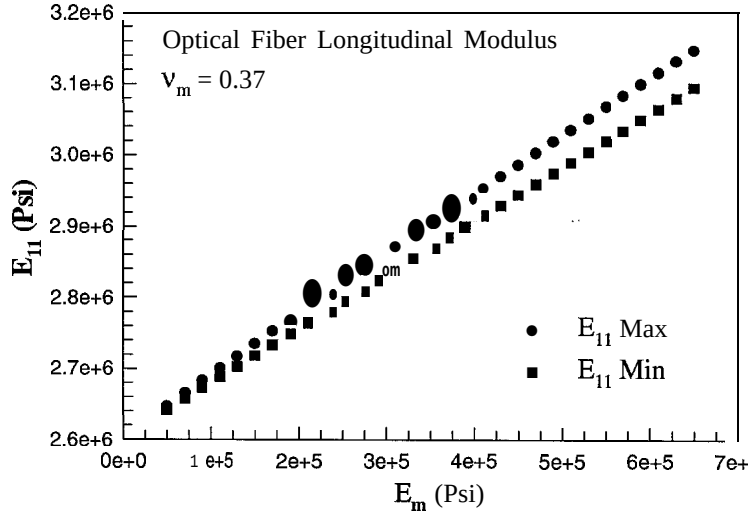


Figure 2. Bounds for the longitudinal modulus, E_{11}, for a Poisson's ratio of the jacket, ν_m = ().37, and a range of values for the Young's modulus of the jacket (E_m).

98 URC041

INTERLAMINAR SHEAR PROPERTIES OF LOW-DENSITY, LOW-CONDUCTIVITY PAN ABLATIVE COMPOSITES

David R. Veazie
Assistant Professor

Ronald O. Grover, Jr.
Research Assistant

High Performance Polymers and Ceramics (HiPPAC) Center
Clark Atlanta University
Atlanta, GA 30314

Abstract

In this study, the mechanical response of six low-density, low conductivity polyacrylonitrile (PAN) based carbon fiber reinforced phenolic resin composites was investigated. Six panels were fabricated at various autoclave cure pressures and ceramic microballoon filler content. Of the six PAN composite panels, two panels each were fabricated with 20°/0, 15°/0 or 10% ceramic microballoon tiller content. Composites with 20% microballoon content were cured at pressures of 1.65 MPa and 0.35 MPa, whereas panels with 15°/0 microballoon content were cured at 1.65 MPa and 0.69 MPa. Composites with 10% microballoon content were cured at 1.65 MPa and 0.35 MPa. The improvement in mechanical properties achieved by the influences of the microballoon content and the cure pressure was assessed by comparing the ultimate interlaminar shear strength of the composite at room temperature. Results included ultimate interlaminar shear strength of the PAN composites, and the description of the unique experimental techniques required by testing of brittle composites.

KEYWORDS: phenolic resin, ablative composite, polyacrylonitrile fiber, Iosipescu method, ply lift, interlaminar shear, ceramic microballoons.

Introduction

Concerns about the stability of the domestic supply of carbonizable aerospace grade rayon for ablative composites have led to interest in an alternate carbon fiber precursor. Polyacrylonitrile (PAN) based carbon fibers have been shown through various studies to be an acceptable alternative. Another approach to developing a replacement for rayon is to reduce the thermal conductivity and density of the composite by the inclusion of filler ceramic microballoons to the PAN prepreg. These microballoons are an evolution from previous rayon low density materials which used carbon microballoons, but provide cost and thermal conductivity advantages.

The ply lift anomaly which occurs in ablative composites causes erratic erosion and spallation, leaving large multiple voids between plies of the material. This can lead to loss of the

thermally protective char cap and result in increased heat-affected depth of the material. The fiber/matrix bond strength, as well as the interply strength, in low-density PAN based carbon fiber reinforced phenolic composites is believed to influence this event. In particular, higher cross-ply tension and interlaminar shear strength properties are necessary to accommodate the high cross-ply thermal expansion (cool-down dilation) experienced in ablative applications.

An experimental investigation was performed to study the mechanical properties thought to initiate ply lifting in PAN ablative composites. Room temperature interlaminar shear tests were conducted on PAN fiber ablative composites panels fabricated at Clark Atlanta University (CAU) and The Thiokol Corporation. Specimens from six PAN fiber ablative composite panels were created at different cure pressures and ceramic microballoon content. Ultimate interlaminar shear strength served as a basis of comparison.

Materials and Processing

Six PAN ablative composite panels were autoclave cured measuring 35.56 cm. long by 44.64 cm. wide by 10.16 cm. thick. Of the six PAN composite panels, two panels each were fabricated with 20%, 15% or 10% ceramic microballoon filler content. Optical microscopy and vibrational spectroscopy were used to study the size, size distribution and integrity of the virgin microballoons prior to layup.

The prepreg broadgoods, obtained from ICI Fiberite and designated MX-134 LDR, included the ceramic microballoon fillers in SC-1008 Phenolic Resin, along with the T300 PAN fibers. Normal certification testing of the prepreg including a minimum of volatile content, resin content, filler content and flow was performed by ICI Fiberite. In addition, flow and Differential Scanning Calorimetry (DSC) testing of the prepreg was performed by Thiokol. The composite panels were fabricated at Thiokol through the Productivity Enhancement Center of the NASA Marshall Space Flight Center (MSFC), and Clark Atlanta University through the High Performance Polymers and Composites (HiPPAC) Center.

From the as-received rolls, 35.56 cm. long by 44.64 cm. wide plies were cut and piled into 8 stacks of 32 plies per stack. Each stack was double sealed in foil bags and placed in cold box storage until layup. After warming to ambient (approximately two hours), these 8 "sub-stacks" were preheated in a convection oven at 60°C for 30 min. To simulate tape wrap debulk, the stacks were then arranged in a hydraulic press with load cell and induction heated platens to form a 256 ply composite. Teflon tape and high temperature, non-perforated release film was used to hold the assembly and to prevent the laminate from sticking to the tool surfaces, respectively. Thermocouples attached to the center of each platen provided feedback to the platen temperature controller. A maximum pressure of 2.76 MPa was applied to the stacks at 60°C for 30 min.

Following the press, the panels were bagged and placed under a minimum of 24,0 in. Hg vacuum for 1 hour. Each panel was monitored to assure that the vacuum drop did not exceed 1.0 in. Hg within 15 min. and held to a minimum of 21.0 in. Hg for a period of 15 min. Each bagged panel was then placed in a programmable temperature and pressure cycle autoclave. Autoclave pressure was held constant throughout the cure cycle. Thermocouples placed directly on top and under the center of the first debulk stack was used to monitor the composite temperature.

Six PAN ablative composite panels were fabricated and tested. Table 1 shows the ceramic microballoon content and autoclave cure pressure of each panel. The microballoon filler contents selected were based on the widest range that are of interest, and are predicted to result in

composite bulk densities from 1.1 glee to 1.4 glee. The skewed distribution to the higher process pressures is prompted by the desire to be able to co-process these materials with standard density materials to further reduce cost.

Table 1. Composite Panels

Panel (#)	Microballoon Content (%)	Autoclave Cure Pressure (MPa)
1	20	1.65
2	20	0.35
3	15	1.65
4	15	0.69
5	10	1.65
6	10	0.35

Specimens were machined according to ASTM Specification D-5379 (V-Notched Beam Method). Figure 1 shows the specimen configuration and dimension, along with the typical thickness variation throughout the specimen.

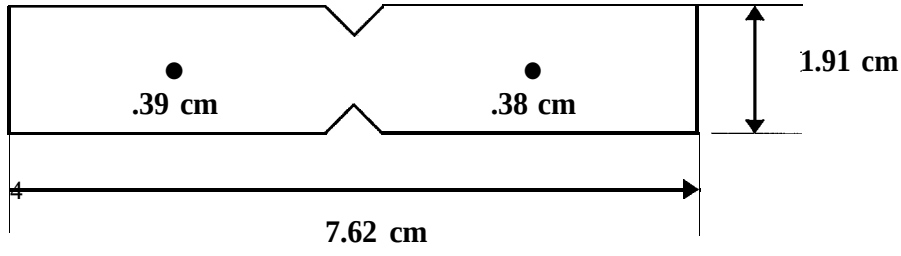


Figure 1. Schematic Diagram of PAN composite specimens. All dimensions in centimeters.

Experimental Procedure

Room temperature tests of PAN ablative composites were performed according to ASTM D-5379 V-Notched Beam Method (also referred to as the Iosipescu Shear Method). Specimens machined from six different panels (described in the above section) were tested. All six panels were fabricated by The Thiokol Corporation. To ensure repeatable processing procedures and mechanical properties, Clark Atlanta University also fabricated Panels 2, 4, and 6. The mechanical properties from these three additional panels were compared to the mechanical properties of the panels fabricated at Thiokol. Each specimen machined from the respective panels was tested to failure to determine the ultimate interlaminar shear strength.

Testing was performed in laboratory air using a 100 kN servo-hydraulic test frame equipped with digital controller and computer data acquisition. Several specimens were tested using the Iosipescu Wyoming test fixture; however, difficulty was encountered due to the

random fracture of the specimen outside of the V-notch gage area. Because of this, the V-Notched Beam specimens were divided into two sections about the gage length, and tested to obtain the ultimate interlaminar shear strength. The "halved" specimens were affixed in the Iosipescu Wyoming test fixture and sheared to failure. Pressure to the test fixture was applied by the hydraulic actuator action of the test frame. Specific procedures and techniques relating to the testing may also be found in Veazie and Vaughn [1].

Results and Discussion

The results of the room temperature interlaminar shear strength tests of the PAN composites are shown in Figures 2 - 4. To reduce experimental errors at least three replicates were tested for each panel. Each figure shows the average maximum shear strength of the corresponding panel. Error bars indicate the deviation of the ultimate strength values of each panel. The ultimate interlaminar shear strength of the Thiokol fabricated panels is represented in Figure 2. The ultimate interlaminar shear strength of the panels fabricated at CAU is shown in Figure 3. For comparison purposes, the Thiokol and CAU Panels 2, 4, and 6 are presented in Figure 4.

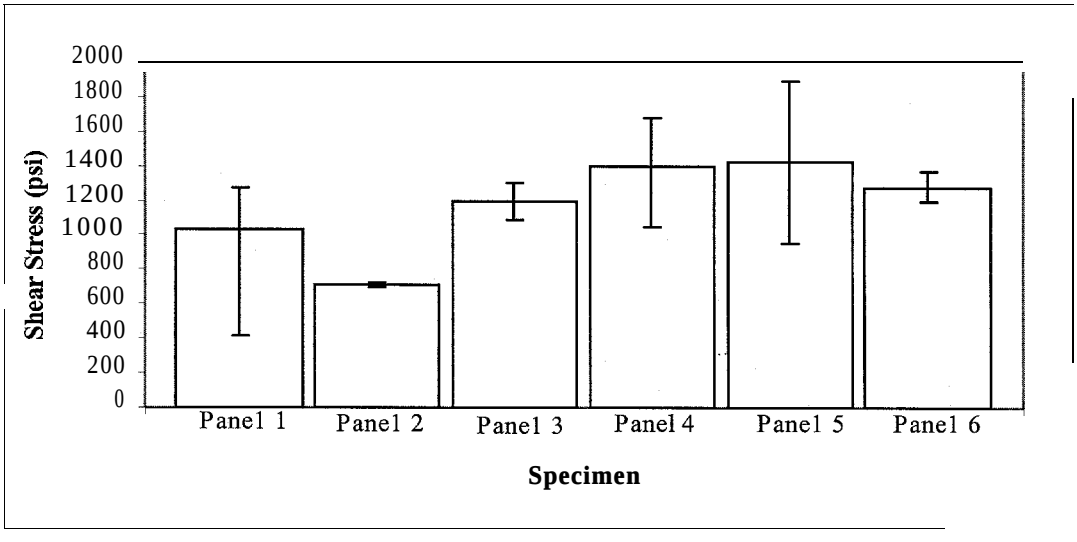


Figure 2. Ultimate interlaminar shear strength of Thiokol Specimens.

Panels 4 and 5 of the Thiokol specimens had almost identical ultimate interlaminar shear strengths and were the highest of the Thiokol panels. Panel 2, however, had the lowest ultimate interlaminar shear strength. Panels made with the 20% microballoon filler content (Panels 1 and 2) had the lowest ultimate interlaminar shear strengths. Panels 1, 3, and 5 were each fabricated at a cure pressure of 1.65 MPa. Of these 1.65 MPa cured panels, the specimens with the lowest microballoon filler content (Panel 5) had the highest ultimate interlaminar shear strength while the specimens with the highest microballoon content (Panel 1) had the lowest ultimate interlaminar shear strength.

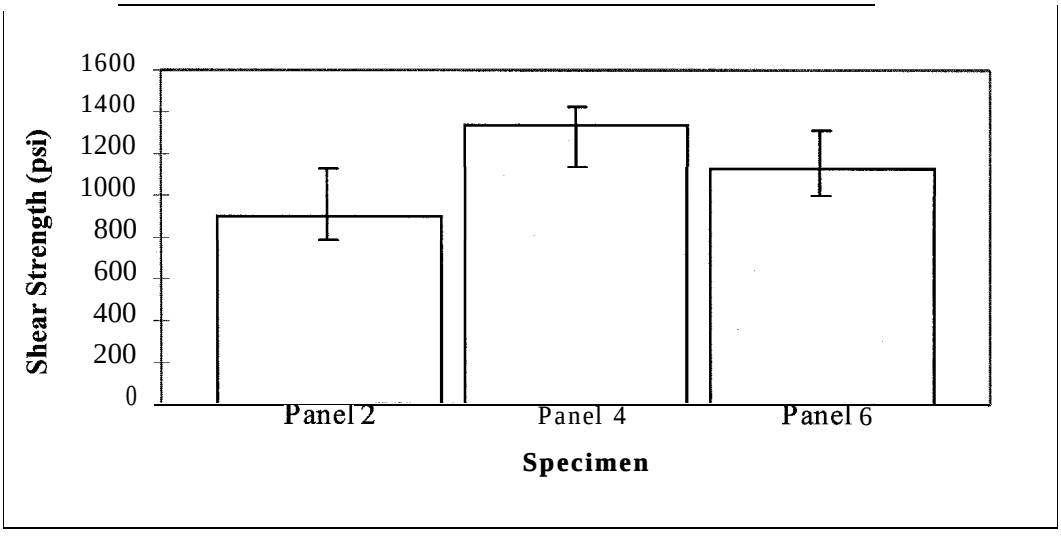


Figure 3. Ultimate interlaminar shear strength of CAU Specimens.

Panel 4 of the CAU specimens had the highest ultimate interlaminar shear strength, while Panel 2 had the lowest (See Figure 3). Panels 2 and 6 were fabricated at a cure pressure of 0.35 MPa. The lower microballoon content of the CAU specimens fabricated at 0.35 MPa had the higher ultimate interlaminar shear strength. The ultimate interlaminar shear strengths of Panel 4 fabricated at both Thiokol and CAU were similar (See Figure 4). In a comparison of the ultimate interlaminar shear strengths from Panels 2 and 6 fabricated at both CAU and Thiokol, differences of approximately 20°/0 and 12°/0 were observed, respectively. On average, Thiokol fabricated specimens experienced higher ultimate interlaminar shear strengths than the CAU specimens, except for those with 20°/0 microballoon content and 0.35 MPa cure pressure.

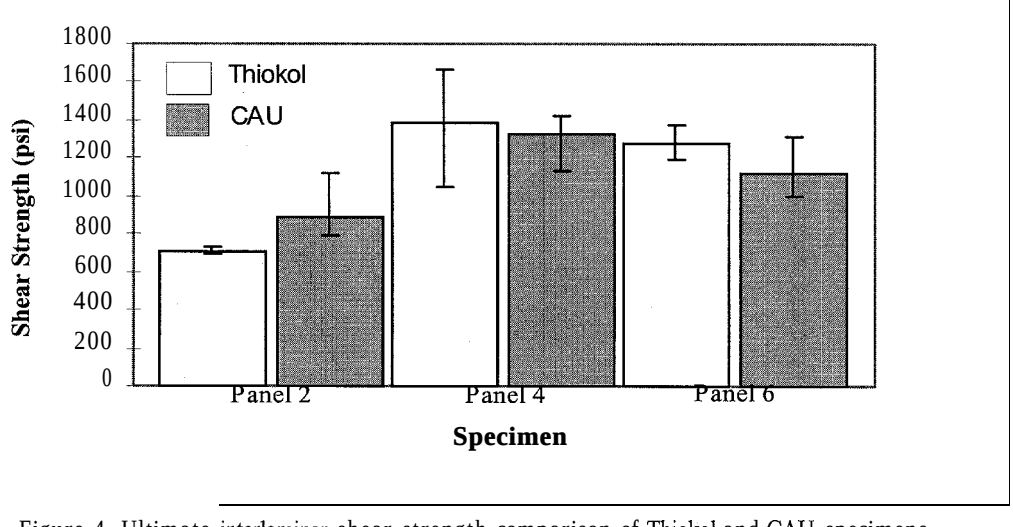


Figure 4. Ultimate interlaminar shear strength comparison of Thiokol and CAU specimens,

Concluding Remarks

An experimental study was undertaken to evaluate the interlaminar shear properties of low density, low conductivity PAN ablative composites. The test results compared the ultimate interlaminar shear strength of PAN ablative composites fabricated by both Thiokol and CAU at different ceramic microballoon tiller contents and cure pressures. Considerable difficulty was encountered due to the random fracture of the specimens outside of the V-notch gage area. The results of this work, albeit not conclusive here, revealed that the V-Notched Beam specimens can be divided into two sections about the gage length, and along with the Iosipescu Method, tested to obtain the ultimate interlaminar shear strength.

Experimental results indicated that at a given cure pressure, the ultimate interlaminar shear strength increased in the specimens as the ceramic microballoon tiller content decreased. There was no consistent relationship between microballoon filler content and the ultimate interlaminar shear strength in specimens cured at the same pressure. The ultimate interlaminar shear strength of the CAU specimens followed the same trends as the Thiokol specimens. It was also observed that the Thiokol and CAU fabricated specimens had comparable maximum shear strengths.

Acknowledgments

The authors gratefully acknowledge the Productivity Enhancement Center of the NASA MSFC, the HiPPAC Center at Clark Atlanta University, and The Thiokol Corporation for providing the materials used in this study.

References

[1] Veazie, D. R., and Vaughn, K., Preliminary Evaluation of Low-Density, Low-Conductivity PAN Ablative Composites, *Proceedings of the First NASA NANURC Conference*, Greensboro, NC, March 31 - April 2, 1996.

"Un-annealed and Annealed Pd Ultra-Thin Film on SiC Characterized by Scanning Probe Microscopy and X-ray Photoelectron Spectroscopy"

W.J. Lu, D.T. Shi, K. Elshot, E. Bryant, K. Lafate, H. Chen, A. Burger, W. E. Collins

Center for Photonic Materials and Devices
Department of Physics
Fisk University,
Nashville, TN 37208

Abstract

Pd/SiC has been used as a hydrogen and a hydrocarbon gas sensor operated at high temperature. UHV-Scanning Tunneling Microscopy (STM), Atomic Force Microscopy (AFM) and X-ray Photoelectron Spectroscopy (XPS) techniques were applied to study the relationship between the morphology and chemical compositions for Pd ultra-thin films on SiC (<30Å) at different annealing temperatures. Pd ultra-thin film on 6H-SiC was prepared by the RF sputtering method. The morphology from UHV-STM and AFM shows that the Pd thin film was well deposited on SiC substrate, and the Pd was partially aggregated to round shaped participates at an annealing temperature of 300°C. At 400°C, the amount of surface participates decreases, and some strap shape participates appear. From XPS, Pd_2Si was formed on the surface after annealing at 300°C, and all Pd reacted with SiC to form Pd_2Si after annealing at 400°C. The intensity of the XPS Pd peak decreases enormously at 400°C. The Pd film diffused into SiC, and the Schottky barrier height has almost no changes. The work shows the Pd sicilides/SiC have the same electronic properties with Pd/SiC, and explains why the Pd/SiC sensor still responds to hydrogen at high operating temperatures.

Key words: SiC, Pd thin film, Scanning Probe Microscopy, X-ray photoelectron spectroscopy.

1: Introduction:

Since commercial SiC wafers became available in 1990, SiC research and device technology have grown rapidly [1]. SiC has well known characteristics such as: wide bandgap, high electrical field breakdown strength, high thermal conductivity, and high saturated electron drift velocity. These characters give SiC devices the capability to operate at high temperature, high radiation field, and high power level.

The SiC device fabrication requires the development of contacts and interconnects which are physically and chemically stable at severe conditions, such as high temperatures. Metal/SiC contact and epitaxial films studies have resulted in commercially available SiC devices including blue light-emitting diodes, high-voltage diode, and hydrogen and hydrocarbon sensors which can be operated at severe environments. The diffusion and reaction between metal thin film and SiC can alter the electrical properties, and even destroy the device.

Most transition metals form good Schottky contacts on SiC. Some metals, such as, Ti, Ni, Al, Ta, and Ni-Al on SiC contact [2] will convert from Schottky contact to Ohmic contact after high

temperature annealing. The high temperatures assure a significant diffusion and reaction with the SiC, and likely create a graded, highly defective interface associated with the ohmic behavior. Therefore, understanding the interracial compositions and electrical properties is very important for both Schottky and ohmic contact. The relationship between the chemical compositions and electrical properties is one of the most critical issues in SiC devices.

Pd/SiC without annealing has been studied [3], and yielded a Schottky barrier height of 1.11 eV for Si-face SiC measured by I-V measurement and XPS. Pai *et al.* **[4]** studied the thermally annealed Pd/SiC by XRD and ion backscattering. SiC starts to react with Pd at 500°C, and form Pd_3Si. When the annealing temperature is above 700°C, Pd_2Si is found on the interface instead of Pd_3Si. No carbon compounds are found. Chen *et al.* [5] reported that the Pd/SiC diode has a broad interface region with Pd_xSi distributed throughout the Pd thin film after annealing at 425°C.

In the literature, the thickness of metal films on SiC used by most researchers is up to 1000 Å, and the interracial compositions are analyzed by sputtering/XPS. This method is unable to determine the stoichiometrical ratio in the interface formed during the annealing because of the effect of the preferential sputtering of light elements. TEM [6] and XRD [4] techniques are unable to determine the amorphous and micro-crystals in the interface region. In this work, we avoid the above problems. The Pd ultra-thin metal film with the 15-30 Å thickness on SiC has been studied by XPS, UHV-STM, and AFM. The advantages for the ultra-thin metal films are:

(1) The effect of the preferential sputtering of light elements is avoided. The Pd-Si compounds can be determined stoichiometrically in the interface directly by XPS. The morphology of these compounds and the thin film also will be directly observed by UHV-STM and AFM.
(2) The role of diffusion for Pd, Si, and C and their relationships with electrical properties will be explored at different annealing temperatures. The diffusion rate can be measured by XPS peak intensities.

2: **Experimental.**

a: The Samples and Surface Cleaning Procedure.

An n-type, Si-face 6H-SiC wafer with 3.5° titled off-axis on Si (0001) substrate was purchased from Cree Research Inc.. The doping concentration was at 2.6 x 10^{18} cm^3.

We have tested the effectiveness of different chemical procedures on Si wafers, and examined them by XPS. The cleaning procedure for SiC is given in the following steps:

(a) $NH_4OH:H_2O_2:H_2O$ (1:1:2) at 85°C for 5 rein;
(b) $HCl:H_2O_2:H_2O$ (1:1:2) at 85°C for 5 rein;
(c) $HF:H_2O$ (20% solution) for 5 rein;
(e) HNO_3 (50%) at 90°C for 5 rein;
(f) $HF:H_2O$ (10% solution) for 2 min;
(g) immerse the sample in CH_3OH before metal thin film deposition.
Between each step the sample was rinsed by de-ionized water.

b: Pd Ultra-thin Film Fabrication.

In this work, RF (radiation frequency) sputtering system (Kurt J. Lesker Company) has been used to fabricate Pd ultra-thin film on SiC. The chemically cleaned SiC sample was put in the sputtering chamber, and heated up to 350°C for 30 minutes (the maximum heating temperature for

this sputtering system). The Pd thin film fabrication conditions on SiC are: the distance between the Pd target and the substrate is 5 inches, the exposure time is 20 seconds, and a 35 W power was applied. The samples were annealed in air at 300°C and 400°C for 30 minutes.

c: Measurements and Characterizations.

(1) X-ray Photoelectron Spectroscopy:

The surface analysis was determined primarily by X-ray photoelectron spectroscopy (XPS). The experiments were performed on a Kratos X-SAM 800 spectrometer with a resolution of 0.1 eV at the pressure of 1 x 108 torr. The Mg Kα ($h\nu$ = 1253.6 eV) radiation was used. The data was processed by the VERSION software associated with the equipment.

(2) UHV-STM (Scanning Tunneling Microscopy):

UHV-STM (Ultra-high Vacuum-Scanning Tunneling Microscopy) was used, with a head from McAllister Inc., and the controlling software from NanoScope II (Digital Instruments, Inc.). The W tip was used, the vacuum pressure was 8 x 10^{-9} torr, the current was 0.5 nA, and the bias voltage is from -1000 to -3000 mV. All measurements are at room temperature.

(3) AFM (Atomic Force Microscopy):

A NanoScope E (Digital Instruments, Inc.) was used in this project.

3: **Results and Discussions.**

a: **Morphology of SiC substrate, and un-annealed Pd/SiC.**

Morphology of the surface state of SiC substrate and the un-annealed Pd/SiC sample was observed by AFM and UHV-STM. Figure 1 shows the morphology of the SiC substrate sample after chemical cleaning procedure. The morphology at the micron scale is shown in Figure 1(a) and (b) by AFM. The surface of the SiC substrate has some straight scratches which were caused by the mechanical polishing process. The surface state of the SiC substrate depends on the commercial process.

UHV-STM has higher resolution than AFM and air-STM. Figure 1 (c), using UHV-STM, shows that there are many nano-scale defects on the cleaned SiC substrate. Some honeycomb-like structures [7] appear on the larger smooth terraces.

Figure 2 shows the morphology of the SiC after Pd thin film deposition by AFM and UHV-STM. There are no significant changes in morphology as shown in Figure 2 (a) and (b) by AFM, and the surface looks smoother than the SiC substrate in micron scale. So the surface of SiC seems well covered by the Pd thin film. Using UHV-STM, Figure 2(c) shows Pd round shaped islands on the SiC surface. The size of the Pd islands on SiC surface ranges from a few nanometers to about 40 nm.

b: Morphology of Pd/SiC samples annealed at 300 “C and 400 “C.

Figure 3 shows the AFM and UHV-STM images for the Pd/SiC sample annealed in air at 300°C for 30 minutes. From the AFM images shown in Figure 3 (a) and (b), the morphological features are that some rounded surface precipitates were formed at 300°C. The size of the rounded surface precipitates is about 30-80 nm. Figures 3(c) shows the morphology at the nano-scale for the sample annealed at 300°C using UHV-STM. As shown, the Pd partially aggregates into the round shaped precipitates on the SiC surface,

Figure 4 shows the AFM and UHV-STM images for the Pd/SiC sample annealed in air at 400°C for 30 minutes. As the AFM images shown in Figure 4 (a) and (b), the number of the surface precipitates decrease after the annealing temperature increase to 400 'C, and some strap shaped precipitates were formed. Using UHV-STM, Figure 4 (c) shows that the surface is covered by the precipitates with unclear edges, it looks as though the Pd precipitates have melted on the surface and formed new compounds on the surface.

c: The relationship between chemical composition and annealing temperatures.

The XPS measurements are very important to understand the interracial electronic properties, such as the Schottky barrier height, the Fermi energy, and the interracial compositions. Figure 5 shows that the Pd ($3d_{5/2}$ and $3d_{3/2}$) X-ray photoelectron spectroscopy for un-annealed Pd/SiC, and Pd/SiC annealed at 300°C and 400°C. The Pd ($3d_{5/2}$) peak shifts from 335.90 eV to 336.59 eV after annealing at 300°C. After annealing at 400°C, this peak is at 336.67 eV and the intensity decreases tremendously. The peak at 336.59 eV is assigned to $Pd3d_{5/2}$ of Pd element [8], and the peak at 336.59 eV is assigned to $Pd3d_{5/2}$ of the silicide Pd_2Si [9]. As shown in Figures 3-5, the intensities of the Pd XPS peaks and the number of the surface precipitates decreases with increasing annealing temperature. This means that the quantity of Pd on the surface decreases with increasing annealing temperature. Obviously, Pd diffuses into the SiC substrate during annealing.

As Chen *et al.* reported [5], the Pd/SiC diode still responds to hydrogen after annealing at 425°C for 140 hours. At this temperature, the interracial compositions have changed, but the hydrogen sensing properties have not been altered. Our work [10] shows that the Schottky barrier height for the Pd/SiC system almost has no changes with annealing temperature, although Pd_2Si has been formed after annealing at 300°C, and most of Pd has reacted with Si to form Pd_2Si after annealing at 400°C. In Ni/SiC, Al/SiC systems [2], the formation of the metal silicides in the interface during annealing decreases the Schottky barrier height, and the metal/SiC contact changes from Schottky contact to ohmic contact. In the Pd/SiC system, the formation of Pd silicides in the interface does not alter the Schottky contact when the annealing temperature is below 400 "C.

Our results show that Pd diffuses into SiC to form Pd_2Si, the Pd_2Si contacts to SiC, and the silicide formation does not alter the Schottky contact properties. This explains why the Pd/SiC diode still responds to hydrogen at high operating temperatures.

Acknowledgment:

This work was supported by NASA under Grant No. NAGW-2925, NAG3-2012, and NAG3-1430.

References.

[1] M.R. Melloch, and J.A. Cooper, Jr., "Fundamentals of SiC-Based Device Processing", *MRS Bulletin,* March, 1997, p42.

[2] L.M. Porter, and R.F. Davis, *Materials Science and Engineering,* **B34**, *83 (1995).*

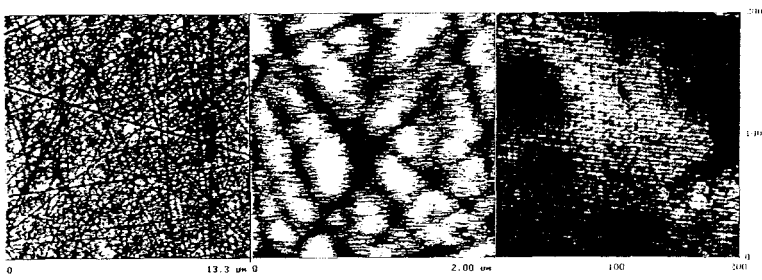
[3] J.R. Waldrop, *J. Appl. Phys., 75(9), 4548 (1994).*

[4] C.S. Pai, C.M. Hanson, and S.S. Lau, *J. Appl. Phys., 57(2), 618 (1985).*

[5] L.-Y. Chen, G.W. Hunter, P.G. Neudeck, G. Bansal, J. B. Petit, and D. Knight, *J. Vat. Sci. Technol.* **A 15(3), 1228 (1997).**

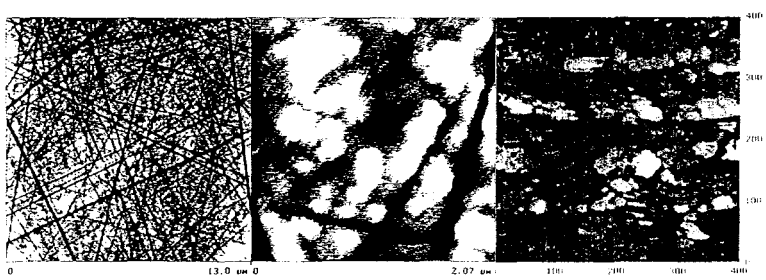
[6] F. Goesmann, and R. Schmid-Fetzer, *Semicond. Sci. Technol.* **10,** 1652(1995).

[7] S. Tanaka, R.S. Kern, R.F. Davis, J.F. Wendelken, and J. Xu, *Surface* Sciences, 350,247 (1996).
[8] "*Handbook of X-Ray Photoelectron Spectroscopy* ", C. D. Wagner *et al.* Perkin-Elmer Co., Eden Prairie, MN. 1979.
[9] B. Schleich, D. Schmeisser, and W. Göpel, *Surface Science,* **191,367 (1987).**
[10] D.T. Shi, W.J. Lu, E. Bryant, K. Elshot, K. Lafate, H. Chen, A. Burger, W. E. Collins, *NASA University Research Center Technical Conference '98,* February 21-26, 1998, Huntsville, AL.



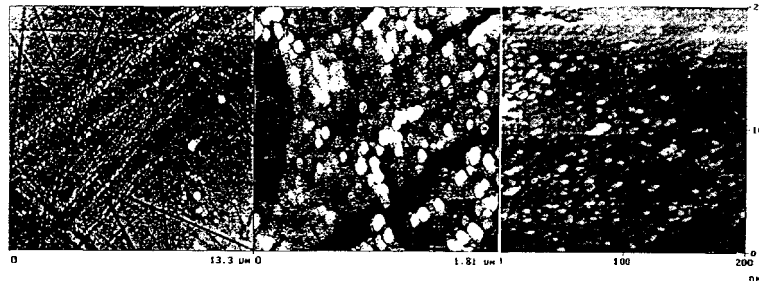
(a) (b) (c)

Figure 1: AFM, UHV-STM Images for SiC substrate.
(a) and (b) AFM images, (c) UHV-STM image



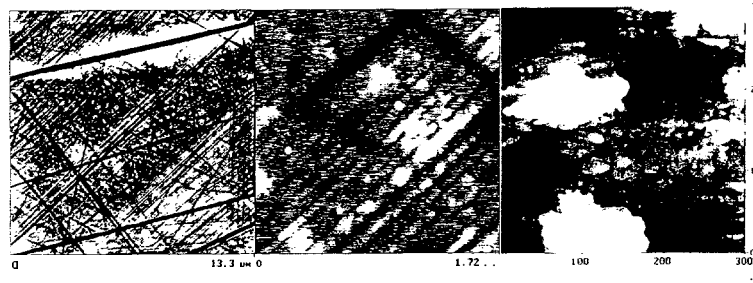
(a) (b) (c)

Figure 2: AFM, UHV-STM Images for un-annealed Pd/SiC sample.
(a) and (b) AFM images, (c) UHV-STM image



(a) (b) (c)

Figure *3:* AFM, UHV-STM Images for Pd/SiC sample annealed at 300 "C. (a) and (b) AFM images, (c) UHV-STM image



(a) (b) (c)

Figure 4: AFM, UHV-STM Images for Pd/SiC sample annealed at 400 "C. (a) and (b) AFM images, (c) UHV-STM image

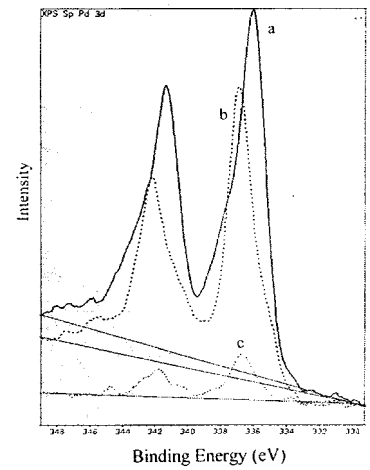


Figure *5:* Pd $3d_{5/2}$ and $3d_{3/2}$ XPS for un-annealed, and annealed Pd/SiC samples. (a) un-annealed Pd/SiC, (b) annealed at 300 "C, and (c) annealed at 400 'C.

98 URC043

Mobile Robot Navigation Enhancement Using Genetic Programming

D. Battle, J. Baghdadchi, A. Homaifar, and E. Tunstel
NASA Center for Autonomous Control Engineering
Department of Electrical Engineering
North Carolina A&T State University
Greensboro, NC 27411

ABSTRACT

The use of the genetic programming (GP) paradigm has shown the potentiality of computers possibly programming themselves. When in combination with fuzzy logic, GP becomes a favorable tool for automated learning in intelligent control systems.

A coding scheme in the port of C is introduced and applied to the problem of discovering a controller for mobile robot path tracking. An improvement in execution speed will be highlighted in comparison to the existing scheme developed in the LISP domain.

1. INTRODUCTION

In recent years, increased efforts have been centered on developing intelligent control systems that can perform effectively in real-time. These include the development of non-analytical methods of soft computing such as evolutionary computation and fuzzy logic, which have proven to be effective among the design of intelligent control systems, and the handling of real-time uncertainty, respectively [4]. In this paper, our efforts focussed on combining these two paradigms to accommodate the development of intelligent control algorithms that will be able to perform in a "real-world" environment. More specifically, genetic programming (GP) is employed to rule learning in the fuzzy domain.

Genetic programming is a method of program induction created by John Koza and James Rice. Because of its adaptive nature, GP can be a useful tool in the realm of artificial intelligence. Tunstel and Jamshidi demonstrated this notion by applying GP to the problem of rule learning to properly steer a mobile robot along a given path. Their efforts highlighted GP's performance to that of tedious trial-and-error from human approaches. Because of its slow execution speed, their implementation was limited to off-line simulations. We intend to improve on this flaw common to highly computational problems by investigating a portable C implementation. This will offer a vast improvement upon the execution speed versus the scenario when LISP (list processing) is used.

The mobile robot-tracking problem will briefly be revisited to exhibit the nature of the problem. A general comparison between the introduced implementation versus the existing implementation within the LISP structure will also be provided.

2. RULE LEARNING

To allow mobile robot control, some type of input information is needed to feed into the system to generate a desired output. Thus the system under control is assumed to have some type of sensory apparatus. For our implementation, we assume that the system has dead-reckoning sensors to provide access to the error states at all times. This sensory input data is then mapped to control outputs according to the desired control policy.

Among the design of controllers with interacting rule-bases, GP has exhibited feasible results among the task of learning rules that make up a given control system. Although an extension-of GA (genetic algorithms), GP is more favorable because it allows rule-bases the freedom to vary in size as well as allow the direct manipulation of linguistic variables.

2.1 Mobile Robot Navigation Problem

The control objective is to find a genetically evolved fuzzy rule-base to successfully navigate a mobile robot along a desired path. We utilize a multiple-input, single-output controller with inputs of measurable position (ε_d) and orientation (ε_θ) errors characteristic of path following. The output is the steering angle (δ) which represents the corrective control action.

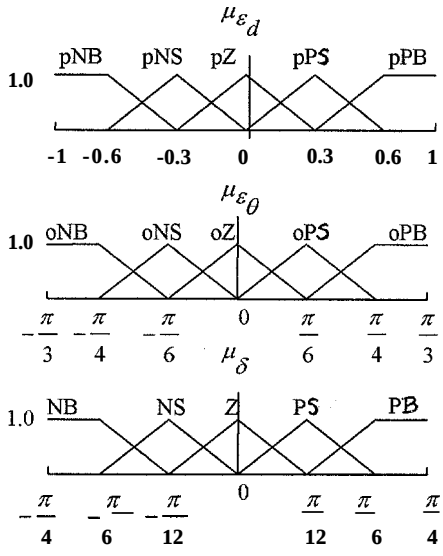
The membership functions are assumed fixed and specified *a priori,* They are reflected in figure 1 along with the hand-derived rule-base generated for the problem.

2.2 Implementation Issues

Functions and terminals. A feature of genetic programming that is important to mention is the minor role of preprocessing of inputs and the postprocessing of outputs [1]. The inputs and outputs are often noted in terms of the terminology within the problem itself Thus, for our implementation, fuzzy linguistic terms and operators are directly imposed. This justifies the use of the following function set,

$$F = \{ANT, CONSQ, f\_AND, IF\text{-}THEN, f\_OR\}$$

Each of these functions requires two arguments with the exception of f_OR which arguments reflect the number of rules within a rule-base.



ε_d \ ε_θ	oNB	oNS	oZ	oPS	oPB
pNB	PB	PB	PB	PB	Z
pNS	PB	p	P	Z	NS
pZ	PB	P	z	NS	NB
pPS	PB	Z	NS	NS	NB
pPB	Z	NB	NB	NB	NB

Figure 1. Hand-derived rule-base and membership functions

The terminal set recognized for this problem is represented by,

$$T = \{pNB, pNS, pZ, pPS, pPB, oNB, oNS, oZ, oPS, oPB, NB, NS, Z, PS, PB, \varepsilon_d, \varepsilon_\theta, \delta\}$$

where *pNB, oNS, Z, etc...* represent the fuzzy attributes (p: position attributes, o: orientation attributes, no suffix: output attributes) and $\varepsilon_d, \varepsilon_\theta$ and δ represent the inputs and outputs.

Syntactic constraints. Due to the manipulation of fuzzy rule-bases, proper selection of functions and terminals is important to ensure correct structure of rules. Hence, rigid constraints must be imposed on the syntax of a genetically evolved rule-base. This is defined by special rules of construction addressed by Tunstel and Jamshidi. In order to avoid severe syntactic violations, they proposed the following rules of construction:

. f-OR can only occupy the root of the tree.
. Only IF-THEN is allowed at the level immediately below the root.
• A left branch of IF-THEN can only be ANT or f_AND.
• A right branch of IF-THEN can only be CONSQ.
• Branches of f_AND can either be ANT or f_AND.
. Branches of ANT can only be input linguistic variables and input fuzzy sets.
. Branches of CONSQ can only be output linguistic variables and output fuzzy sets.

If followed, these rules will insure the proper format for individuals within the initial population, however, we must modify our genetic operators to ensure proper structure among succeeding generations. These resulting operators are referred to as structure-preserving crossover and mutation.

Structure-preserving operators. Structure-preserving crossover prohibits mating individuals to transfer information at different non-root points. The only exception is the validity between the functions *ANT* and *f_AND* provided no violations develop with respects to specifications on the maximum depth of the tree.

Selecting (at random) a non-root point and eliminating all the information succeeding it performs structure-preserving mutation. The discarded information will then be replaced by a random subtree that must also conform to the syntactic regulations.

Fitness evaluation. The fitness measure for a given problem can be described as the driving mechanism for the evolutionary process, Considering that it is fully defined, the fitness measure will be capable of evaluating any individual that it encounters within a run. For our needs, the fitness measure will be measured by how well an individual (rule-base) performs based on several initial conditions. Each rule-base, within a population, will be subjected to different orientation initializations. From Tunstel's analysis, eight was a reasonable choice due to the pair-wise symmetry of the fitness cases, Also, we have to take into account the time-consuming simulation associated with each initialization.

The fitness function used is given by the following equation

$$Raw\ Fitness = \sum_{i=1}^{8} \sqrt{(\varepsilon_d^2 + \varepsilon_\theta^2)_i}$$

In our case, since we are trying to minimize error, the standardized fitness is equivalent to the raw fitness. Thus the lower the fitness measure, the better.

2.3. **Program Performance**

Because of the high computational costs associated with GP, its manipulation in the port of C proved to be very beneficial. In most simulated cases, execution time was decreased significantly, while maintaining the established performance level associated with LISP. Time calculations were done relative to the amount of time it takes the program to find a highly fit rule-base or reaches its termination criterion. Due to the lack of internal time calculations within the provided LISP programs, time was measured by inspection.

Because of to the inaccuracies associated with human measurement, as well as the random search nature characteristic of GP, we can only estimate the improvement in execution. Over a range of 15 consecutive runs for both the LISP and C programs, the average rate of execution speed improvement was approximately 40°/0 in the cases where the termination criterion was reached. This is with respect to using identical parameters that must be provided to the programs,

3. SUMMARY

A mobile robot steering control problem has been revisited for the purpose of demonstrating how genetic programming can be beneficial within the fuzzy domain. Basic implementation issues associated with this were also emphasized to illustrate the importance of special rules.

Some imprecise approximations were made to demonstrate the significance of the investigated code utilization. Overall, the use of C provided noticeable improvement,

4. FUTURE WORK

It is obvious that we have great expectations for the use of genetic programming in discovering fuzzy rule-bases. Due to the improvement in execution speed, as a result of our C implementation, the idea of some actual on-line processing has now been put into perspective.

We intend to use this notion to eventually implement GP in a practical nature. Continued efforts will be made to further improve the execution speed of our GP.

ACKNOWLEDGEMENTS

All runs for the discussed problem were done with a restructured version of Walter Alden Tackett's Simple Genetic Programming in C (SGPC) system.

Thanks to Eddie Tunstel for his help with implementation issues as well as encouragement to use his work as a foundation to build upon. Note that Tunstel's name is referred to often in this paper, thus we do not intend to take acclaim for his efforts.

This work is partially funded by a grant from NASA Autonomous Control Engineering Center (ACE) at North Carolina A&T SU under grant # ACE-48 146. The authors wish to thank ACE Center for its financial support.

REFERENCES

1. Koza, J. R *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press, Cambridge, MA (1992).
2. Koza, J. R *Genetic Programming II: Automatic Discovery of Reusable Programs.* MIT Press, Cambridge, MA (1994).
3. Koza, J. R. "Survey of Genetic Algorithms and Genetic Programming."
4. Tunstel, E. and Jamshidi, M. "On Genetic Programming of Fuzzy Rule-Based Systems for Intelligent Control." *International Journal of Intelligent Automation and Soft Computing, Vol. 2, No. 2 (1996).*
5. Tunstel, E., Akbarzadeh, M., Kumbla, K., and Jamshidi, M. "Soft Computing Paradigms for Learning Fuzzy Controllers with Applications to Robotics."

6. Tunstel, E. "Coordination of Distributed Fuzzy Behaviors in Mobile Robot Control." *IEEE Intl. Conference on Systems, Man, and Cybernetics.* (Oct. *1995), 4009-4014.*
7. D'haeseleer, P. "Context Preserving Crossover in Genetic Programming." *1st IEEE Conference on Evolutionary Computation.* (June 1994), 256-261.

98 URC044

Adaptive Fuzzy Control of a Direct Drive Motor: Experimental Aspects*

E. Medina†, M.-R. Akbarzadeh-T., & Y. T. Kim
NASA Center for Autonomous Control Engineering
Electrical Engineering and Computer Engineering's Building
University of New Mexico, Albuquerque, NM 87131, USA
E-mail: emed@eece.unm.edu

Abstract

This paper presents a state feedback adaptive control method for position and velocity control of a direct drive motor. The proposed control scheme allows for integrating heuristic knowledge with mathematical knowledge of a system. It performs well even when mathematical model of the system is poorly understood. The controller consists of an adaptive fuzzy controller and a supervisory controller. The supervisory controller requires only knowledge of the upper bound and lower bound of the system parameters. The fuzzy controller is based on fuzzy basis functions and states of the system. The adaptation law is derived based on the Lyapunov function which ensures that the state of the system asymptotically approaches zero. The proposed controller is applied to a direct drive motor with payload and parameter uncertainty, and the effectiveness is experimentally verified. The real-time performance is compared with simulation results.

Introduction

Direct drive motors have received increasing attention since they do not have a backlash or dead zone which are caused by gears. The high torque of the direct drive motor allows for the direct connection of the load to the motor axis. Since they are used in high-precision robot and machine tool applications, they must have high resistance to external disturbances. The absence of gear reduction leads to great sensitivity for the motor to variations in the load inertia. In fact, direct drive motors require more robust torque control than the conventional servo motors.

As a result, a linear controller cannot provide a good response under varying load conditions. Variable Structure System (VSS)-type self-tuning control [1], Bang-Bang control [2], and adaptive control [3, 4] have been proposed to handle such problems. However, none of the above mentioned approaches have taken advantage of the robustness of fuzzy logic in the controller design of direct drive motors.

This paper presents a state feedback fuzzy adaptive control method for position and velocity control of a direct drive motor for more robustness to system disturbances. The proposed control scheme is based on a general method outlined in [5]. This method does not require an accurate mathematical model of the system. It allows for integrating heuristic knowledge with mathematical knowledge. It performs well even when mathematical model of the system is poorly understood. The controller consists of an adaptive fuzzy controller and a supervisory controller. The supervisory controller requires only knowledge of the upper bound and lower bound of the system parameters. The fuzzy controller is based on fuzzy basis functions and states of the system. The adaptation law is derived baaed on the Lyapunov function which ensures that the state of the system asymptotically approaches zero. The proposed controller is applied to a direct drive motor with payload and parameter uncertainty, and the effectiveness is verified experimentally.

The motivation for the experiment is to experimentally verify simulation results previously presented in [6]. The necessity of reading the spatial signals, calculating the control signals and implementing them quickly meant a very fast processor needed to be utilized to handle the adaptive real-time control scheme. A Texas Instruments TMS320C30 digital signal processor (DSP) was chosen to handle the immense numerical calculations needed. The DSP chip resides on a Tiger 30 DSP board produced by DSP Research, Inc.

*This work was supported in part by NASA contract # NC CW-0087
† Currently a member of technical staff, Z-TECH Inc., Albuquerque, New Mexico

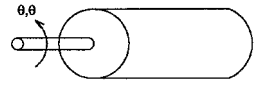


Figure 1: Direct-drive motor and the control parameters

The real-time system consists of three parts: a control mechanism, a data transfer mechanism, and the experimental testbed for the direct drive motor. Our objective is to control a Direct Drive motor (DD) to follow a desired trajectory. The high torque of DD motors allows for the direct connection of the load to the motor axis. Because of this, the motor becomes very sensitive to the load inertia applied. Therefore, we must design a controller that is robust with respect to the applied loads to the motor. In this paper a first-type adaptive fuzzy controller,i.e. one in which the adaptive parameters are linear, is used to meet our objective.

The control mechanism is programmed in the C language, compiled and downloaded to the DSP board. The control scheme utilizes Wang's method of adaptive control [5]. The objective is to control a direct drive motor to follow a desired trajectory. The adaptive controller minimizes the error between the desired and actual trajectories using fuzzy logic methods. The data transfer mechanism is a custom designed data acquisition (DAQ) system, which facilitates rapid parallel data transfer. The data acquisition system consists of IO ports, along with digital-to-analog (DAC) and analog-to-digital (ADC) converters and signal conditioning circuits. The spatial data from the control system is read using an optical encoder located on top of the direct drive motors shaft. Data read in from the control system is manipulated by the control mechanism on the DSP to produce control signals. Control signals are transferred to the system via the DAQ system.

In the following section a linear model of a direct drive motor is presented. The next section shows the derivation of the first-type fuzzy adaptive controller, where stability is guaranteed by the Lyapunov baaed supervisory controller. Then simulations of the DD motor which shows robustness of the fuzzy controller to parameter variations within the model of the DD motor is shown. Finally, experimental aspects of the design are discussed.

Mathematical Model

A DD motor is modeled and a state-space model is developed. Figure 1 shows the control parameters, θ and θ, which correspond to $\dot{x}_1$ and $\dot{x}_2$, respectively. The mathematical model of a DD motor is represented by the following differential equations,

$$J\ddot{\theta} + D\dot{\theta} = \tau \tag{1}$$

where,

J - Inertia moment of the system load and rotor
D- Coefficient of viscous friction term
θ - Angular displacement of the motor (output)
τ - Output torque of the motor (control input, $u = \tau$)

Describing (1) in state space form we have a second-order system of the form,

$$\dot{x}_1 = x_2 \tag{2}$$

$$\dot{x}_2 = f(x, \dot{x}) + bu \tag{3}$$

where f is an unknown function, b *is an* unknown constant, and u and y are the input and output, respectively. After substitution, the above equation become,

$$\dot{x}_1 = x_2 \tag{4}$$

$$\dot{x}_2 = \frac{-D}{J}x_2 + \frac{1}{J}u \tag{5}$$

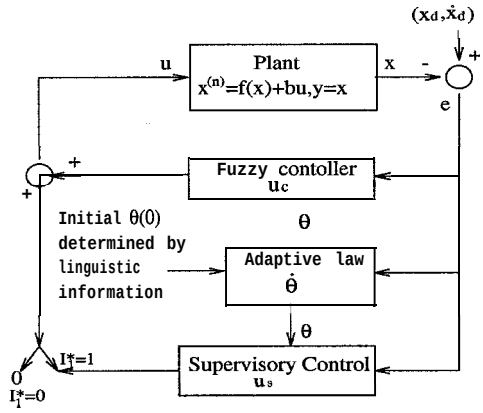


Figure 2: Block diagram of direct adaptive fuzzy control system

from which,

$$f(x_1, x_2) = \frac{-D}{J} x_2 \tag{6}$$

$$b = \frac{1}{J} \tag{7}$$

b is lower bounded and f is upper bounded and both exist.

Control Mechanism

BACKGROUND

The adaptive fuzzy controller designed for the control objective follows the work of Wang [5], refer to figure 2 for a block diagram of the control scheme. The resulting control law is the summation of a basic fuzzy controller, u_c, and a supervisory controller, u_s,

$$u = u_c(\underline{x}|\underline{\theta}) + u_s(\underline{x}) \tag{8}$$

The basic fuzzy controller, $u_c(\underline{x}|\underline{\theta})$ is a fuzzy logic system of the form,

$$f(\underline{x}) = \underline{\theta}^T \underline{\xi}(\underline{x}). \tag{9}$$

where $\underline{\theta}$ are the adjustable parameters, and $\underline{\xi}(\underline{x})$ is known as the fuzzy basis function. It can be shown that there exists an adaptive law and a supervisor controller such that the closed loop system is globally stable. For more detailed treatment of this topic, one may refer to [5, 6]. Furthermore, it is shown through simulations that the resulting adaptive controller is robust to system disturbances.

DESIGN

The first-type direct fuzzy adaptive controller, is designed to control the direct drive motor. The controller is designed to be stable in the sense of Lyapunov, where the Lyapunov equation must be satisfied. The matrix P was arbitrarily chosen to satisfy the Lyapunov equational,

$$P = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}. \tag{lo}$$

The Lyapunov equation was solved using matlab with the P and Λ_c matrices as defined earlier. In our case, we chose $k_1 = 2$ and $k_2 = 1$ for the Λ_c matrix, which placed the roots of $s^2 + k_1 s + k_2 =$ Oat $s = -1, -1$. The resulting matrix Q was

$$Q = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \tag{11}$$

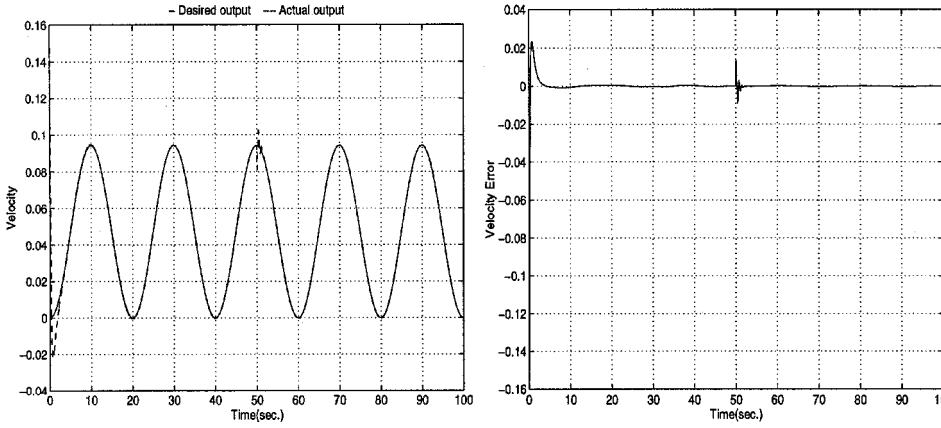


Figure 3: Plot of the velocity and velocity error response of the motor

a symmetric positive definite matrix, thereby satisfying the criterion for stability in the sense of Lyapunov.

The values of $f^U(\underline{x})$ and b_L are calculated. From the state equations we see the values of $f^U(\underline{x})$ and b_L are dependent on the .7 and D parameters of the motor. The parameters are motor dependent. In our case the values of J and D were chosen, respectively as $0.3 \leq J \leq 1.0$ and $0.001 \leq D \leq 0.1$. Therefore the values for $f^U(\underline{x})$ and b_L are $0.001x_2$ and 1, respectively.

Fuzzy basis functions preconstructed using the above five fuzzy membership functions according to the relation (10).

The fuzzy basic controller is now defined as

$$u_c(\underline{x}|\underline{\theta}) = \underline{\theta}^T \underline{\xi}(\underline{x}). \tag{12}$$

The following adaptive law adjusts the state vector O

$$\dot{\underline{\theta}} = \gamma \underline{e}^T \underline{p}_n \underline{\xi}(\underline{x}) \tag{13}$$

where γ is a constant, $\underline{e}^T$ is the desired trajectory minus the actual trajectory, and $\underline{p}_n$ is the last column of P.

Simulation

The adaptive fuzzy controller was simulated on Matlab using the ODE45 command to solve the differential equations. An M-file was written to describe the system of ordinary differential equations. A simulations was run for 100 seconds using zero for the initial conditions of x_1 and x_2. A disturbance, a change in the load, was introduced after the simulation was started.

In order to simulate a change in load, the values of J and D were initally 0.31 and 0.0077 then were changed 50 seconds after the simulation started to 0.5 and 0.1, respectively. The desired trajectory is defined as

$$x_d = 0.15(0.1\pi t - sin(0.1\pi t)) \tag{14}$$

$$\dot{x}_d = 0.015\pi(1.0 - cos(0.1\pi t)) \tag{15}$$

for x_1 and x_2, respectively.

Figure 4, shows the difference between the angle of the actual position and the desired position. It is shown that the difference in the desired and actual output is negligible. After the disturbance is applied, referring to Figure 5, the controller is able to adapt within 5 seconds.

In Figure 3, it is shown that the actual velocity response differs only at the start of the simulation and when the disturbance is introduced to the system. Figure 3 shows the adaptive fuzzy controller is robust with respect to variations in the motor load.

Figure 5, shows the adaptive fuzzy controller is robust with respect to variations in the motor load. Looking at the graph we are able to deduce the adaptive fuzzy controller is able to stabilize the system after the disturbance is introduced within 5 seconds. The controller quickly adapts to compensate for the changing load/inertia condition.

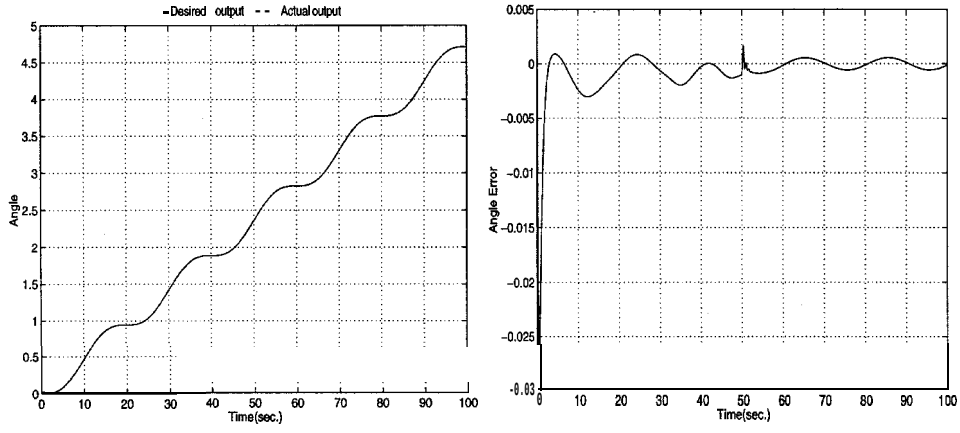


Figure 4: Plot of the angular and angular error response of the motor

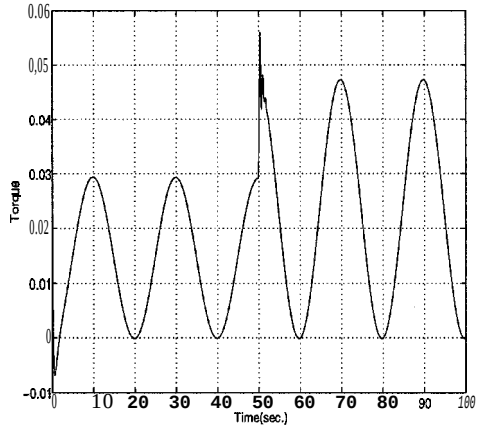


Figure 5: Plot of the control input used to stabilize the DD motor

Experimental Aspects

The experimental testbed is a direct drive motor used to control a flexible link manipulator. Optical encoders are used to sense the position of the motor shaft. The signals from the optical encoders are sent to the signal conditioning circuits of the DAQ system. The position of the motor shaft is also displayed on one of the DAQ boards. The position data is used to calculate the velocity of the motor.

Data Transfer Mechanism

Some of the issues in successful hardware implementation of fuzzy logic controllers are the need for immense computational capability and dual processing of the optimizing algorithm as well as of the fuzzy controller. To remedy this situation, a Digital Signal Processor (DSP) board may be used in combination with a Pentium processor. Specifically, DSP Research's TIGER 30 board utilizing Texas Instruments TMS320C30 DSP is used which is fully compatible with a Pentium- 133 Processor board. The TMS320C30 has a powerful instruction set and is capable of 40 million floating point operations per second (MFLOPS). The TIGER board has the ability to be interfaced with a PC allowing experimental data to be saved in a file for analysis.

A medium for data transfer between the control system and the DSP board was necessary. A generic data acquisition system was designed to implement this need. The capability of asynchronous and 16-bit data transfers were two features the DAQ system needed. The DAQ system designed consists of: eight ports -four inputs and four outputs, digital to analog converters, analog to digital converters, and signal conditioning circuits.

A program written in C is used to transfer data to and from the control system. The DAQ system is assigned an 1/0 memory location in the TMS320's memory map. The memory location assigned to the board in 0x800400 with every other port offset by 0x02. Reading or writing to the above memory locations in the DSP memory accesses the ports of the DAQ system. When the DSP board is accessing the exteral memory its IOSTRB output is asserted low. The IOSTRB output in combination with a 3-to-8 decoder (SN54S138) is used to select which board as well as which of the eight ports is to be selected.

Each input port of the system consists of a pair of octal tri-state transceivers (SN54KS245) in series with a set of octal D-type flip-flops (SN54273), allowing 16-bit data transfers. The tri-state transceivers were necessary in order to isolate the ports from one another, because they are all connected in parallel to the expansion bus of the DSP board. The flip-flops are used to store the data, thereby keeping the input signal from being compromised [7].

The output ports are also pairs of octal tri-state transceivers. The inputs to the transceivers are connected to the output of the DSP boards expansion bus. Control signals generated by the DSP are transmitted from its expansion bus to the inputs of the transceivers. The output port is decoded and the control signal is output to the control system.

Conclusion

In most practical systems, some type of a mathematical model is available. Although this mathematical model is often corrupted by unknown parameters, disturbances, etc., it still constitutes a very important portion of a systems engineer's knowledge base. In this paper, heuristic knowledge (fuzzy logic rule base) is allowed to combine with a model-based controller to achieve a more robust control system.

A first-type fuzzy adaptive control method was introduced for position control of a direct drive motor. The control method presented does not require an accurate mathematical model of the control system, it is capable of incorporating fuzzy control rules directly into the controller and guarantee global stability of the resulting closed-loop system in the sense that all signals are asymptotically convergent. The controller was adaptively changed by monitoring the difference between the desired trajectory and the actual trajectory.

The control method discussed was shown to be robust with respect to the load on the motor, i.e. the system can be adaptively stabilized even though the load on the motor changes. Currently, the software for experimental testing of the control algorithm is being debugged.

References

[1] K. Furuta, K. Kosuge, & K. Kobayashi, "VSS-Type Self-tuning Control of Direct Drive Motor," *Proceedings of the IECON*, pp. 281-286, 1989.

[2] K. Furuta & S. Kobayashi, "Bang-Bang Position Control of Direct Drive Motor," *Proceedings of the IECON*, pp. 148-153, 1990.

[3] Y. T. Kim, B. C. Kim, & H. D. Lee, "Adaptive Control of a Direct Drive Motor," *Proceedings of the KIEE*, Korea, 1994.

[4] K. Furuta, K. Kosuge, & K. Kobayashi, "Self-tuning, Sliding Mode and Adaptive Control for Direct Drive Motor,"PToceedings *of the IECON*, pp. 459-465, 1988.

[5] L.-X. Wang, "Adaptive Fuzzy Systems and Control: Design and Stability Analysis," *Prentice Hall* International *Inc.* Englewoods Cliffs, NJ, 1994.

[6] E. Medina. Y.T. Kim, & M.-R. Akbarzadeh-T., "Adaptive Fuzzy Control of a Direct Drive Motor," Proceedings *of the NASA University* Research *Center Technical* Conference, Albuquerque, NM, 1997.

[7] M-R. Akbarzadeh-T., M.E. Medina, and M. Jamshidi, "DSP Implementation of Evolutionary Fuzzy Control," Proceedings *of the NASA University Research Center Technical Conference,* Greensboro, NC, 1996.

98 URC045

HYBRID FUZZY-PD CONTROL FOR A DOCK MOUNTED PANTRY CRANE

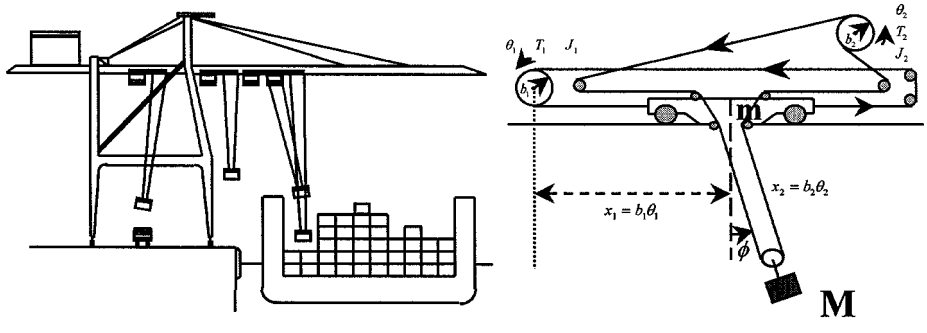
Bahrain Kimiaghalam, Tel: (336) 334-7760 x259, Email: bahram@ncat.edu
Abdollah Homaifar, Tel: (336) 334-7760 x221, Email: homaifar@ncat.edu
Marwan Bikdash, Tel: (336) 334-7760 x225, Email: bikdash@ncat.edu
NASA-Autonomous Control Engineering Center
Department of Electrical Engineering
NC A&T State University, Greensboro, NC

Abstract We study in this paper the performance of a hybrid fuzzy/Proportional-Derivative controller used to control and plan the motion of a heavy payload carried by a dock-mounted pantry-type crane. The fuzzy controller decides about the desired speeds of the load based on its present location and the PD controller translates the desired speeds to necessary torques to get to the desired speeds. By changing the shape and range of the membership functions and /or rules using common sense and trial and error, we can adjust the performance of controller to get to a reasonably acceptable result.

KEYWORDS: Crane control, Hybrid control, Fuzzy control, Proportional-Derivative Control

1. INTRODUCTION

Dock mounted cranes are supposed to move very heavy loads under performance and safety requirements on driver riding comfort, carrying time and collision prevention. Fast changes in load, trolley positions and velocities are ill advised. Another difficulty is due to the nonlinear dynamics of the crane. In this paper, we design a fuzzy controller for a dock mounted container crane to be used as a basis for ship crane control. Since this is a difficult nonlinear control problem, it is not easily solvable with conventional control methods. The solution of this problem by optimal control theory would lead to solving a two-point boundary value problem (Sakawa and Shindo, 1982). This problem has been approached by methods such as predictive fuzzy method (Yasunobu and Hasegawa, 1986) and it seems that simple fuzzy is not been used yet because it hasn't been thought to be powerful enough for this problem. Here the simplicity of the design was one of our objectives but to improve the performance one can increase the number of inputs or attributes of the Fuzzy Inference System (FIS). Our fuzzy logic process is similar to human decision making. We discuss a Hybrid Fuzzy/PD Controller design that determines the desirable speeds of the trolley and hoist from which the controls can be computed using a PD controller. The hybrid Fuzzy/PD controller is similar to what an operator does. In these controller, the horizontal and vertical positions each have four attributes and there are sixteen rules. The membership functions used are Gaussian with different standard deviations. Each output has three attributes with Gaussian memberships. We have simulated and illustrated the results for different adjustments.



a: The dock based container crane

b: Coordinates and Torques

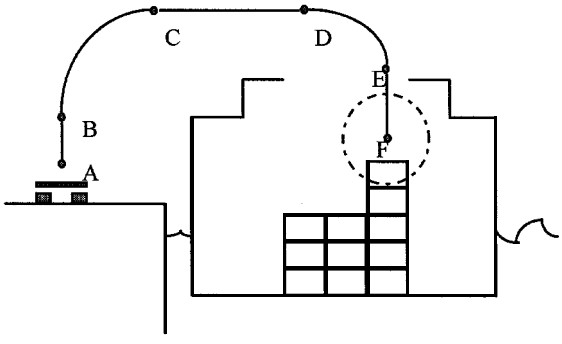
Figure 1: Crane geometry

Deriving the states for the system has been done in [1] and [8]. We just here mention that the state-space equations where ***Xi*** are the state-variables, and ***v,*** and v_2 are the controls, are:

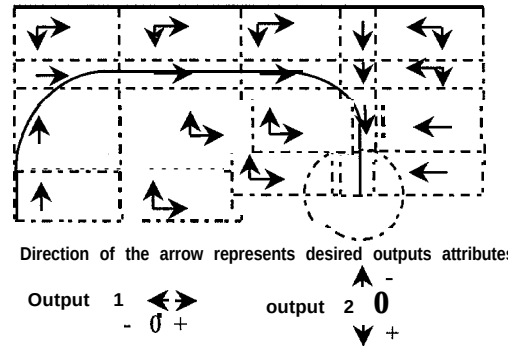
$$x_1 = x_4,\ \dot{x}_4 = v_1 - \delta_1 x_3 v_2 + \delta_1 g x_3$$

$$x_2 = x_5,\ \dot{x}_5 = -\delta_2 x_3 v_1 + v_2$$

$$\dot{x}_3 = x_6,\quad \dot{x}_6 = -\frac{1}{x_2}[v_1 - \delta_1 x_3 v_2 + (1+\delta_1) g x_3 + 2x_5 x_6]$$



a: A" desired" trajectory

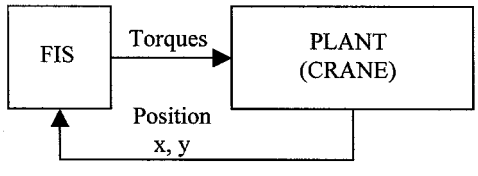


b: Conceptual rule assignment strategies

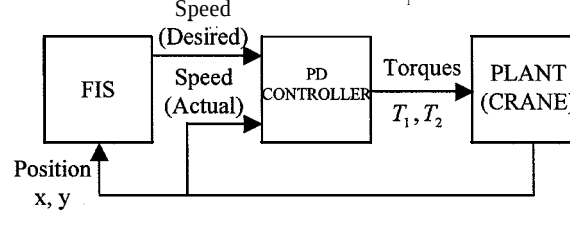
Figure 2: Crane operation

The objective of the control system is to move the load, in a relatively short time, from one side of a fixed obstacle to a known destination on the other side, without having the load collide with the obstacle and. We can make the control problem easier by separating the operation into two phases: a hoist-lowering phase and a translation phase. Then we could assume that the length is a known constant during the translation phase. It would save time if the two phases of operation could be done concurrently. As shown in figure 2.A, there is no collision with obstacles between points A and F in a typical desired trajectory of the load and there are concurrent vertical and horizontal motions from point B to C and D to E. Of course, swinging may cause a big delay in coming to rest at the destination.

2. HYBRID FUZZY/PD CONTROLLER



Input: Position of the load
Output: Trolley and hoist torques



Input: Position of the load
Output: Trolley and hoist speeds

Figure 3: Simple fuzzy and Hybrid Fuzzy/PD designs

We frost started by designing a simple fuzzy controller as in figure 3.a, but the result was not satisfactory and we had to increase the number of attributes to get a better result, which was against the simplicity objective. Then we tried to mimic the operators' action and we designed a Hybrid Fuzzy/PD controller as shown in figure 3-b. The FIS is responsible to find the appropriate speed of the load based on its position and the PD controller will find the amount of torques needed to achieve the desired speed as fast as possible. The Hybrid Fuzzy/PD controller will lead the load until it reaches the dotted circle shown in figure 2 at which point a linear controller takes over. Figure 2.b shows the

rule assignment based on the position of the load. The directions of the arrows represent the desired attributes for the outputs.

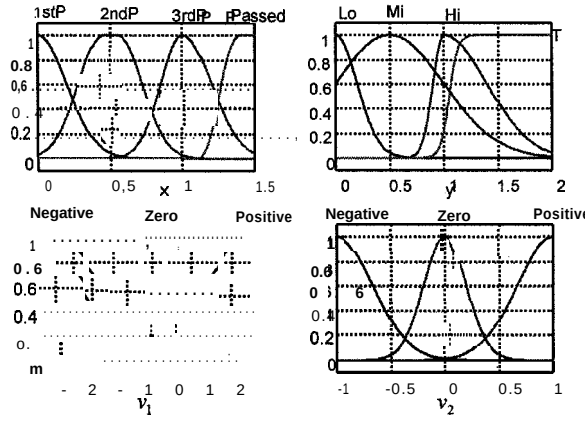


Figure 4: Membership functions for inputs (x and y) and outputs (v_1 and v_2)

The FIS is relatively simple and uses Mamdani's operators. A sample rule is:

If" x " is in its "First part" and " y " is "High" then " v_1 " is "positive" and " v_2 " is "zero"
Which can also expressed as
$x = x^1$ and $y = y^3 \Rightarrow v_1 = v_1^3$ and $v_2 = v_2^2$
Another example rule is
If" x" is in its "Third part" and" y " is "High" then" ***VI" is*** "Zero" and" v_2 " is "Positive"
Or $x = x^3$ and $y = y^3 \Rightarrow v_1 = v_1^2$ and $v_2 = v_2^3$

The rule base can be obtained from figure 2.b. that will be based on the position of the load,

The membership functions chosen for FIS are mostly Gaussian, as shown in figure 4.

3. RESULTS

We simulate the motion of the load until it reaches the dotted circle, shown in figure 2, so the timing history that we reported here depends on the time that it takes for the controller to transfer the load within the dotted circle.

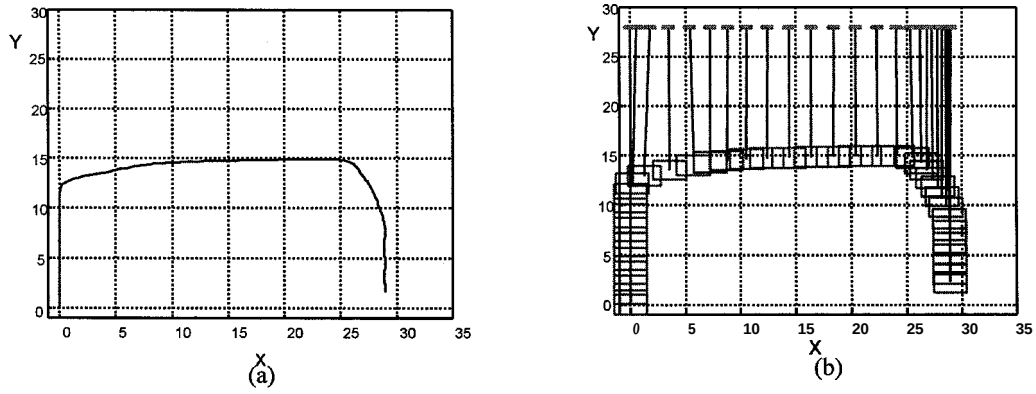


Figure 5: Trajectory of the load

Figure 5 shows the frost result in two graphs. Figure 5 .a shows the trajectory of the load is very similar to the desired trajectory shown in figure 2.a. Remember here we are dealing with container crane carrying huge and heavy payloads. For example this simulation is done for a 45 tons load. The small differences comparing to desired trajectory are acceptable. As far as the control objective is concerned, we just need to move the load over the obstacles and suppress the swinging of the load. Swinging of the rope can be seen at Figure 5.b, which shows the swinging is very small and negligible.

Unfortunately this control is slow. The whole process from origin to inside the dotted circle around the destination takes about 74 seconds and that is about twice longer than what we expected. To speed up the system we changed the range of output membership functions as shown in figure 6. All the memberships are the same as the previous one, except the output membership functions have a bigger range on the x-axis. It does not change the attributes but does have an effect similar to multiplying the output of the previous FIS by a constant or in another word amplifying the output.

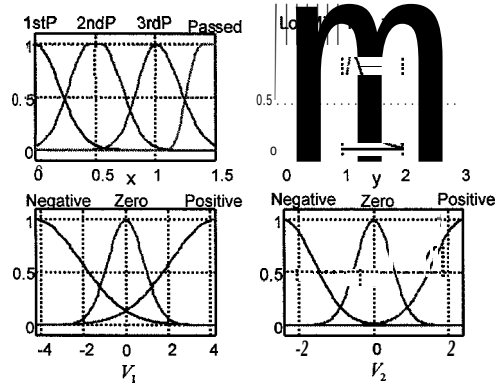


Figure 6: New memberships with expanded range for the outputs

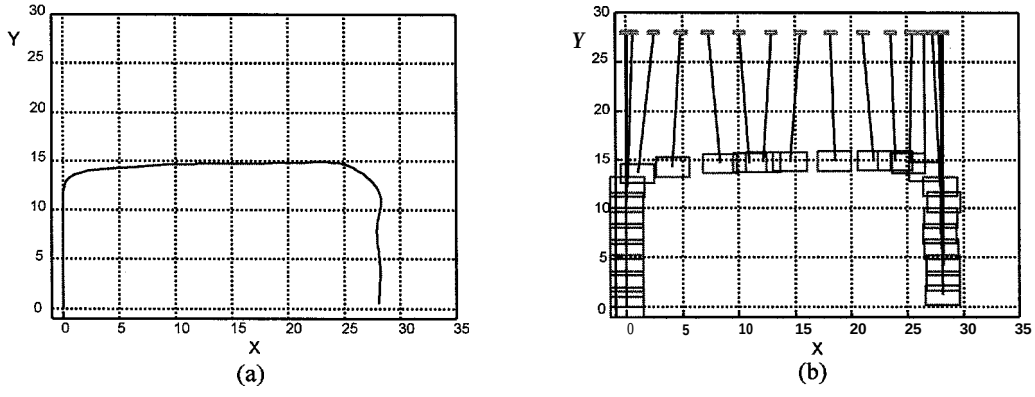


Figure 7: Effect of expanding the output range

After this change the result as seen in figure 7 is still satisfactory. The trajectory is very similar to shape of the desired trajectory. The swinging observed from figure 8.b are slightly more than the fist case, however it is still negligible. The whole process in this case takes about 44 seconds is faster by $^{0}/_{0}40$ as compared to 74 seconds in the previous case and is reasonably acceptable.

Speeding up the system should be done to the degree that will not create undesired trajectory and swinging. Figure 8 shows a case with overspeeded outputs. The trajectory close to destination is distorted and the swinging is not acceptable. This control may cause collision between the load with some obstacles. The control is fast and will take less than 44 seconds to reach to dotted area. Unfortunately there will be swinging towards the end and we have to wait longer time till we can release the load from the hook.

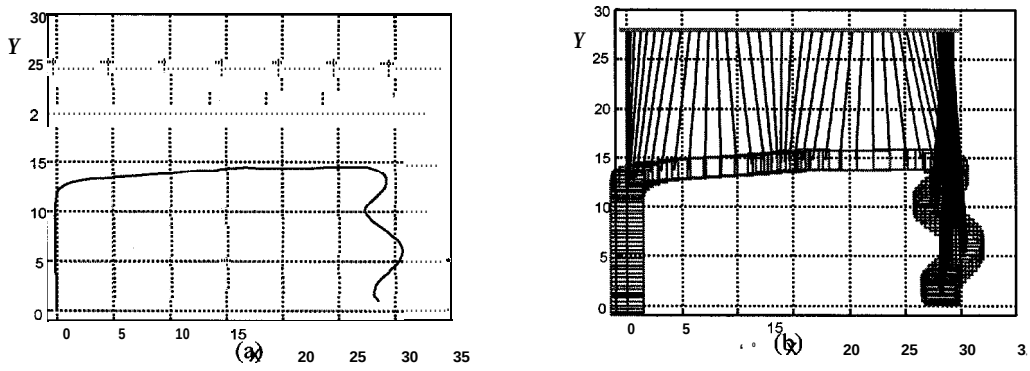


Figure 8: Effect of overspeeding the system

4. CONCLUSION

We developed a Hybrid Fuzzy/PD controller for a dock-mounted pantry crane. There are no external forces and disturbances to the system. Considering the difficulty of the problem and the simplicity of our design, the result was satisfactory. We increased the performance by minor changes in the output membership functions. It can be improved further by increasing the number of attributes of inputs and outputs and/or replacing the PD controller by a fuzzy-PD controller, which in payoff will increase the complexity of the design. This experience shows the power of fuzzy logic in the area of nonlinear control and it is a good starting point to design a ship mounted crane control, which is in demand. In that case one needs to control the swinging caused by ship motion. For such problems however, more complex designs may be needed.

5. ACKNOWLEDGEMENT

The U.S. Office of Naval Research through contract No. N00014-96-1-1123 supports this work. The authors would like to thank them.

REFERENCES

1. Sakawa, Y., and Y. Shindo "Optimal Control of Container Cranes," ***Automatic,*** Vol. 18, No. 3 (1982), pp. 257-266
2. Moustafa, Kamal and A. M. Ebid "Nonlinear Modeling and Control of Overhead Crane Load Sway," ***Transaction of the ASME,*** Vol. 110 (September 1988), pp. 266-271.
3. Filess, M. and J. Levine and P. Rouchon "A Simplified Approach of the Crane Control Via a Generalized State-Space Model," ***Proceedings of the 30th conference on decision and control,*** Brighton, England (December 1991), pp. 736-741.
4. Meirovitch, L. Analytical Methods in Vibrations, Macmillan, New York (1967).
5. Yasunobu, S., and T. Hasegawa "Evaluation of an Automatic Container Crane Operation System Based on Predictive Fuzzy Control," ***Control-Theory and Advanced Technology,*** Vol. 2, No. 3 (September 1986), pp. 419-432
6. Yasunobu, S and T. Hasegawa (1987), "Predictive Fuzzy Control and its Application for Automatic Container Crane Operation System," ***Preprints of second IFSA Congress,*** Tokyo, July 20-25 (1987).
7. Lee, C.C. "Fuzzy logic in control systems: fuzzy logic controller: Part I & Part II, " ***IEEE Trans. Syst. Man and Cybernetics,*** Vol. 20 (March /April 1990), pp. 404-435.
8. Kimiaghalam, B., A. Homaifar, and M. Bikdash, "Using Genetic Algorithms for optimal crane control" NASA-URC Conference (February 1998)

Fuzzy Control of Flexible-Link Manipulators: A Review[1]

M.-R. Akbarzadeh-T, S. Quintana, and M. Jamshidi
Autonomous Control Engineering
EECE Building
University of New Mexico,
Albuquerque, NM 87131, USA
E-mail: akbazar@eece.unm.edu

Abstract

Several recent research efforts are reviewed here which have applied fuzzy logic in control of flexible-link manipulators. A flexible robot is a distributed parameter system represented by complex nonlinear dynamics, its actuator and the control parameters are non-colocated, and lastly, unstructured/unknown parameters play a significant role in model dynamics of a flexible robot operating in the real world. As a result, control of flexible robots is considered a promising area for application of intelligent control methodologies such as fuzzy logic, genetic algorithms, and neural networks.

I. Introduction

Flexible robot control has received widespread attention within the control community for the past two decades. Existing rigid and lumped models of robotic systems and their corresponding controllers have performed well in industry since the assumption of rigidity is validated by using bulky and rigid material and limiting the speed of the manipulator. This has caused a low power efficiency, large weight to payload ratios, typically in the range of 30-50, and slow movement of robotic arms. These poor characteristic are typical among the existing robotic systems in factories and space applications. There is a clear need for improving system performance. For example, the robotic arm of the NASA space shuttle, Columbia, moves at a staggering rate of 0.00088 rad/s [1,2]. Due to shuttle's loading constraints, the arm is designed to be large, light weight and slim, while maintaining the ability to operate on heavy payloads. Consequently, it has to move very slowly not to excite any of the flexible modes. Another example is the Adept II. Adept II is a popular industrial robot with an admirable repeatability rate. However, it weighs 400 pounds whereas its nominal payload capability is only 13.2 pounds [3]. Obviously, future generations of robotic systems will need to be designed to move faster, be more power efficient, be more lightweight and slimmer in order to satisfy the increasingly fast and precise requirements of the future tasks. Ignoring the distributed flexibility in robot links can result in instability and undesirable performance in applications where precision, reliability and repeatability are extremely important.

Flexibility is usually an undesirable behavior and is a direct consequence of slimmer design of links, faster robot movement and/or higher payload ratio. Only in few instances, it is a built-in feature used to enhance system performance such as in [4] where barrel flexation is used to speed up the response of a tank gun-pointing problem. In most situations, such as the treatment in this paper, the feedback control system is designed to dampen/eliminate the motions resulting from flexible modes and to move the tip of the link in a desirable fashion.

From the perspective of the control theory, the non-colocated nature of the actuator (motor) and the link's tip is an interesting problem. The technology of achieving stable control using non-colocated sensors and actuators is a fundamentally very challenging control problem. In this paper, we will discuss the current and the past research activities in the field of fuzzy control of flexible robots.

IL Fuzzy Logic and Flexible Robot Control

[1] This work was supported in part by NASA contract # NCCW-0087 and by Waste-management Research and Education Consortium nuder contract # WERC/NMSU/DOE Amd 35.

One of the papers which coincided with a surge in interest in this area was a paper by Vidysagar in 1987 [5] in which he identified several uprising issues in control system technology, one of which was the field of flexible robotics. One of the issues addressed by Vidyasagar in [5] was in regards to the modeling of flexible robots. Although certain mathematical models of flexible robots are feedback linearizable [6], other models are not [7]. This raised an interesting distinction between the physical object, the flexible robot, and its mathematical model. What does it mean when one mathematical model is feedback linearizable and the other is not? When should one use a pinned-free model, or a cantilevered model? A given system can best be represented by both of these models at different times. What does this conclude about the exact nature of the system?

This uncertainty in mathematical modeling enhanced by the existence of unstructured and/or unknown parameters in a manipulator's environment raises the question of the true applicability of model-based conventional control techniques in the real-world environment. A given system which is proven to be stable based on rigorous theoretical calculations and mathematical derivations and even which performs well in a laboratory environment, may not necessarily exhibit the same desirable behavior in the real world when faced with heat, dust, and aging of system components. This is why the introduction of fuzzy logic to control of flexible robots was almost inevitable.

Fuzzy controller which are designed for flexible robots can in general be divided in two categories: hybrid and pure. Most of the research in this area has been devoted to hybrid combinations of classical control techniques with fuzzy logic expert knowledge \cite{mccullogh,liu,tzes1,tzes2,garcia-benitez}. Even though these approaches usually require some kind of a mathematical representation of the flexible robot, they still exhibit robustness to system uncertainty due to their ability to incorporate expert knowledge in their control. In the following sections, we will discuss and categorize various approaches which use fuzzy logic in some form in the control of flexible robots.

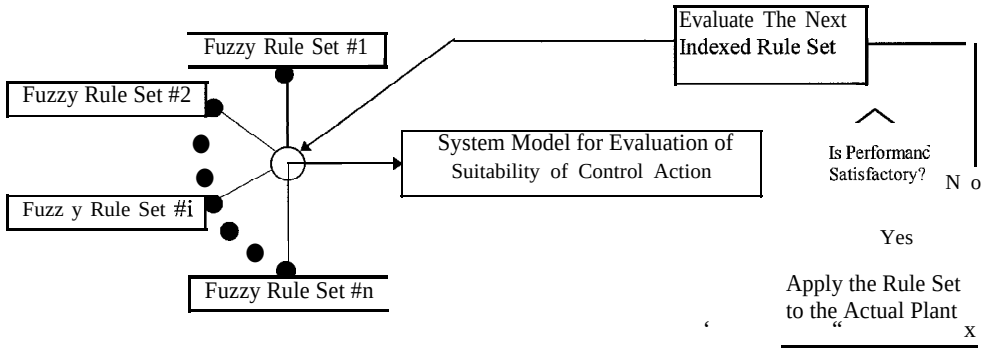


Figure 1. Anticipatory Fuzzy Logic: Selection of appropriate fuzzy rule set by evaluating the performance using a simulated system

III. Hybrid Fuzzy Controllers

Anticipatory Fuzzy Control

Among the proponents of fuzzy logic, it is a commonly held belief that fuzzy logic has the capability to handle uncertainty and to exhibit robustness to unknown system parameters. In 1994, McCullough [8] verified this belief for the case of flexible robots by comparing three controllers: an "anticipatory" fuzzy logic controller, a NN-based controller which is trained to emulate the fuzzy logic controller, and a Linear Quadratic Regulator (LQR) type controller. The LQR Controller was chosen as a benchmark controller to compare the results of other two controllers. The "anticipatory" fuzzy controller was developed by combining traditional fuzzy logic with a system model to predict system behavior. A Neural Network was

then trained to emulate the fuzzy controller. In the abscence of noise, all three controllers performed equally well. However, when the system was injected with noise, the LQR showed large oscillations (6 in magnitude) while the fuzzy controller showed only small oscillations (0.2 in magnitude), and the NN showed almost no oscillation (not visible). In this study, the fuzzy logic controller and the Neural Net emulator showed a strong sense of robustness to system noise which was not shared with the standard LQR controller.

Model-based Controllers with Fuzzy Logic Supervisor

In [9], a model-based feedback linearizing controller with an embedded PID control law is used to linearize the nonlinear dynamics of a flexible robot. The gain parameters of a PD controller (in an external control loop) are adjusted by a fuzzy expert system. Therefore, the fuzzy expert system can adjust for uncertainties and model approximation.

In [1 O], a hierarchical control strategy is used where the fuzzy "supervisor" is a simple rule-based system which chooses from the following three model-based control strategies: input shaping, tip acceleration feedback, and shaft position-endpoint acceleration feedback. The input shaping technique has shown a great deal of promise and has performed very well in simulations and also in some controlled laboratory experiments. However, it is a model-based method and can easily lose performance if unstructured parameters are introduced to the environment. The addition of the other control loops using fuzzy logic adds a degree of robustness and disturbance rejection to this technique.

Neural Network and Fuzzy Logic Controllers

In [11], a series combination of NN and fuzzy logic is used to eliminate vibrations due to backlash and non-linearities introduced by motor's amplifier in saturation. The NN resembles a Pulse Width Modulated (PWM) controller which is optimized by a back propagation algorithm to minimize the square of the error in tip deflection. The fuzzy logic gain controller adjusts the NN output based on l_2 and l_{infity} of the error. The final error converged to zero although oscillations sustained for several seconds. In [12], Tzes developed a hybrid scheme involving both frequency and time domain techniques. The frequency identification algorithms estimate the plant's open loop eigen-values . Based on this information, a fuzzy rule-set chooses the feedback gains of a PD control such that the closed loop eigen-values of the system are moved toward desirable values.

Genetic Algorithm and Fuzzy Logic Controllers

Genetic Algorithms (GA) are optimization routines modeled after Darwinian theory of natural evolution. GAs have 'shown the ability to optimize multi-objective functions with several_ variables such as in parameters of fuzzy logic controllers. In [13,14], the ability of GA is exploited in optimization of a fuzzy logic controller for a flexible robot. Parameters of membership functions are optimized and system response is shown to improve. The GA contains a population of potential solutions and uses a system model to evaluate potential solutions to the optimization problem. As a result, it is similar to the anticipatory fuzzy approach where a number of static solutions are simulated iteratively until a suitable rule set is found. The difference is that in GA, the set of solutions are dynamic, i.e. they change in time based on the genetic operators of mutation, crossover, and reproduction.

IV. Purely Fuzzy Logic Controllers

The second category of application of fuzzy logic in flexible robot control is where only fuzzy system and structures are used to control the system [15-19]. In [15], a fuzzy inverse model-based controller is proposed by using the converted fuzzy relation matrices of the relational equation. The resulting algorithm is applied to vibration suppression of the flexible modes.

The paper by Moudgal, Passino and Yurkovich [16] uses purely fuzzy heuristics to design a fuzzy controller for a flexible manipulator. In [17], tip acceleration and motor's angular position sensory data are incorporated in a fuzzy hierarchical structure to determine proper control action. By monitoring the error,

The higher level hierarchy of fuzzy controller adjusts for input and output scaling of the lower level fuzzy controller. The simulation and experimental results demonstrate that this controller is superior to all of the existing model based conventional controllers in terms of speed of response, minimal residual vibrations, overshoot, and smoothness of response.

In a later paper, Moudgal, et. al. [17], expanded the hierarchical structure to include a reference model, a fuzzy inverse model, and a knowledge-base modifier. The proposed architecture provides for learning and adaptation to the desirable reference model. Even though the controller initially begins with no expert knowledge, the learning mechanism is able to determine proper rules such that the system behaves well after only a short delay. Here again, the flexural effect of the link is ignored and the rule set is only a function of the sensors at the tip and the base of the arm.

All of the above papers have one common theme in that the non-colocated sensory information from the sensors at the tip and the base of link are used to form the basis of the control decision. In [18, 19] spatial features of the arm are considered in the decision making process. Due to wave nature of a distributed parameter system, the response may be improved if the spatial features of the arm are considered in addition to the position of the base and tip of the arm. In other words, the points between the base and the tip contribute to the behavior of the system and should be considered in the overall control strategy. In [18,19], the distributed nature of the flexible link is exploited through a hierarchical fuzzy structure and by installing several strain gauges throughout the length of the arm. The higher level of the hierarchy determines spatial features of the arm such as *Oscillatory, Curved,* and *Straight* The lower hierarchy determines the proper control action depending on the response from the feature extraction routine and position measurements from the tip and base of the arm.

IV. Conclusion

There are several approaches which have utilized fuzzy logic in control of flexible robots. In this review, we have mentioned combinations of classical model-based techniques with fuzzy logic, combinations of Neural Networks with fuzzy logic, Genetic Algorithms with fuzzy logic, anticipatory fuzzy logic, and purely fuzzy logic controllers. The variety of approaches illustrates the suitability of flexible robotic systems as a desirable platform for testing and comparing novel control strategies.

V. References

[1] Nguyen, P. K., Ravindran, R., Carr, R., Gossain, D.M., "Structural Flexibility of the Shuttle Remote Manipulator System Mechanical Arm," *Proceedings of the Guidance and Control Conference, AIAA*, Paper no. 82-1586, New York, New York, August 1982.

[2] Cetinkunt, S. and Yu, W.-L., "Closed-Loop Behavior of a Feedback-Controlled Flexible Arm: A Comparative Study," *The International Journal of Robotics Research,* 10(3):263-275, June, 1991.

[3] Adept Technology, Inc., AdeptTwo Robot Manual.

[4] Lewis, F.L., Dawson, M.D., Lin, J., and Liu, K., "Tank Gun-Pointing Control with Barrel Flexation Effects," *Proceedings of The ASME Winter Annual Meeting,* Dec. 1-6, Atlanta, 1991.

[5] Vidysagar, M., "System Theory and Robotics," *IEEE Control Systems Magazine,* pp. 16-17, April 1987.

[6] Spong, M. W., "Modeling and Control of Elastic Joint Robots," *Proceedings of the ASME Winterr Annual Meeting,* Anneheim, CA, Dec. 1986.

[7] Cesario, G. and Marine, R., "On the Controllability Properties of Elastic Joint Robots, " *Proceedings of the International Conference on Analy. and Opt. Systems,* INRIA, Nice, 1984.

[8] McCullough, Claire L., "Neural Net Control Vs. Anticipatory Fuzzy Control For a Flexible Beam : A Comparison," *Proceedings of the IEEE International Conference on Neural Networks,* PP. 2350-2354, Orlando, Florida, 1994.

[9] Liu, K., and Lewis, F., "Fuzzy Logic Control of a Flexible Link Manipulator, " *Journal of Intelligent and Fuzzy Systems,* 2:325-336, *1994.*

[10] Garcia-Benitez, E., Yurkovich, S., and Passino, K. M., "Rule-Based Supervisory Control of a Two Link Flexible Manipulator; *Journal of Intelligent and Robotic Systems, 7:195:213,* 1993.

[11] Tzes, A., Peng, P.-Y., Khorrami, F., "Fuzzy Neural Network Control for a Single Flexible-Link Manipulator," *Proceedings of IFS, 1993.*

[12] Tzes, A., and Kyriakides, K., "A Hybrid Frequency-Time Domain Adaptive Fuzzy-Control Scheme for Flexible-Link Manipulators," Journal of Intelligent and Robotic Systems, 10(3):283-300, 1994.

[13] M.-R. Akbarzadeh-T. &M. Jamshidi, "Evolutionary Fuzzy Control of a Flexible Link," Proceedings of the IEEE International Conference on Robotics and Automation, Albuquerque, New Mexico, 1997.

[14] M.-R. Akbarzadeh-T. & M. Jamshidi, "Incorporating A-Priori Expert Knowledge in Genetic Algorithms," Proceedings of the IEEE International Conference on Computational Intelligence in Robotics and Automation, Monterey, California, 1997.

[15] Yoo, B., Jeong, Sacheul, and Ham W., "Hybrid control of Flexible Manipulator Based on Fuzzy Relations," *Proceedings of the IEEE International Conference on Fuzzy Systems, New* Orleans, LA, Sep 8-11, 817-823, 1996.

[16] Moudgal, V. G., Passino, K. M., and Yurkovich, S., "Rule-Based Control for a Flexible-Link Robot," *IEEE Transactions on Control Systems Technology, 2(4):392-411,* 1994.

[17] Moudgal, V. G., Kwong, A. W., Passino, K. M., and Yurkovich, S., "Fuzzy Learning Control for a Flexible-Link Robot," *IEEE Transactions on Fuzzy Systems,* 3(2):199-210, 1995.

[18] M.-R. Akbarzadeh-T., M. Jamshidi, and N. Vadiee, "A Hierarchical Fuzzy Controller Using Line-Curvature Feature Extraction for a Single Flexible Arm," *Proceedings of the Third IEEE Conference on Fuzzy Systems,* Orlando, Florida, June 26-29,524-529, 1994.

[19] M.-R. Akbarzadeh-T., "A Fuzzy Hierarchical Controller for a Single Flexible Arm, " *Proceedings of the International Symposia on Robotics and Manufacturing*, Maui, Hawaii, August 14-17,411-418, 1994.

GAS SENSORS BASED ON SINGLE-ARM WAVEGUIDE INTERFEROMETERS

Sergey Sarkisov, Michael Curley, Darnell Diggs, and Grigory Adamovsky*

Department of Natural and Physical Sciences (Physics), Alabama A&M University, Normal, Alabama 35762-1268
Telephone: 205-851-5306, Fax: 205-851-5622, E-mail: sergei@caos.aamu.edu

*Instrumentation and Control Technology Division, NASA Lewis Research Center, Cleveland, Ohio 44135

Introduction

Various optical technologies can be implemented in chemical sensing. Sensitive, rugged, and compact systems will be more likely built using interferometric waveguide sensors. Currently existing sensors comprise dual-arm systems with external reference arm [1], dual-arm devices with internal reference arm such as integrated Mach-Zehnder interferometer, and single-arm systems which employ the interference between different waveguide modes [2-9]. These latter ones are the most compact and rugged but still sensitive enough to monitor volatile pollutants such as NH_3 coming out of industrial refrigerators and fertilizer plants and stocks, NO, NO_2, SO_2, emitted by industrial burning processes. Single-arm devices in planar waveguide configuration most frequently use two orthogonally polarized modes TE_i and TM_i of the same order *i* [5-8]. Sensing effect is based on the difference in propagation conditions for the modes caused by the environment. However, dual-mode single-order interferometers still have relatively low sensitivity with respect to the environment related changes in the waveguide core because of small difference between propagation constants of TE_i and TM_i modes of the same order. Substantial sensitivity improvement without significant complication can be achieved for planar waveguide interferometers using modes of different orders with much greater difference between propagation constants.

Single-arm dual-mode waveguide interferometer

Fig. 1 presents possible set-up of such novel single-arm dual-mode interferometer. The waveguide is made as an open asymmetric structure containing a dye-doped polymer thin film onto a substrate with the refractive index lower than that of the film. The waveguide is equipped with input and output prism couplers. Simultaneous excitation of two modes of orders O and 1 is achieved by a focusing lens placed in front of the input coupler. In this configuration of the interferometer, mode combining after outcoupling from the waveguide is done by a mirror and a beam splitting cube. Photo detector reads the light intensity of interference pattern. Actually, there are four modes in the device: two orthogonally polarized modes for each order. All these modes can be employed in different types of interferometry. Combinations (TE_0, TM_0) and (TE_1, TM_1) will correspond to traditional polarimetric interferometers. These combinations are achievable with polarizers oriented at 45° and placed in separate beams for each mode order. Combinations (TE_0, TE_1) and (TM_0, TM_1) correspond to pure dual-mode interferometry which is more sensitive for the environment change which affects differently the propagation conditions for the modes of different order. One can also utilize combinations (TE_0, TM_1) and (TM_0, TE_1) that are the hybrids of polarimetric and dual-mode interferometry. Thus, we can expect the

combination of good sensitivity to the change of chemical environment as well as to the change of physical conditions. Various types of mode combinations in the interferometer can be implemented with output polarizers. For example, the hybrid interferometer uses the first polarizer with 0° or 90° orientation placed separately to select TE or TM mode of a particular order, the second polarizer with 45° orientation is common for both mode orders. Combination of polarimetric and dual-mode interferometry in this novel system adds more flexibility to its performance.

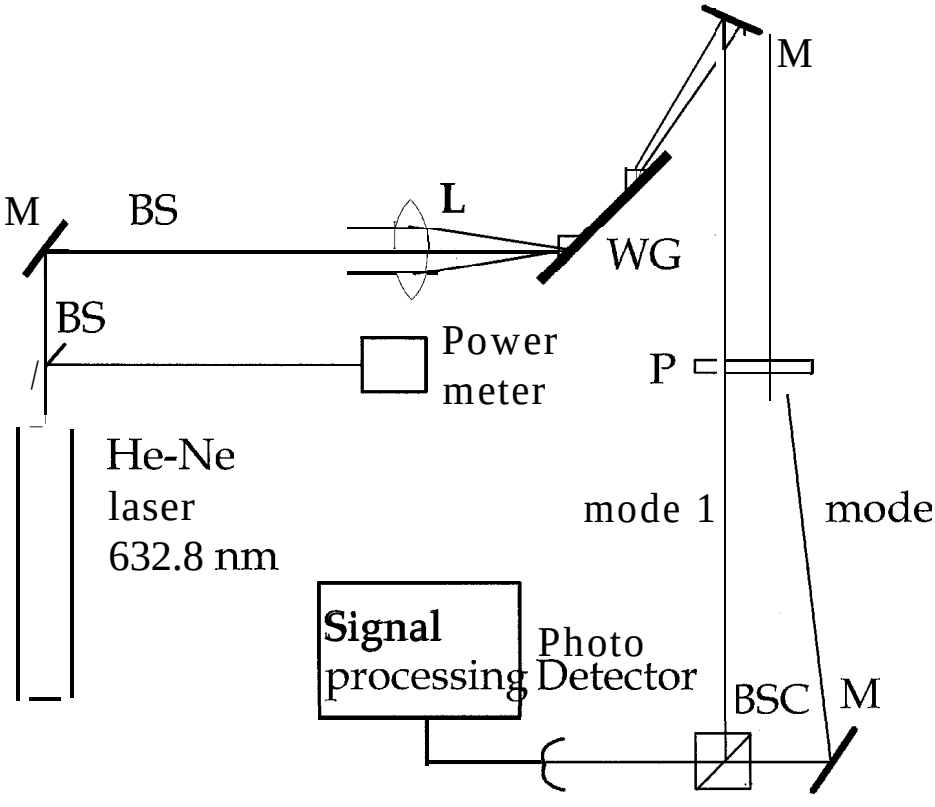


Fig. 1. Double-mode single-arm interferometer. M-mirror, L - lens, WG-waveguide, BS - beam splitter, P - polarizer, BSC - beam splitting cube.

Material selection

The actual working prototype of the waveguide sensor will use thin film of high temperature polyimide doped with metal substituted phtahlocyanines as indicator dyes. We have already shown that polyimide Ultradel 9020 from Amoco can be potentially used for planar waveguides delineated by direct writing with medium power UV source [1 O]. Metal substituted phthalocyanines are thermally stable organic compounds with distinctive optical absorption spectrum giving them deep blue or green color [11]. Preliminary study of a number of metal substituted phtahlocyanines has shown their potential use as indicator dyes [12]. For example, 1,4,8,11,15, 18,22,25-Octadecyloxy copper phthalocyanine changes significantly its optical absorption spectrum in a reversible manner after being exposed to various acid and alkali solutions. When several drops of acetic acid were added (3% acid concentration) to initially

green solution of the phthalocyanine in ethanol, the color changed to brown. However, when several drops of this brown solution were added to acetone, the color changed back to green. The same behavior accompanied by the refractive index change could be expected for phthalocyanines being exposed to NH_3, NO, NO_2, CO, CO_2, or SO_2 in the presence of water vapor.

Experimental gas chamber

Experimental set-up for testing a single-arm dual-mode interferometric waveguide sensor is presented in Fig. 2. It includes gas chamber where the sensor is exposed to various gas mixtures. The gas supply unit includes a combination of gas cylinders and water bubbler connected through controlled pipe lines to a mixing chamber. The purpose of the mixing chamber is to prepare a mixture of different gases at a certain proportion before exposing the sensor. This approach allows to investigate time response of the sensor to a particular gas mixture. The main gas chamber is equipped with an internal heater which provides the opportunity to characterize the performance of the sensor at elevated temperatures (up to 250^0C). The light from the external laser source (green He-Ne laser) is fed into the chamber through specially designed vacuum sealed multimode optical fiber connector. The multimode optical fiber line is coupled to the waveguide using focusing optics and a prism coupler. The same approach is used to outcouple the light from two propagating modes into the output multimode optical fiber line. The interference pattern is created by the light beams from the modes in the plane of the input terminal of the optical fiber line. The output light is fed through the second vacuum sealed optical fiber connector to a photo detector. Signal measurement is done using a chopper an lock-in amplifier. The pressure during mixing and filling the gas chamber is monitored by a vacuum gauge connected to the chamber.

Results and discussion

In our preliminary experiments we used the optical set up shown in Fig. 1. The double mode waveguide was fabricated by spin coating of poly(methyl methacrylate) (PMMA) film on preoxidized silicon wafer (silicone dioxide thickness was 1.5 micrometer, refractive index was 1.46). PMMA film refractive index was 1.5 at 633 nm wavelength. No substantial film birefringence was observed. The interference of TM_0 and TM_1 modes was used to produce sensing effect. The reaction of the interferometer to ambient temperature change is illustrated in Fig.3. The temperature change was induced by an external heater placed at a distance of 10 mm from the surface of the waveguide. The top scaling of the graph shows the heating cycle time. The heating rate was approximately 2^0/min. The interferometer sensitivity can be estimated as an average temperature increment causing phase shift 2π between interfering optical waves. The phase shift 2π corresponds to the distance between adjacent peaks or valleys of the sensor signal (Fig.3). Within this approach, estimated sensitivity is about 2^0C.

We also performed preliminary experiments on gas sensing with a planar waveguide made of dye-doped polymer (particularly, PMMA doped with laser dye DCM) thin film. This experiments showed that the waveguide being exposed to atmospheric gases experiences the increase of its refractive index. After annealing the refractive index returns to its previous level. Annealing of the waveguide was done with UV radiation from a 500-W mercury arc lamp efficiently absorbed by the dye. The refractive index of the waveguide was measured with a prism coupler. For example, keeping the waveguide in open air for 15 hours increased its refractive index by approximately 0.001 while the heat generated by the UV radiation (even at doses less than 5 J/cm^2) removed the adsorbed gases from the waveguide and returned the index

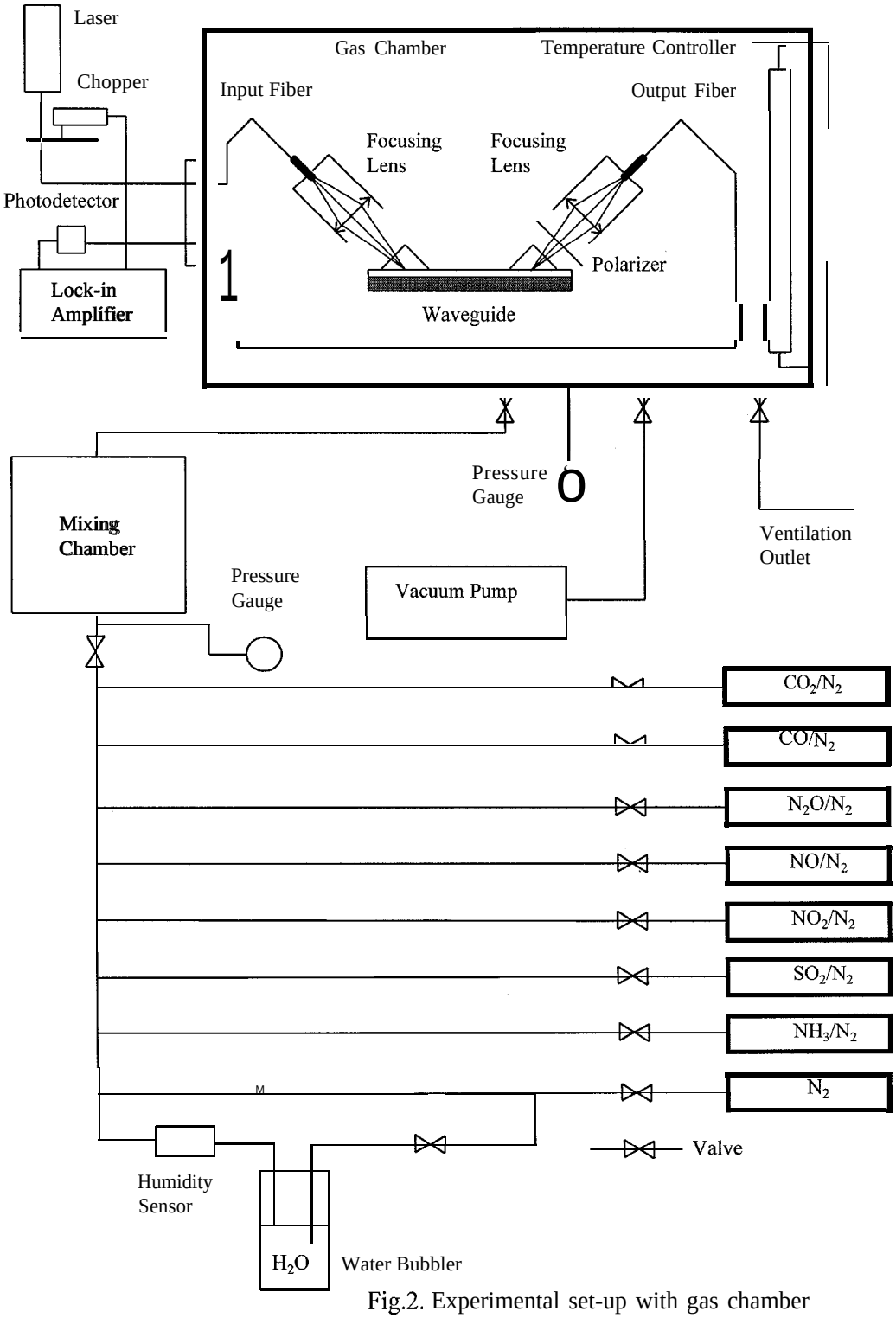


Fig.2. Experimental set-up with gas chamber

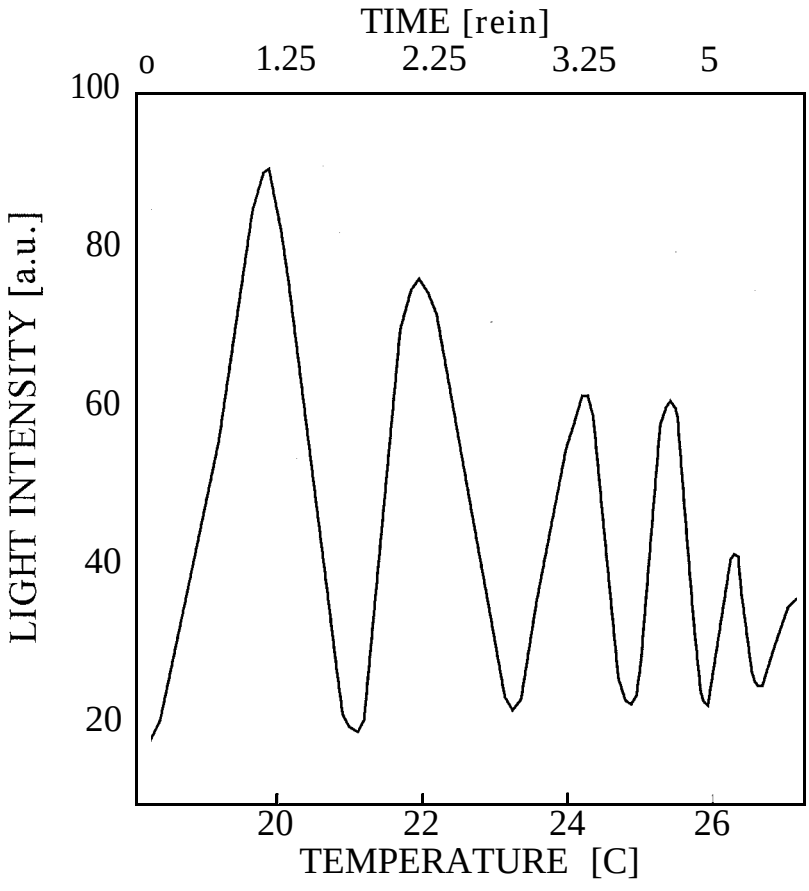


Fig. 3. Dual-mode interferometer response to the change of environment temperature.

to the previous level. This preliminary experiment shows that reading the refractive index change in the waveguide using single-arm dual-mode interferometry could give the information about the presence of various gaseous pollutants in atmosphere.

Conclusions

We have proposed and successfully tested the single-arm dual-mode thin film waveguide interferometric sensor. The sensor demonstrated 2π-phase shift per 2^0C-change of the ambient temperature. The experimental gas chamber has been designed to expose the sensor based on high temperature polyimide film doped with metal substituted phthalocyanine to NOX gaseous pollutants. During preliminary experiments dye-doped polymer film demonstrates substantial change of its refractive index after being exposed to atmospheric gases. The proposed interferometer is able to detect with high sensitivity the change of the refractive index and can be used as an efficient environmental sensor for monitoring NOX pollutants.

References

1. K.E. Rochford, R. Zanoni, G.I. Stegeman, W. Krug, E. Miao, M.W. Beranek, Pulse-modulated interferometer for measuring intensity-induced phase shifts, IEEE J. Quant. Electronics, Vol. QE-28, No. 10 (1992) pp. 2044-2050.

2. P. Hlubina, P. Prochazka, Sensor application of two-mode fiber in the Michelson interferometer configuration, International Conference "Interferometry' 94: Interferometry Fiber Sensing", Warsaw, Poland, 1994. Proceedings of International Society for Optical Engineering (SPIE), Vol. 2341 (1994) pp. 202-211.

3. L.M. Lechuda, A.T.M. Lenferink, R.P.H. Kooyman, J. Greve, Feasibility of evanescent wave interferometer immmunosensors for pesticide detection: Chemical aspects, Proceedings of the 5-th International Meeting on Chemical Sensors, Part.2, Rome, Italy, 1995, Sensors and Actuators,B; Chemical, Vol. B25, No. 1-3, Part.2 (1995) pp. 762-765.

4. R.M. Jenkins, R. W.J. Devereux, J.M. Heaton, Novel waveguide Mach-Zehnder interferometer based on multimode interference phenomena, Optics Commun., Vol. 110, No. 3-4 (1994) pp. 410424.

5. P.Plizka, W. Lukasz, Integrated-optical acoustical sensor, Proceedings of the Conference on Eurosensors VII, Sensors and Actuators, A: Physical, Vol. 41, No. 1-3 (1994) pp. 93-97.

6. Ch. Fattinger, H. Keller, D. Schlatter, P. Wehrli, Difference interferometer: A highly sensitive optical probe for quantification of molecular surface concentration, Biosensors and Bioelectronics, Vol. 8, No. 2 (1993) pp. 99-107.

7. W. Ecke, W. Haubenreisser, H. Lehmann, S. Schroeter, G. Schwotzer, R. Willsch, Phase-sensitive fibre-optic monoptodes for chemical sensing, Proceedings of the 1 -st European Conference on Optical Chemical Sensors and Biosensors - EUROPTODE 1, Sensors and Actuators, B: Chemical, Vol. B11, No. 1-3 {1993) pp. 475-479.

8. A.A. Bolarski, J.R. Busch, R.S. Brody, W.P. Altman, Integrated optic sensor for measuring aflatoxin-B 1 in corn, SPIE Symposium "Photonics West' 96", 27 January-2 February 1996, San Jose, CA, USA, will be published in Proceedings of SPIE, Vol. 2686 (1996).

9. A. Sharma and R. Posey, Dual-mode fiber-optic interferometer: a novel design, Opt. Lett., Vol. 20, No.22 (1995) pp. 2345-2347.

10. S. Sarkisov, Z. Teague, P. Venkateswarlu, H. Abdeldayem, D. Frazier, and G. Adamovsky, Formation of graded-index waveguide in UV exposed polyimide, J. Appl. Phys., Vol. 81, No. 6 (1997), p. 2889-2891.

11. Phtahlocyanines: Properties and Applications, Editors: C.C. Leznoff and A.B.P. Lever, VCH Publishers, Inc., New York, 1989.

12. Huaisong Guo, MS Thesis, New Mexico Highlands University, Las Vegas, New Mexico, 1997.

98 URC048

Effects of Personality and Gender on a Simulated Flight Task

D. Strickland, H. Johnson-Dunnings, S. Wright, C. Ntuen and M. Portalatin.
North Carolina A & T State University, Greensboro, North Carolina

Traditionally, studies in the area of human factors and supervisory control have focused on the perception and cognitive aspects of the controller in the system. This research was conducted to determine possible effects of personality and gender on an individual's ability to respond to the system, as well as the individual's perception of workload. For this study, a flight simulation prototype with five separate gauges was created to act as the system. Potential participants between the ages of 17 and 30 were given questionnaires to determine whether they were one of the following personality types: Introvert/Extroverts or TypeA/Type B. Based on their responses to the questionnaires, five individuals of each personality type were chosen to participate in the testing (1 O males and 10 females). They were exposed to two 30 minute simulated flight sessions, during which time they monitored the gauges for automation failure. Response times to failures were collected and analyzed using analysis of variance. Each participant also filled out workload surveys to determine his/her perceived workload for each session. Statistical analysis indicated significant differences (at an alpha level of 0.05) in response times and workload measures based on gender and personality, as well as the interaction of the two variables. Further research using more subjects is recommended to validate results obtained. Validation may lead to changes in pilot training programs.

INTRODUCTION

New technology always comes at a price. The more a technology advances, the more problems arise that need to be studied and corrected. This is the case in the aviation industry, as can be seen in the increasing number of airplane accidents that are tied to pilot error.

An example of a major technological advance that has been made in aircrafts in recent years is the automatization of most cockpit tasks. This advance has reduced the need for pilot interaction with the plane controls. As a result of this, the pilot has become a supervisor of the system as opposed to being the initiator of all tasks. This is often referred to as the pilot being "out-of-the-loop".

In the cockpit, the pilot is exposed to both high and low levels of workload. This can be attributed to automation and numerous tasks involving multi-functional displays. By analyzing the factors involved in operating a supervisory control system, the operator's workload can be examined.

This research attempted to determine the results of personality and gender on performance of simulated flight tasks and perceived workload. In this experiment, 20 subjects were tested in two simulation sessions, with their perceptions of various components of workload recorded after each session.

METHOD

Participants. Several volunteers from North Carolina A&T State University agreed to take the Eysenck Personality Questionnaire (EPQ), Form T of the Jenkins Activity Survey (JAS) (Krantz, Glass, and Synder, 1974), and a short user profile survey. The EPQ was used to determine whether a subject was a strong introvert or extrovert or neither. The JAS was used to determine whether the subject was a strong Type A personality, a strong Type B personality, or neither. After the personality questionnaires were scored, five individuals for each personality type (introvert/extrovert/type a/type b) were chosen to participate in the experiment. All subjects had normal or corrected-to-normal vision.

Apparatus. A flight simulation prototype was developed to investigate an operator's performance and workload in a supervisory control system. It was developed using

Rapid Prototyping and Design 3.5 software. This program required a desktop computer with color graphics and VGH color monitor and a mouse.

The flight simulation prototype resembled the Multi-Attribute Task Battery Test (MAT), designed by Comstock and Arnegard (1992). The flight simulation prototype consisted of three tasks: engine temperature gauges, speedometer, and attitude indicator (Figure 1). The simulation of these three tasks represents only part of the functionality found in an aircraft cockpit.

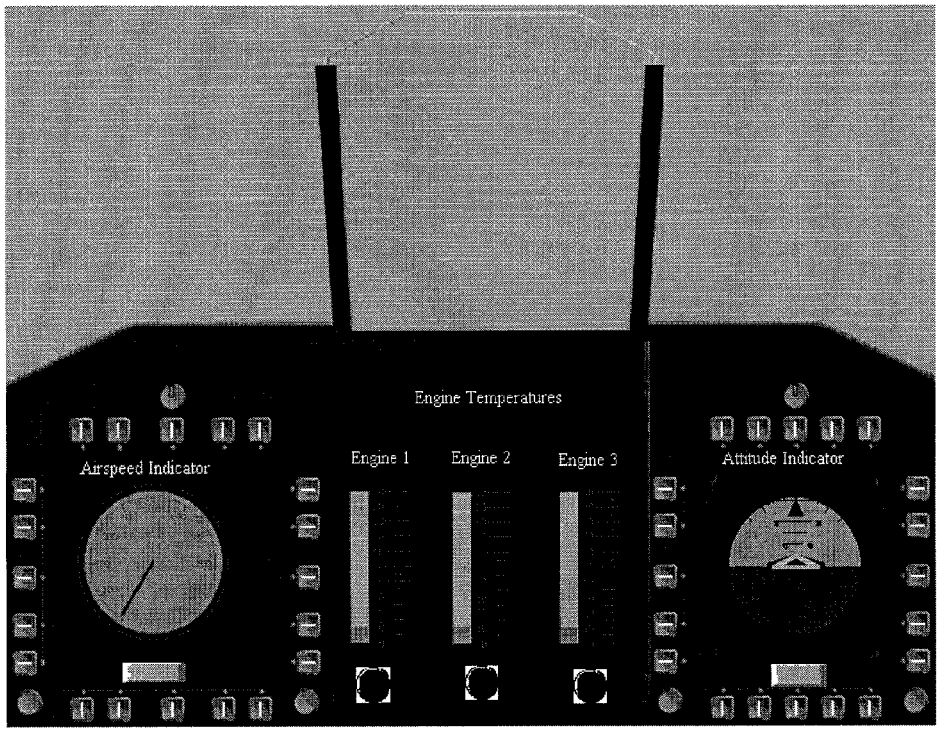


Figure 1: Flight Simulation Protytpe Design with no warning signals.

A workload rating scale was also used to determine participant perceived workload. The components of workload to be rated were mental demand, physical demand, temporal demand, performance, effort, and frustration.

Procedure. Participants were tested individually after filling out a consent form. Each participant had two simulation sessions of approximately 30 minutes each. For each session, participants were required to monitor all the components of the display for automation failure. Automation failure occurred when the indicator on the monitor went above or below a given range. All tasks failed at variable rates. The participant was required to click on the button below the monitor to indicate failure. The monitor then reset itself in the acceptable range. After completing each session, the participant then rated his or her perceived workload and performance during the session.

RESULTS

Analysis of variance was performed for the following measures: engine 1, engine 2, engine 3, speedometer, attitude, mental demand, physical demand, temporal demand, performance, effort, and frustration. Gender was examined with respect to personality type. Significant differences due to personality and gender were found for a number of the measures using an alpha level of 0.05. Statistically significant results are shown in Table 1.

TABLE 1

Significant Results of Statistical Analyses

Independent Variable	*Measure*	*F value*	*Prob>F*
Type A/Type B personalities			
Gender	Engine 2	4.3282	0.0395
Gender X Personality	Speedometer	12.1943	0.0006
Personality	Mental Demand	4.6276	0.0471
Personality	Frustration	5.5658	0.0314
Introvert/Extrovert personalities			
Gender X Personality	Engine 2	5.1922	0.0244
Gender	Engine 3	4.1588	0.0439
Gender X Personality	Speedometer	9.1627	0.0028
Gender	Attitude	5.0384	0.0256
Gender X Personality	Attitude	5.7775	0.0169
Personality	Frustration	9.2378	0.0078

Type A/Type B personalities. For engine 2, males responded to the failures significantly faster than women, as can be seen in Figure 2. This is a somewhat expected result since previous research has found men to have faster reaction times than women.

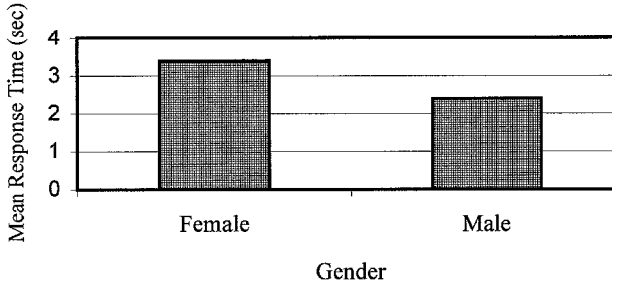


Figure 2. Mean Response Times for Engine 2 by Gender.

The Speedometer measure resulted in significant differences for the interaction of gender and personality. As can be seen in Figure 3, females had the shortest mean response time, while Type A males had the largest.

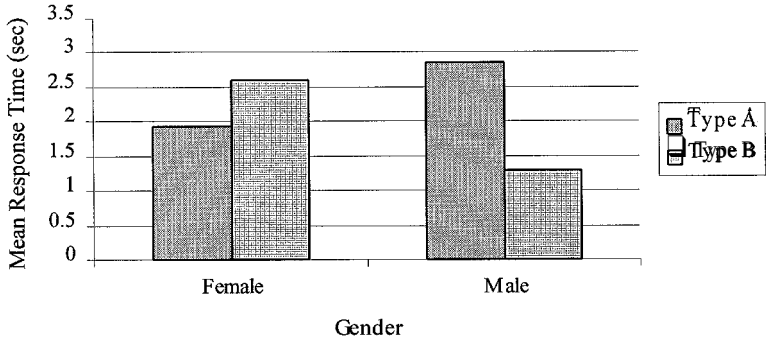


Figure 5. Mean Response Times for Speedometer by Gender and Personality,

Significant differences in perceived mental demand and frustration were seen due to personality type (Type A/Type B) (see Figures 4 and 5). Type A personalities perceived greater mental demand and frustration, which is to be expected given characteristics associated with these personalities (such as impatience and competitiveness).

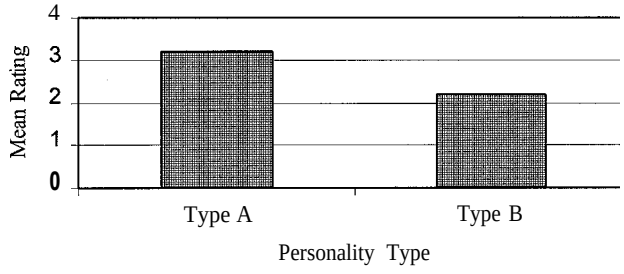


Figure 4. Mean Ratings for Mental Demand by Personality Type.

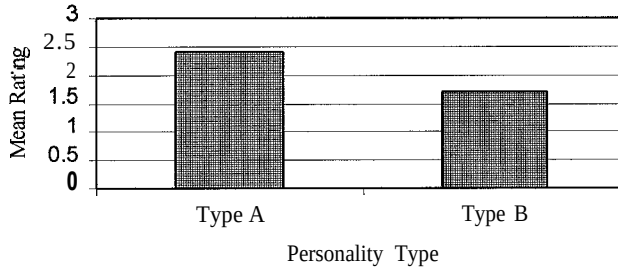


Figure 5. Mean Ratings for Frustration by Personality Type.

Introvert/Extrovert personalities. For engine 2, extroverted females had the shortest response times and extroverted males had the longest (see Figure 6). This same trend can be seen in Figure 7 for the speedometer responses.

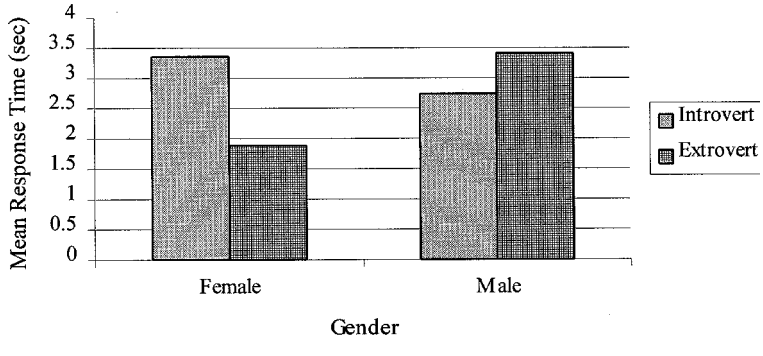


Figure 6. Mean Response Times for Engine 2 by Gender and Personality Type.

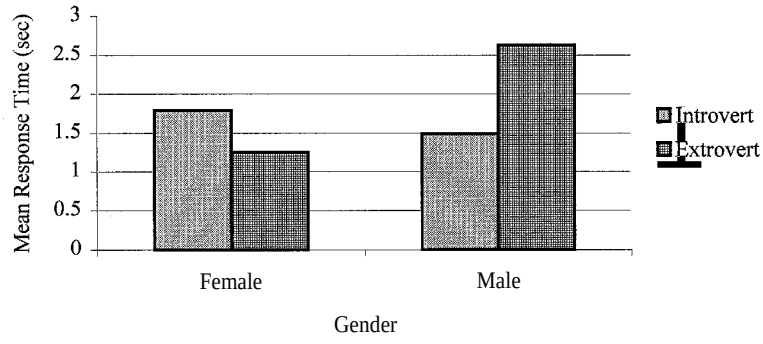


Figure 7. Mean Response Times for Speedometer by Gender and Personality.

Significant differences were seen between females and males in response times to engine 3 and the attitude indicator (see Figures 8 and 9). This also is in contrast to expected results (males are usually expected to be quicker than females).

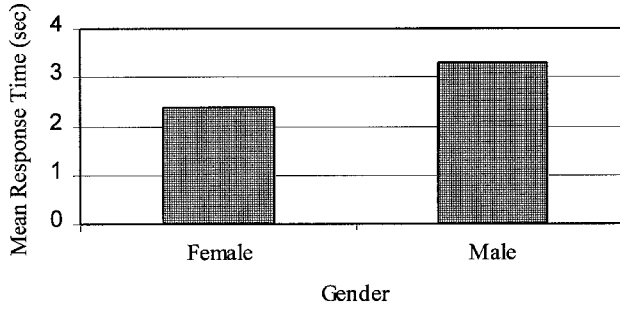


Figure 8. Mean Response Times for Engine 3 by Gender.

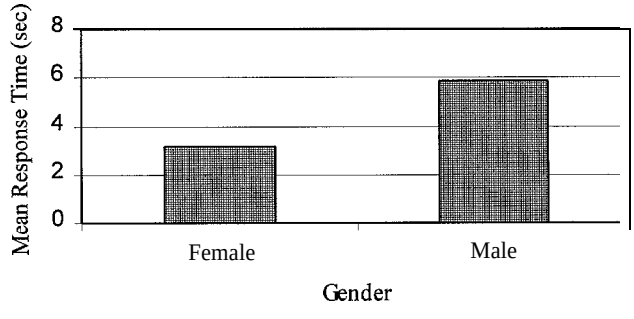


Figure 9. Mean Response Times for Attitude by Gender.

As can be seen in Figure 10, the response times for attitude also showed an interaction effect for gender and personality type. Extroverted females had the shortest response times while extroverted males had the longest. This data follows the trend seen earlier for response times to engine 2 and the speedometer.

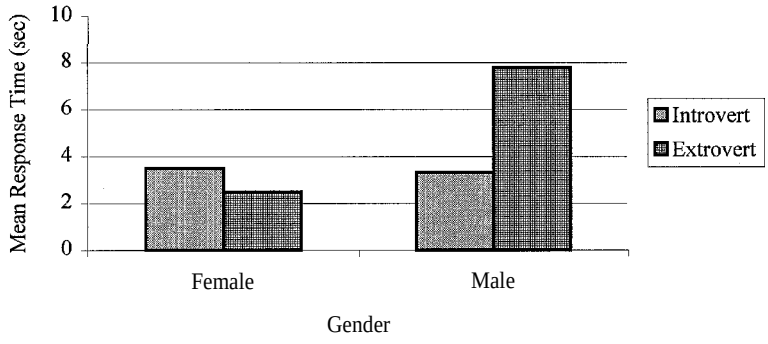


Figure 10. Mean Response Times for Attitude by Gender and Personality.

The only workload component which showed a significant difference was frustration. Introverts had significantly higher ratings than extroverts (see Figure 11).

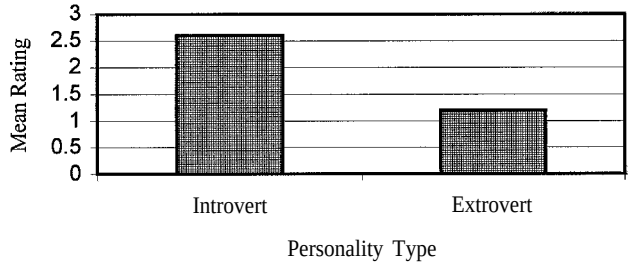


Figure 11. Mean Rating for Frustration by Personality Type.

CONCLUSION

The results discussed above suggest gender and personality should be taken into account for assignment to monitoring tasks, which includes. many piloting tasks. There were a total often significant differences found in this research due to gender, personality or the interaction of gender and personality for the eleven different measures.

There is a great deal of interest at this time in research on personality traits and pilots. Further research should be conducted with a greater number of subjects to validate the findings of this study, as well as to explore other personality traits. Research should also be conducted using actual pilots. Validation of this data may be used to enhance pilot training programs, as well as in pilot employment selection processes.

REFERENCES

Comstock, and Arnegard (1992). *Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Patterns.* Hampton, VA: NASA Langley Research Center.

Frantz, Glass, and Synder (1974). Helplessness, stress level, and the coronary-prone behavior pattern. *Journal of Experimental Social Psychology, 65 (2), 399-409.*

98 URC049

ASSESSING THE IMPORTANCE OF THE EVAPORATION-WIND FEEDBACK MECHANISM IN THE MODULATION OF SIMULATED MADDEN-JULIAN OSCILLATIONS

EDWARD COLON *AND JAMES LINDESAY
The Center for the Study of Terrestrial and Extraterrestrial
Atmospheres, Howard University, Washington D.C. 20059
MAX J. SUAREZ
Laboratory for Atmospheres
NASA Goddard Space Flight Center, Greenbelt, Maryland 20771

Abstract

An examination of simulated Madden-Julian Oscillation (MJO) response to active and suppressed air-sea interactions is made using an aquaplanet model employing a realistic representation of the hydrologic cyle. In general, the evaporation-wind feedback (EWF) results from a coupling between tropical zonal surface wind stresses and evaporation anomalies. Recent observational and theoretical studies have questioned the significance of EWF in sustaining the predominantly wavenumber 1 eastward propagating mode commonly attributed to the interaction between large scale convergence and cumulus-scale convection (conditional instability of the second kind, CISK). To ascertain the nature of the EWF dependance on lower boundary conditions and thus quantify its effect on MJO development, a series of numerical experiments were conducted employing various zonally symmetric sea surface temperature (SST) distributions with active and suppressed EWF mechanisms. Results suggest that a correlation exists between tropical SSTs and the efficacy of the EWF in vertically redistributing heat acquired through surface wind stresses. It has been determined that the removal of the EWF is not a crucial factor in the dampening of the simulated MJO at high equatorial S STs. The additional energy fed into the developing convective mode by the E WF selectively amplifies higher order wave modes in all numerical experiments thus boosting overall variances in oscillatory responses.

1 Introduction

Of the several types fluctuations observed in the tropical atmosphere, the 30-60 day or Madden-Julian oscillation (MJO) has remained a topic of intense interest since its initial detection over 25 years ago. [1, 2]. The manner in which it originates, the reason for its preferred intraseasonal timescale and its regional and global manifestations remain issues to be resolved. Its effects have been directly linked to the onset and interruption of the Asian monsoon [3, 4, 5] to the atmospheric angualr momentum (AAM) cycle. The MJO contributes to nearly half of the (AAM) variability observed at intraseasonal timescales through couplings of circulation responses resulting from latent heat released by the convective mode associated with this oscillation to zonally varying ambient flows and mountain torques [6]. Although not as well substantiated, a correlation has also been proposed between the MJO and the onset of El Niño/Southern Oscillation (ENSO) [4, 7]. It has been recognized that the MJO is primarily a wave number one phenomenon possessing a phase speed of 5-15 m s^{-1} and a meridional extension of 10° north and south of the equator. The oscillatory signal is most pronounced in barometric pressure, tropospheric temperature and zonal wind fields using earth-b ased observing stations and outgoing longwave radiation (OLR) fluctuations which often serve as a proxy for precipitation using satellite-breed remote sensing [8, 9, 10].

The overall structure and behavior of this tropical phenomenon suggests that the MJO may be interpreted in terms of equatorially trapped Kelvin waves. However, real difficulties are encountered when we try to rectify discrepancies in wave phase speeds. From the wave equation derived from inviscid, shallow water conditions, we find that for the nondispersive dry Kelvin wave, the phase speed $c = \sqrt{gh_o} \simeq 60\ \mathrm{ms}^{-1}$ where g is gravitational acceleration and h_o is the equivalent depth of the vertical mode. This rate of propagation is four times faster than the highest phase speeds observed for the MJO and has spurred the development of several schemes which incorporate moisture and boundar y layer friction in order to mediate the rapid progression of these equatorially trapped waves. The evidence suggests that tropical moist convection may provide the impetus for the creation of instability leading to wave generation

*PhD candidate, corresponding author, em ail: ecolon@branson. huphysics.howard.edu

and growth through a coupled interaction between large-scale circulation and cumulus-scale convection. The release of latent heat within the convecting cells drives mid-level ascent in the atmosphere and thus serves to intensify the initial low-level moisture influx. This process is dubbed conditional instability of the second kind (CISK) [12] and in its original form, relies on boundary layer friction to supply the low-level forcing necessary for cumulus-scale moisture convergence.

Current applications of CISK theory in MJO studies utilize equatorially trapped Kelvin waves in lieu of boundary layer processes (wave-CISK) [13]. Most wave-CISK type parameterizations tend to hinder the progression of equatorially trapped Kelvin waves in the troposphere. Current general circulation models employing appropriate convective schemes are capable of reproducing some kind of MJO but they tend to overestimate phase speeds. The only way to obtain realistic periodicities in these models is by lowering the level of maximum cumulus heating to the lower troposphere which conflicts with observational profiles [14]. Throughout this discussion, surface heat fluxes from the tropical oceans have not been considered although they may in fact play a vital role by providing additional energy to the MJO by vertically redistributing heat acquired from the sea surface thoughout the troposphere. In section 2, we will briefly examine air-sea interaction theory and in section 3, we will discuss the series of numerical experiments designed to asertain the significance of air-sea interactions. A discussion and summary will be presented in section 4.

2 Air-Sea Interactions

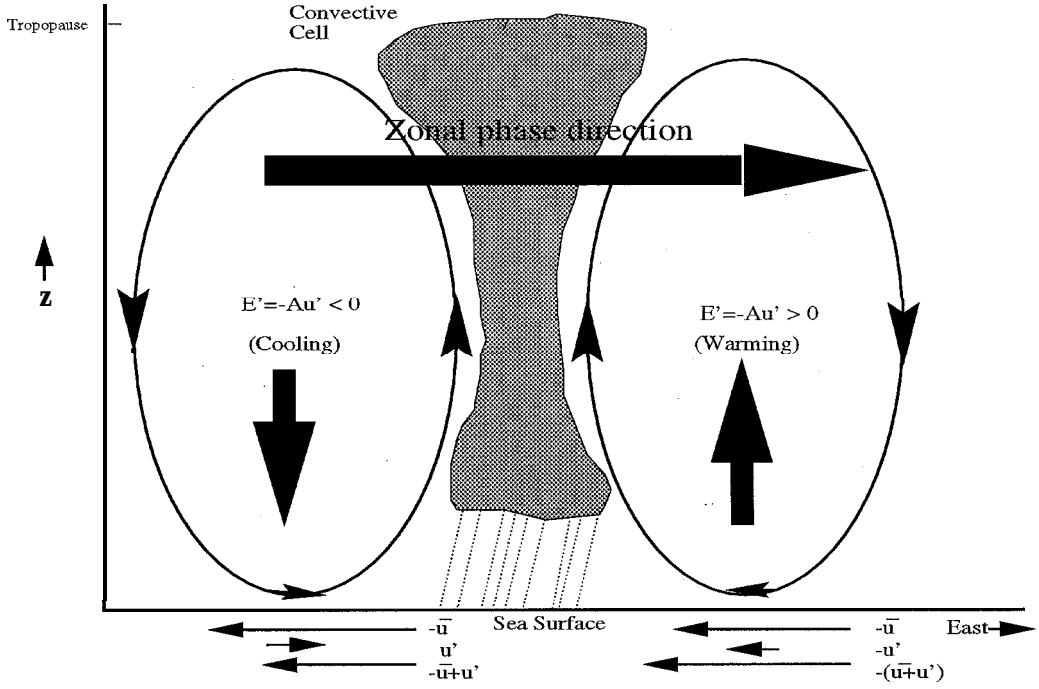


Figure 1. Conceptual diagram illustrating the forcings responsible for the EWF processes. Zonal wind perturbation (u') due to convective flow tend to strengthen the mean easterly circulation to the east of a convective cell and diminish circulation to the west leading to enhanced evaporation upwind followed by wave growth and westerly mode propagation

Of particular interest in this study is a class of air-sea interaction dubbed the evaporation-wind feedback (EWF) mechanism. This process spurs the development of evaporation and latent heat flux anomalies as a result of perturbations in mean surface easterly flow in the tropics induced by convergent mass fluxes associated with deep cumulus convection. Figure 1. graphically illustrates the manner in which strong easterlies give rise to enhanced evaporation to the right of the low-level convergence regime and suppressed evaporation to the left. Since the temperature anomaly associated with latent heat release through condensation is in phase with the vertical mass flux associated with the convective cell, growing baroclinic instability results and the subsequent EWF wave begins to propagate eastward while undergoing substantial growth. Previous numerical and theoretical studies have examined the manifest ation of the EWF modes in the absense of CISK and have generated oscillatory structure and phase speeds compatible

with observed MJO data [15, 16, 17]. All studies concede that the essential criterion governing EWF wave growth is directly linked to surface evaporation (E) so that

$$E = \rho C_D |\vec{V_h}|(\epsilon_s - \epsilon) + E' \quad (1)$$

where ρ is the lower troposphere density, C_D is the drag coefficient, $|\vec{V_h}|$ is the magnitude of the surface wind velocity vector, ϵ_s is the saturation specific humidity at the surface, and E' =-Au' where A represents the E WF coefficient and u' is the zonal wind perturbation brought about by large-scale convective flow. EWF mode behavior is determined by the following conditions: when (1) A >0 wave growth and propagation results from mean zonal easterly surface winds and when (2) A < 0 mean zonal westerly surface winds lead to coupled mode damping [16] Observational studies by Wang [1 S] and Lin and Johnson [19], have brought the relevance of the EWF into question in the maintenance of the MJO. The essential difficulty relates to observed westerlies during episodes of MJO development and propagation which conflicts with the basic feature of EWF processes. Coupled Ocean Atmosphere Response Experiment (C OARE) data reveals that the regions to the east of developing MJO episodes where characterized by weaker zonal winds, higher SSTs, and anomalously low evaporation which contradicts the expected EWF response [19]. This remains an area of contention which we have probed with the use of a simple two-level general circulation model.

3 Experiment

A series of numerical ten year simulations were performed using an aquaplane model employing a simple cumulus convective scheme and a full hydrologic cycle [20, 21]. Several simplifications were made including the removal of land-sea contrasts and seasonality while fixed cloud distributions were employed. The purpose of these modifications was to insure a steady state model environment which facilities the examination of simulated MJ O characteristics. Zonally symmetric SST distributions were employed to ascertain the EWF boundary condition dependencies. Profiles varied from a realistic 28°C maximum equatorial SST to a very low 1 8° C . The simulated MJO signal analysis involved comparing EOF (empirical orthogonal function) variability obtained from χ_{250} (250 mb velocity potential) fields. MJO structure was studied using composite maps derived from χ_{250} principal component data for the first and second (PC 1, PC2) eigenvector which make up the wavenumber 1 oscillatory response. Using a one standard deviation cutoff criterion, composite dates were extracted corresponding to minimum and maximum PC1 and PC2 time series signals. The extracted dates then served as a basis for developing grand composite maps of several fields of interest averaged with ± 1 day lag/lead time indices [10]. Frequency spectra were obtained using multitaper and hamonic tests to determine the definitive periodicities in the model data [22].

4 Discussion and Summary

Figure 2 illustrates this relationship between simulated MJO signal strength and maximum equatorial SST for the first 3 wavenumbers (EOFs 1-6) which account for most of the signal variance and the first 25 EOFS. A logarithmic scale was utilized to facilitate direct comparisons. The increase in equatorial SSTs is direct] y correlated to saturation specific humidity for the coupled mode and it is also evident that the growth rate of the wavenumber 1 is favored over lower order oscillations for SSTs below 23° C. Scale selection then appears to favor lower order modes which conforms with tropical energetic principles requiring that short waves develop and intensify as a result of barotropic instability y brought on by zonal mean flow and wave-wave (nonlinear cascade) interactions between long and short waves. The coupled versus the CISK-only experiments also reveal that the EWF plays an increasingly significant role in maintaining the simulated MJO at the low equatorial SSTs. An observed steady dominance of the CISK mode in relation to the EWF is apparent with increasing equatorial SSTs and is in accord with the results of previous studies [18, 23].

The horizontal structure of the simulated MJO is represented in Figure 3 and was obtained using the compositing procedure described in section 3. A side by side comparison is made between the coupled and CISK-only model runs in order to denote structural disparities for the simulations employing an 18° C equatorial SST distribution. Wind vector plots reveal a quasi-zonal response for the coupled run while significant subtropical eddies were manifest in the CISK-only simulation. The lack of meridional structure in the coupled response illustrates that frictional moisture convergence and the associated heating are not sufficent to allow conspicuous generation of westward propagating Rossby wave signatures. The CISK-only wind vector plot does show considerable subtropical cyclonic flow but this is due in part to the suppression of wavenumber 1 variability and the amplification of higher frequency modes with associated noise. The superimposed total precipitation anomaly represented by thick contours reveal equatorially symmetric structure which correlates with vertical velocity and upper and lower-level divergence fields (not shown).

The plots representing evaporation related to surface wind stresses reveal the very nature of the EWF mechanism. As shown in equation 1, the surface latent heat flux is dependent on time-mean evaporation plus a time-varying perturbation. By taking the zonal and time average of the surface wind stresses, we can remove the transient latent heat flux eddies and are left with a fixed zonally symmetric distribution [16]. The cycle of simulated MJO production in the coupled experiment can be deduced by examining plots (b), (d), and (c) in sequence. A transient perturbation

in evaporation to the east of the propagating wave (see figure 1) vertically mixes moisture into the troposphere. The approaching convergent wave trough triggers condensation and the release of additional latent heat such that an upper level temperature anomaly results and the pre-existing baroclinic instability y intensifies. The resulting moisture then precipitates out at focus of low level convergence. This scenario is absent from the CISK-only experiment for the obvious reason that there is no mechanism for vertically mixing transient moisture fluxes.

Figure 4. depicts the harmonic power and amplitudes spectra obtained for the same experiments used in the composite analyses. To isolate 250 mb zonal wind peaks, multitaper metho ds are employed using a 95 % confidence level cutoff. A 31.8 day peak is observed in the coupled mode experiment which corresponds with an oscillatory phase speed of 14.5 ms^{-1}. This falls at the upper limit of observed oscillatory phase speeds. The harmonic peak test corroborates the dominant wavenumber 1 period. The CISK-only experiment experiment yields a triplet of spectral peaks at 43.1 days, 36.4 days, and 31.3 days. The wavenumber 1 periodicity present in the coupled simulation is approximated by the 31.3 day spike, however, the two additional harmonics appear to be associated with the dominant mode as well since the peak test confirms their robustness. The overall spectral power drops by nearly a factor of three between the coupled and CISK-only experiment which again illustrates the substantial difference in simulated MJO signal strength.

To conclude, the fact that the EWF was necessary for the maintenance of simulated MJOs for relatively cool equatorial SST profiles and that this dependence tended to decline as realistic equatorial SSTs were utilized may support some of the contentions expressed in studies questioning the overall importance of air-sea interactions in sustaining intraseasonal variability. Since as a whole, observed tropical SSTs tend to approach the upper limit simulated temperatures, it is likely that the growing CISK mode can remain dominant without the additional latent heat flux provided by the EWF. The EOF variances obtained from low temperature limits point to a dominating EWF mode which selectively amplifies the wavenumber 1 component while suppressing higher frequency waves. This implies that there may be inst antes when regions of anomalously cool equatorial SSTs could in principle sustain EWF dominant intraseasonal oscillations. Future work will examine observational data as well as employ coupled ocean -atmosphere general circulation models including land-sea contrasts to obtain a more realistic picture of MJO processes.

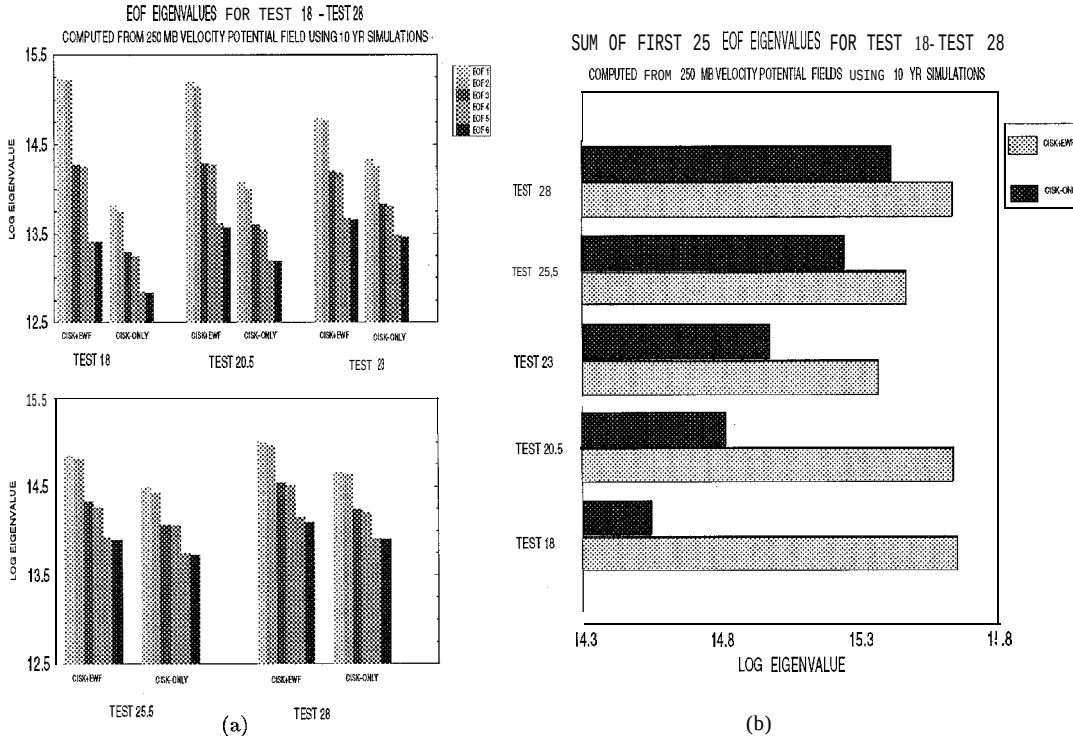
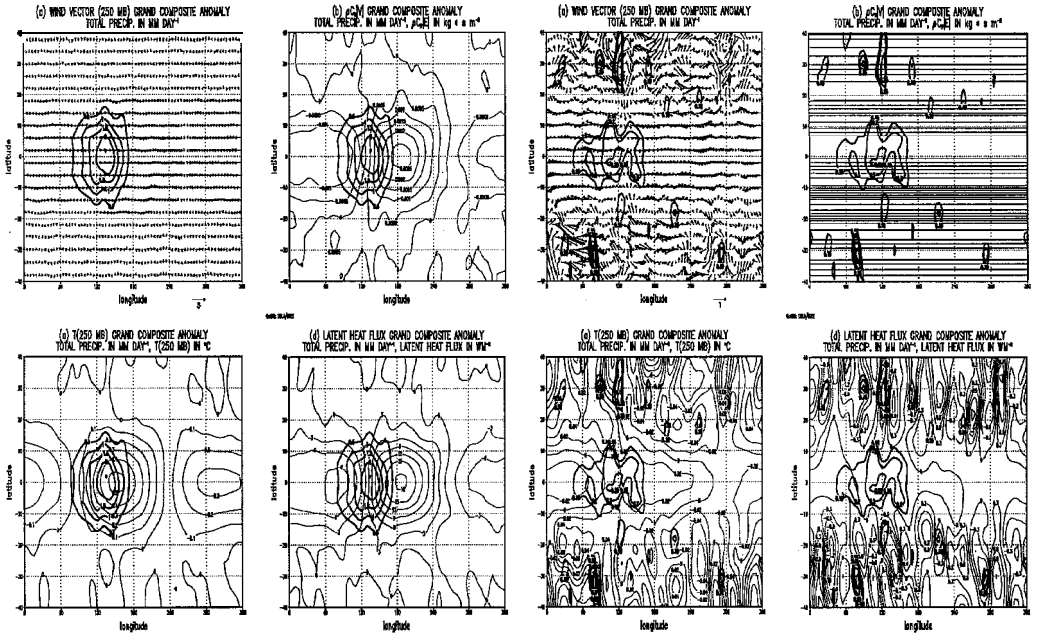
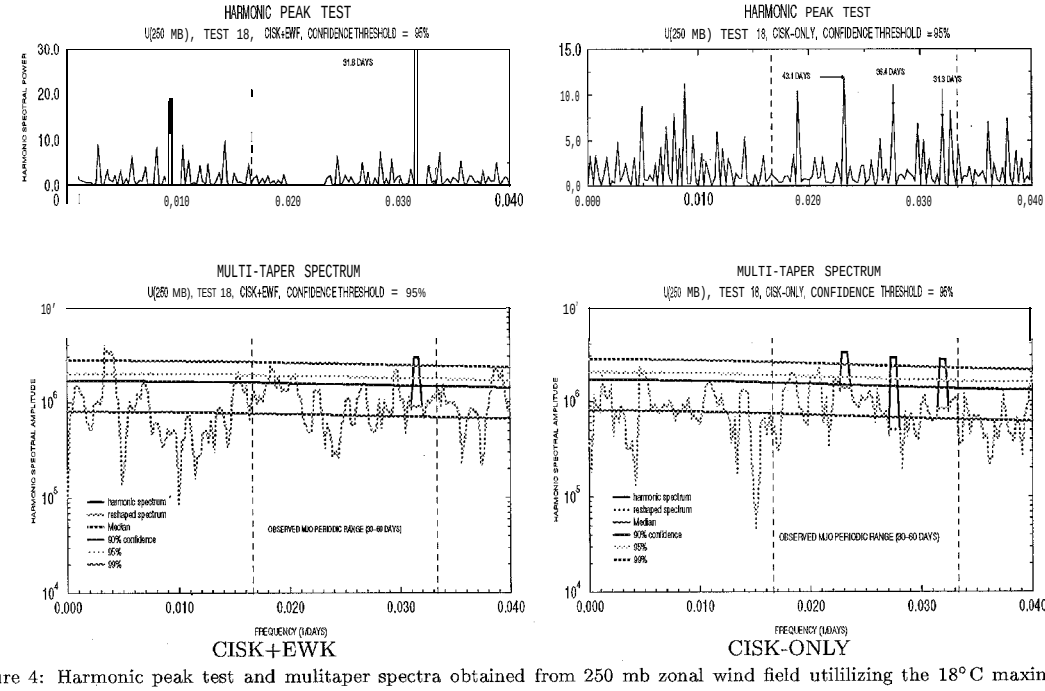


Figure 2. Log of EOF eigenvalues obtained from χ_{250} fields for (a) first three zonal wavenumbers (principal components 1-6) and (b) the sum of the first 25 principal components for 10 year simulations utilizing zonally symmetric SST distribution.



CISK+EWK CISK-ONLY

Figure 3. Composite plots obtained using 18° C maximum equatorial SST distribution for 10 year simulations with/without EWF mechanism in operation. A χ_{250} EOF temporal basis was utilized with 1 σ cutoff and the total precipitation anomaly was superimposed for (a) wind vector anomalies 250 mb, (c) surface wind stress anomalies, (c) temperature anomalies at 250 mb, and (d) evaporation anomalies at the surface.



CISK+EWK CISK-ONLY

Figure 4: Harmonic peak test and mulitaper spectra obtained from 250 mb zonal wind field utililizing the 18°C maximum equatorial SST distribution for 10 year simulations with/without the EWF mechanism in operation.

Acknowledgements—We would like to thank the NASA Goddard Space Flight Center Office of University Programs for support under the Graduate Student Researchers Progam Fellowship and the NASA Goddard Space Flight Center Laboratory for Atmospheres for technical assistance.

References

[1] Madden, R. A. and Julian, P. R., Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, 1971, 28, 702-708.

[2] Madden, R. A. and Julian, P. R., Description of global-scale circulation cells in the tropics with a 40-50 day period. *J. Atmos. Sci.*, 1972, 29, 1109-1123.

[3] Krishnamurti, T. N. and Subrahmanyam D., The 30-50 day mode at 850 mb during MONEX, *J. Atmos. Sci.*, 1982, 39, 2088-2095.

[4] Lau, K.-M., and Chan, P.- H., The 40-50 day oscillation and the El Niño-Southern Oscillation: a new perspective. *Bull. Amer. Meteor. Sot.*, 1986,67, 533-534.

[5] Chen, T.-C., and Weng, S.-P., Some effects os the intraseasonal oscillation on the equatorial waves over the Western Tropical Pacific-South China Sea region during the Northern summer. Mon. *Wea. Rev.*, 1995, 124, 751-756.

[6] Weickman, K. M., and Sardeshmukh, P. D., The atmospheric angular momentum cycle associated with a Madden-Julian oscillation, *J. Atmos. Sci*, 1994, 51, 3194-3208.

[7] Lau, N-C., and Lau, K-M., The structure and propagation of intraseasonal oscillations appearing in a GFDL general circualtion model, *J. Atmos. Sci.*, 1986, 43, 2023-2047.

[8] Weickmann, K. M., 1983: Intraseasonal circulation and outgoing longwave radiation modes during Northern Hemisphere winter. *Mon. Wea. Rev.*, 1983, 111, 1838-1858.

[9] Weickman, K. M., Lussky G. R., and Kutzbach J. E., Intraseasonal (30-60 day) fluctuations of outgoing longwave radiation and 250 mb streamfunction during northern winter. *Mon. Wea. Rev.*, 1985, 113, 941-961.

[10] Knutson, T. R., and Weickmann, K.M. , 30-60 day atmospheric oscillations: composite life cycles of convective and circulation anomalies. Mon. *Wea. Rev.*, 1987, 115, 1407-1436.

[11] Rui, H., and Wang B., Development characteristics and dynamic structure of tropical intraseasonal convection anomalies. *J. Atmos. Sci*, 1990, 47, 357-379.

[12] Charney, J.G. and Eliassen, A., On the growth of the hurricane depression. *J. Atmos. Sci.*, 1964, 21, 68-75.

[13] Lindzen, R. S., Wave-CISK in the tropics. *J. Atmos. Sci.*, 1974, 31, 156-179.

[14] Lau, N-C., Held, I. M., and Neelin, J. D., The Madden-Julian Oscillation in an Idealized General Circulation Model. *J. Atmos. Sci.*, 1988, 45, 3810-3832.

[15] Emmanuel, K. A., An air-sea interaction model of intraseasonal oscillations in the tropics. *J. Atmos. Sci.*, 1987, 44, 2324-2340.

[16] Neelin, J. D. , Held, I. M., and K. H. Cook, K. H., Evaporation-wind feedback and low frequency variability in the tropical atmosphere, *J. Atmos. Sci.*, 1987, 44, 2341-2348.

[17] Kirtmann, B., and Vernekar, A., On Wave-CISK and the Evaporation-Wind Feedback for the Madden-Julian Oscillation. *J. Atmos. Sci.*, 1993, 502811-2814.

[18] Wang, B. ,Comments on "An air-sea interaction model of the intraseasonal oscillation in the tropics." *J. Atmos. Sci.*, 1988, 45, 3521-3525.

[19] Lin, W., and Johnson, R. H., Heating, moistening, and rainfall over the Western Pacific warm pool, *J. Atmos. Sci.*, 1996, 53, 3367-3383.

[20] Held, I. M. and Suarez, M. J., A two level primitive equation model designed for climatic sensitivity experiments. *J. Atmos. Sci.*, 1978, 35, 206-229.

[21] Suarez, M. J., and Takacs, L. L., Documentation of the Aires/ GEOS Dynamical Core Version 2, Nasa Technical Memorandum 1040606, Goddard Space Flight Center, Greenbelt, MD 20771.

[22] Ghil, M., and Mo, K., Intraseasonal oscillations in the global atmosphere. Part I: Northern Hemisphere and Tropics. *J. Atmos. Sci., 1991,* 48, 752-779.

[23] Li, T., and Wang, B., The influence of sea surface temperature on the tropical intraseasonal oscillation. Mon. Wee,. *Rev.* , 1994, 122, 2349-2362.

Specifications for the Design of a Low Cost TDRSS-Compatible Portable Earth Terminal

Eliud Bonilla-González and Rafael Fernández-Sein
NASA-UPRM-GSFC Partnership for Space Telecommunications Education
Electrical and Computer Engineering Department
University of Puerto Rico, Mayagüez Campus
PO Box 5000, Mayagüez, Puerto Rico 00681-5000
Tel. (787)834-7620 Ext. 2263 or 2264- Fax (787)832-2485
ebonilla@sark.upr.clu.edu rafaelf@exodo.upr.clu.edu

INTRODUCTION

Antarctica provides an incredibly rich variety of unspoiled settings for the installation and operation of scientific instruments to study both our planet as well as the cosmos. The lack of spectral pollution, as well as unique atmospheric and environmental conditions, provide a great setting for earth-based experiments. The international scientific community has recognized Antarctica's potential and is increasing its presence. One of these organizations, the Center for Astronomical Research in Antarctica (CARA), under the sponsorship of the National Science Foundation has established the Automated Astronomical Site-Testing Observatory (AASTO) at the South Pole.

The AASTO is a self-contained observatory, currently located at the South Pole, and proposed to be moved to the high plateau of Antarctica. Its purpose is to collect atmospheric and environmental data for the determination of the optimum location for a permanent multinational astronomical observatory. The AASTO itself is derived from the successful US Automated Geophysical Observatory (AGO) program.

The same conditions that make Antarctica attractive for an observatory pose a significant challenge in the establishing of communication links to allow command and control of the instruments, as well as the gathering of data in a timely fashion. Currently, personnel must use low data rate links with some satellites or travel to the South Pole to retrieve the data for later analysis at their home institutions. A high data rate two-way communication link between the scientists and the instrumentation would greatly enhance the productivity and efficiency of these projects. This issue is being addressed for other South Pole experimenters at Goddard Space Flight Center (GSFC) by the South Pole TDRSS Relay Project.

SOUTH POLE TDRSS RELAY (SPTR)

The South Pole TDRSS Relay (SPTR) Project is a Proof-Of-Concept activity that will provide high data rate communications links to the South Pole via the TDRSS to allow Antarctica experimenters interactive Internet-type connectivity to and from their scientific experiments. It will provide fill duplex communications channel at a 1.024 Megabit/second data rate in both directions while allowing the experimenters to utilize standard TCP/IP protocol. Also to be provided will be a simplex communications channel from the South Pole to White Sands Complex (WSC) for data rates between I and 50 Megabits/second. Initial capability for file transfers will be constrained by the SPTR processors to 10 Megabit/second.

The highly inclined (9.5 degrees) geosynchronous TDRS-1 spacecraft, currently located at 49W longitude, will be used on a daily basis to open up a 2-2.5 hour window of communications with the South Pole when it is at its most southerly inclination each day. Its inclination will increase over the years, thus increasing the access time for the South Pole. It should be noted that TDRS-1 is the only TDRS capable of covering the South Pole but a large part of Antarctica is in view of at least one TDRS.

The AASTO will apply the lessons learned from SPTR. It will have a similar communications and operational architecture (see Figure 1). An earth terminal will be designed to provide the AASTO access to TDRS-1 and NASA's Space Network. It will be called the TDRSS AASTO Small Terminal (TASTE).

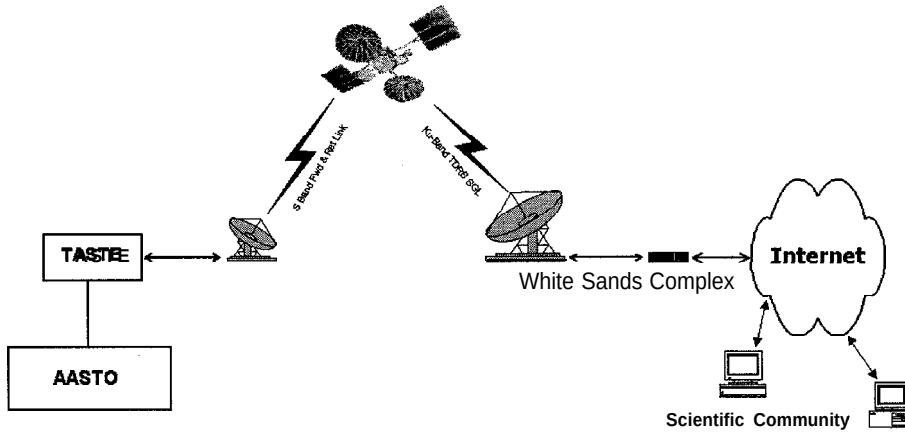


Figure 1 AASTO Communications Architecture Concept Diagram

TDRSS AASTO SMALL TERMINAL (TASTE)

The first step in the design of TASTE is to determine the technical, operational, and economic requirements, constraints and options of the AASTO project. This sounds much simpler than what it really is since the process is an iterative one; requiring input from potential customers, engineers, and managers spread out throughout all the world. At the moment of writing this report, input of various types had been received from: Antarctica, Australia, Canada, Belgium, and the United States. Most of the communications has been via email, thus making this a world-wide telecommuting project from the beginning.

AASTO REQUIREMENTS AND CONSTRAINTS

Communication Links

At present there is no communication link from the scientists to the AASTO. There is only a Housekeeping and Status Report that is periodically relayed via the ARGOS low-earth-orbit (LEO) satellite. This is a low bit rate, packet store-and-forward mode, method which does not provide much timeliness in data acquisition. A full-duplex link of only 300 baud would be a significant operational improvement over current capabilities and a maximum daily transfer of around 500 Kbytes is projected.

TASTE will provide data rates from 9.6 kbps up to 150 kbps. The lower limit, 9.6 kbps, should provide a reasonable rate for the transfer of 500 Kbytes. The upper limit, 150 kbps, was selected so as to have sufficient flexibility to comply with future TDRSS service plans.

TASTE will only use the TDRSS S-Band Multiple Access Mode (SMA) for both Forward and Return links. Multiple Access allows for easier scheduling of Space Network time and will provide compatibility with the Demand Access Service Concept to be implemented in the future.

Low Power Consumption

Electrical power is a very scarce resource. A propane thermoelectric generator provides all the power for heating the AASTO as well as for electrical power. A total of 50W, regulated to 28V, is available for all the instrumentation. The resulting power budget is very tight, allowing for about 5W to be available for the TASTE. Since current TDRSS-compatible transceivers can consume over 50W when transmitting, a group of batteries (6V, 12 AH) will

be supplied to configure a regulating/charging power subsystem. A 1/20 duty cycle seems to be adequate for reliable operation.

Reliable and Autonomous Operation

AASTO instruments have been designed to function in an autonomous mode and this standard must also be applied to TASTE. It should be extremely reliable and have the capability to work independently. Sufficient "intelligence" should be incorporated to allow continuous operation and recuperation from potential anomalies (discharged batteries, temporary loss of communications, etc.).

Low **Costs**

If TASTE is to become a viable alternative for AASTO operations it must be cost effective, both in the initial expenses as well as in daily operations. Current TDRSS-compatible transceivers are too expensive (over $ 100K) for AASTO and other scientific observatories. A judicious selection of specifications and requirements for TASTE should improve the probability of designing an acceptable product.

TASTE PRELIMINARY SPECIFICATIONS

Enough information has been gathered to permit the publication of a reasonable set of TASTE Specifications with a conceptual block diagram.

TASTE General Features and Specifications

Return Link Transmitter Characteristics

Operating Frequency	2287.5 MHz (Multiple Access)
RF Power Output	1.0/5.0 W (Computer controlled)
Antenna Gain	26.5 dB
Total EIRP	26.5 dBw/ 33.5 dBw
Antenna 3 dB Beamwidth	8.1 degrees
Carrier Modulation	BPSK
Selectable PN Code	TDRSS Compatible
Chip Rate	3.08 Mcps
Data Format	NRZ
Encoding Rate	1/2 Viterbi Encoding (Computer controlled)
Data Rates	9.6 kbps up to 150 kbps (300 ksps)
TCXO Stability	< .3 ppm
On/off	Computer controlled

Forward Link Receiver Characteristics	
Operating Frequency	2106.40625 MHz (Multiple Access)
Antenna Gain	26.5 dB
Antenna 3 dB Beamwidth	8.1 degrees
Carrier Demodulation	BPSK
Selectable PN Code	TDRSS Compatible
Chip Rate	3.08 Mcps
Date Format	NRZ
Decoding Rate	1/2 Viterbi Decoding (Computer controlled)
Data Rates	Up to 150 kbps (300 ksps)
Acquisition Threshold	>45 dB-Hz in RF
Noise Figure	< 2.5 dB
Implementation Loss	3 dB

TASTE General Features and Specifications (cont.)

S-Band Antenna Interface	
Diplexer to Antenna	50 Ohms, N-type
Personal Computer / LAN Interface	
Interface Type	RS-232 and Ethernet
Protocol	Asynchronous (max rate TBD) for RS-232 TCP/IP for Ethernet
Connector Type	DB-9 and Coax
Physical and Environmental	
Temperature	TBD
Supply Voltage	28 Vdc
DC Power Consumption	<5 watts for Receiver only <50 watts for Receiver and Transmitter
Size	TBD
Weight	TBD

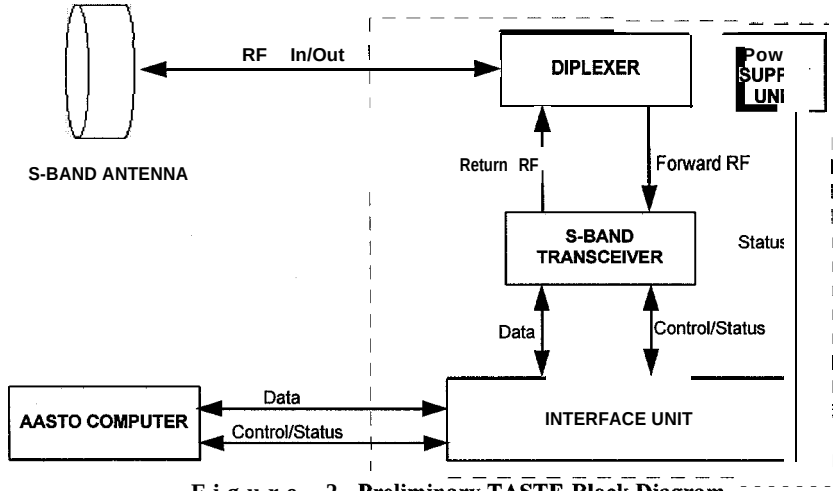


Figure 2 Preliminary TASTE Block Diagram

TASTE SUBSYSTEMS

TASTE will be comprised of the following subsystems: S-Band Antenna, Transceiver, Power Supply, and the Interface Unit. Although not part of the original configuration, TASTE will be installed inside the AASTO equipment shelter with exception of the S-Band antenna subsystem.

S-Band Antenna Subsystem

A covered 4-ft. diameter parabolic antenna has been selected for the initial TASTE configuration. This will provide adequate RF gain while providing a 3 dB pattern wide enough to function in a fixed position. Proper transmission lines will connect the antenna to the TASTE main unit inside the AASTO equipment shelter. All parts of this subsystem are available from commercial sources.

Transceiver Subsystem

This subsystem is composed of an S-Band TDRSS-compatible diplexer and transceiver unit. The diplexer provides a hi-directional interface to the RF signals between the antenna and the transceiver. Various commercial sources are available and lead times are relatively short.

The transceiver is the heart of the TASTE. It provides frequency up/down conversion, spread-spectrum modulation and demodulation, bit-synchronization, and Control/Status functions, among other features. It will be the most critical part of the design from a technological and economic point of view. Various commercial manufacturers have been invited to review and comment on the Preliminary TASTE Specifications so as to evaluate potential designs. In-house development will also be studied since a low cost TDRSS-compatible transceiver could have an immense positive impact on many current and future programs.

Power Supply Subsystem

Power conditioning and battery charging/discharging control will be the main features of this subsystem. Status information will be available to the Interface Unit Subsystem to aid in the autonomous decision-making capabilities that TASTE should possess. In-house development is the most likely option.

Interface Unit Subsystem

This subsystem will have two major functions: provide necessary signal interfacing between the TASTE and the AASTO computer, and control operation parameters and modes with "intelligent" capabilities. A single board computer is the most likely alternative for the implementation of this subsystem. An expansion board might be necessary to accommodate a variety of communication ports which could include RS-232 Serial, Parallel, and Ethernet. A major challenge will be to obtain the necessary functionality within the tight power budget that currently exists. In-house development is the most likely option.

CURRENT/FUTURE ACTIVITIES

Tasks/Schedule

A computer model is currently being developed, using SystemView™ software, to simulate and validate both system and subassembly performance of TASTE within NASA's Space Network. The. model will include parameters usually found in budget link calculations but with the enhanced capability of dynamic variations to approximate as closely as possible the conditions to be found in au Antarctic setting. Completion and demonstration of the model is planned for early Spring of 1998.

Test data from NASA's SPTR Winter 1997 Antarctic testing activities will be incorporated into the model. One of the most important aspects is the validation of signal attenuation assumptions due to multipath cancellation and atmospheric refraction. The line-of-site acquisition of TDRS- 1 is extremely close to the horizon and the landscape can greatly influence the test results. It is estimated that the main TASTE subassembly to be impacted by these results will be the antenna.

Prototype design has begun at the University of Puerto Rico, Mayagüez Campus, Wireless Communications Laboratory. Special emphasis is being place in using commercial of the shelf (COTS) components and subassemblies as much as possible to produce a low cost design. With the use of several CAD tools, together with our TASTE computer model, a prototype design for a low cost TDRSS-compatible portable earth terminal should be completed by late Spring of 1998. Once the prototype design has been completed we will proceed with construction, testing, and validation as a TDRSS-compliant terminal. Most of these activities are planned to take place at GSFC beginning Summer of 1998.

Related Technology Development

The design and prototype construction of TASTE can produce a variety of developments in related technologies. Some areas with high potential include:

1) TCP/IP for Satellite Applications: Optimization of current methods, protocols, and algorithms to increase efficiency of satellite Internet-compatible applications.

2) Low-cost TDRSS-compatible Transceivers: Development of spread-spectrum modem with COTS modules could increase the use of TDRSS services for many projects and experiments which currently find themselves "priced out" of NASA's Space Network.

3) S-Band Antennas: Low profile, high gain array antennas would contribute to TASTE and other projects.

4) Development of a "product line" of low-cost earth terminals. Flexible, modular designs could allow for the manufacture of many units, thus bring the costs down and developing a wider base of TDRSS users worldwide.

REFERENCES

1) NASA. November 1995. *Space Network User's Guide, Revision 7.* GSFC, Greenbelt, MD.
2) Larson and Wertz. 1992. *Space Mission Analysis and Design*, Microcosm, Inc. Torrance, CA.
3) Israel, David. May 1996. *McMurdo TDRS Relay Station.* NASA Networks '96 Workshop, GSFC. Greenbelt, MD.
4) NASA. *South Pole TDRSS Relay Home Page:* http://rodent.gsfc.nasa.gov/sptr/index.htm/, GSFC. Greenbelt, MD.
5) JACARA. *The AASTO Home Page:* http://www.phys.unsw.edu.au/~mcba/aasto.html/, University of New South Wales, Australia.
6) NASA. *Tracking and Data Relay Satellite System (TDRSS) On-line Information Center:* http://www530.gsfc.nasa.gov/tdrss/, GSFC. Greenbelt, MD.
7) Institute for Physical Science and Technology. *Upper Atmospheric Physics Home Page:* http: //www.polar.umd.edu/, University of Maryland, MD.

98 URC051

Automated TDRSS RF Relay System through CTV

Javier Díaz-Serrano* and Rafael Fernández-Sein
UPRM-NASA GSFC
University of Puerto Rico

1 Abstract

Checkout of spacecraft (S/C) telemetry and command capabilities w-as performed on a satellite prior to being launched into orbit. The Compatibility Test Section has a Compatibility Test Van (CTV) that verifies Ground Network (GN) and Space Network (SN) compatibility. In this last mode of testing, the van relays the forward link command and ranging signals from Tracking and Data Relay Satellite System (TDRSS) to the Spacecraft Under Test (SUT) and the return link telemetry and ranging signals from SUT to the TDRSS following a procedure outlined by Compatibility Test Section (CTS). This procedure is rather involved, so there is a requirement for computer-supported calculations and control. The scope of this project was to produce a user-friendly control environment with software (S/W) to support CTV TDRSS Relays. The man-machine interface (MMI) should improve methods for functions such as link calculations, antenna positioning, sun or TDRSS pointing angle calculation, and calibration procedures. This automated system was created using LabVIEW environment and programming language, which supports graphical user interface (GUI), input/output instrument access in addition to multitasking capabilities. This automated environment improved procedures in the following areas: 1) link calculations, 2) networking capability, 3) sun angle calculations, 4) TDRSS angle calculations, 5) calibration of instrumentation and testing path, 6) antenna positioning control, 7) G/T figure of merit calculation. This automated system was based on a two-way design method, top-down design and bottom-up design. The top-down design is concerned with the creation of the front panel interface, callers/callees front panel interaction, and canonical program structure. In the other direction, the bottom-up method deals with modular code creation like I/O drivers, file drivers, mathematical subroutines, and any other tasks. The general idea is to reach a user-friendly GUI and expandability issues covered. Considering the detail that an automated system is like a stand-alone application, which is intended to be error-proof, by creating exceptional condition handling routines *Then* LabVIEW seems to have advantages because is top-dawn oriented by the concept of virtual instrument (VI) -front panel- and graphical language –coding- approaches.

2 Keywords

Antenna Positioning, GPIB, Ground Station, GUI, Link Calculation, MMI, Spacecraft Validation, TDRSS, VI

3 Introduction

In general, the Compatibility Test Van (CTV) verifies that spacecraft (S/C) telemetry, tracking, and command parameters, as well as equipment and operational procedures [1], are adequate for flight mission requirements (refer to the *Generic CTV Equipment Block Diagram*). There are two modes of

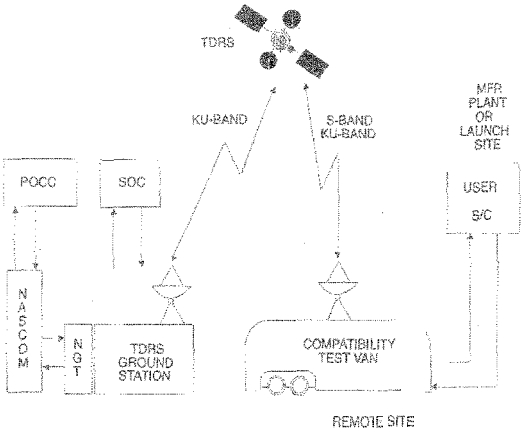


Fig. 1 Generic CTV Equipment Block Diagram

operation, Ground Network (GN) mode, and TDRSS mode, also called Space Network (SN) mode. In the particular case of this project, the TDRSS RF relay configuration will be discussed. This mode adds some additional issues to be addressed. antenna handling, and all required performance determination of the CTV as a receiver Ground Station (GS). Accordingly, the scope of this project was to produce a user-friendly control and measurement environment by S/W to support CTV TDRSS Relays. Making this automated system ensure reduction of testing time, analysis time, and instrument handling errors. Subsequently, this computer controlled system guarantees accurate and repetitive measurements. The intended automated environment should improve or create methods and/or procedures in the areas enumerated in the abstract. The phases that compose this project are: Antenna control manual mode or open loop, GPIB measurement equipment and close the antenna control loop, and finally reach remote

control by Internet, and adding more functions or features.

4 Methodology

Before beginning the discussion of the project first part let us focus on the connotation of each part implicated. The first part is concerned with the antenna control manual mode solely which comprise the familiarization with antenna controller interface (ACI), the RS-232 driver creation to communicate with ACI (send angle commands (CMDs) and read actual angles from pedestal encoders), QBASIC S/W functional and critical evaluation, and the structural creation of the LabVIEW/Windows NT based integration of the Display and Control Modules (DCMs) already available in the CTV (i.e., ANTCNTR, SUNANGLE, etc.). The following part has the purpose of the complete access to H/W equipment as HP 8563A spectrum analyzer and HP 34401A multimeter by AT-GPIB+/TNT card controller. Their drivers creation include calibration routines of instruments, This second part will permit to close the control loop of the antenna pointing routine, call Tracking Mode thanks to the multimeter measurements. Opening also the viability to determine performance figure of merit of the CTV as a receiver station: G/T. This parameter is found by solar calibration test method. Basically at the end of the second phase it is desirable to have a functional automated system of at least the same performance and capabilities of the QBASIC/DOS based MMI already implanted. The third phase could include features such as networking, automatic data entry Vs program update, pass summary Data/CMD quality, and as a future concern it is useful to have different predict (Keplerian element sets [2]) formats (i.e., NASA, AMSAT format. etc.) access to ensure compatibility with commercial-of-the-shelf (COTS) software available.

4.1 Review of Specifications

Accordingly to the first phase, the Antenna Controller Interface chassis translates inputs and outputs between RS232 and a non-standard format compatible with the "remote" control interface at the Rotating Precision Mechanisms, Inc. Antenna controller model CD-2002. The azimuth and elevation angles are parallel 12-bit binary values with 360 degrees as the most significant bit. They are transferred to the controller using time-shared command angle data lines. The 12-bit parallel antenna position data from the shaft angle encoders is accepted from the controller on time shared actual angle data lines, The Antenna Controller Interface chassis provides all appropriate controls and handshakes for these transfers. The ACI chassis acts as the data communications equipment (DCE) on the RS232 Additionally, the second phase required some changes because the actual H/W configuration has a HPIB card (HP 82990A) and a internal card multimeter. This instrument organization has been replaced to only one GPIB card to access the instruments.

The development of the NT platform automated system was based in two-way design method, top-down design and bottom-up design. The idea of the top-down design is to begin with the big picture. R deals with the creation of the front panel interface-which is called VI-, callers/callees front panel interaction, and canonical program structure. The other approach is the bottom-up method that is comprised of modular code creation. This method accounts for variable structure definition (e.g., cluster, string, integers, logical, etc.) and their usage among the subVIs; it also regards about any subroutine variable transfer arrangement and allowed features (i.e., error event handling, debugging mode operation, and by-pass execution).

4.2 Top-Down Design

It was decided to emulate the actual DCMs due to the relationship between such programs and the procedures outlined in the reference [3]. Therefore, it was created block diagrams shown in the fig. 2. As

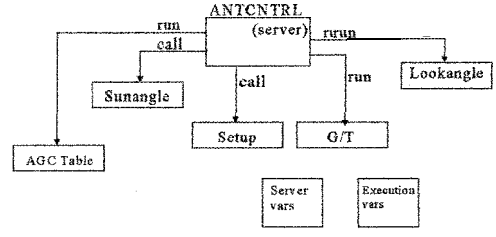


Fig. 2 Display and Execution Hierarchy

shown, the ANTCNTRL or main program is like a server; the user will be able to tail or run any other subprogram needed. Some of them could be ruining in parallel due to the multitasking nature of LabVIEW in the Windows platform. Each one of those boxes shown will have its own front panel as independent windows or pages. The boxes with the name of Server *wars.* and *Execution vars. are* global variable banks, they deal with internal program data. The Server vars. are variables that are common between more one program including the server. In others words, this is the channel by all programs can send/receive data between themselves. Another advantage of this approach is in the debugging process, because it allows change any parameter that affect all programs, to maintain essential data surveillance, and also to allow future server-client program functionality improvement and/or expansion. Meanwhile, the other bank is exclusively

used to confront internal data handling of the server and its subroutines. Its content could be number of elements to read from the setup file, flags that allow bypassing some routines (applicable to a debugging mode), storage of the last read angle antenna position, among other uses. in general the execution hierarchy can be classified as follows: 1) Menu-driven: It has the concept of "open-wait-close" (O-W-C) avoiding the crowding of the virtual control panel, also making interface more functional organized. The concept O-W-C works as follow: the main program calls a sub-program, waits until it finish and then close it. This method is similar to the DOS based technique: menu inside of another menu. Sometimes it is necessary for programs dependent of other program sequences. 2) Client-server: R could be call also as rue-driven. It is different from O-W-C because the main only wants to open the sub-program and make it run, let call it: "open-run" (O-R). This architecture allows real multitasking, for programs that are mutually exclusive.

Now looking at the Fig. 3, it is evident that a VI concept is well understood. Clearly it is complementary way to make top-down design. There

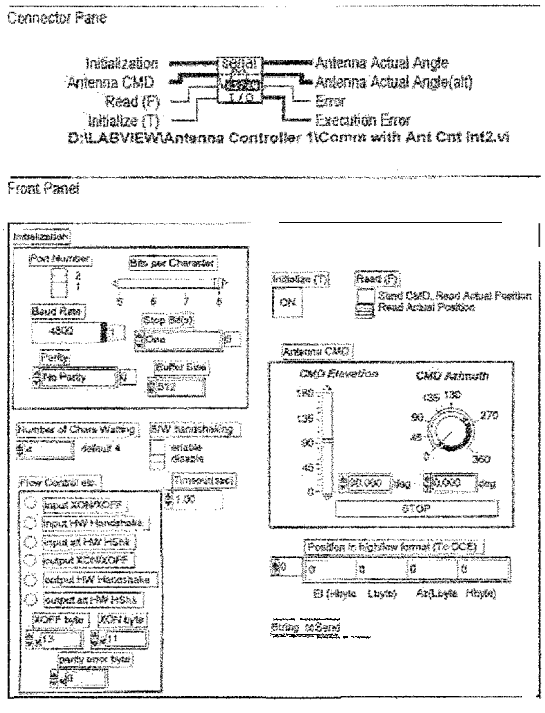


Fig. 3 Antenna Controller SubVI

is also show the Vi connector pane, which show input/output descriptions related to this particular VI.

Now explaining the coding design, still from the top-down approach. It was decided to maintain more or less the same resemblance of the DOS DCMs. The translation of this model to LabVIEW was done by the state machine canonical structure. Each group of states will have a different rate of execution, in other words, the keyboard state could be running a slower pace that port reading state, for example. This ensures that the time slicing of the computer to be controlled, a feature fully supported by LabVIEW. By this way, it is possible to put some kind of priority in the program execution sections.

The advantage of the state machine programming philosophy are the definition of the borderline between the text based language (sequential programming) and the G language (data flow programming or parallel execution capability); and an enhanced capability and flexibility of special handling subroutines creation.

4.3 Bottom-up Design

This approach is based on the philosophy of divide and conquers. To illustrate this point, please refers to fig. 4: *Hierarchy VIs Relationship*. It describes the

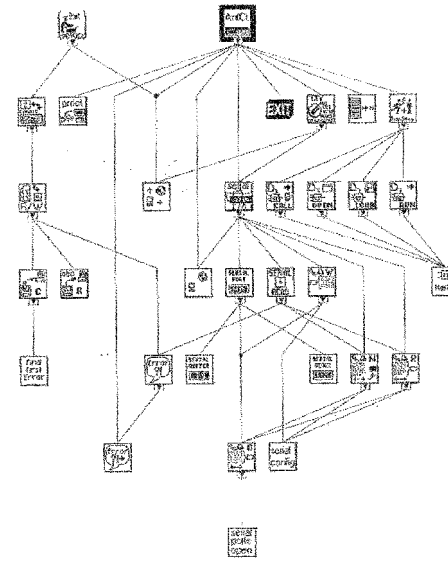


Fig. 4 Hierarchy VIs Relationship

program arrangement of the first phase of this project. In general, each VI created is briefly -in columnwise direction- explained as follow:

SETUP.vi is the program that allows the user to enter a new initial configuration on S/W or H/W parameters required to achieve the testing event.

ANTCNTRL.vi is the main program where all other programs communicate through.

SETUP STORAGE.vi deals with the download or upload process within the Setup cluster definition and the file ANTCTRL.SUP, additionally it makes the necessary formatting PREDICT.vi is intended to deal with the file that contains the predicts created by any Kepplerian set COTS software developer,

EXIT.vi is the routine program ensures that the closing procedure to be accomplished.

HARDWARE INIT.vi do the opposite to EXIT.vi but only concerning with instrument or port configurations.

Which Botton.vi is used as a subroutine that indicates which button was pressed in the Calls Area.

CALLER.vi is the programs that permit dynamic loading execution of the program selected *in* the Calls Area. Dealing with the loading properties O-P. and O-W-C

READ_WRITE SETUP.vi is merely a subroutine of SETUP STORAGE.vi that read or write the setup contain an ASCII format to ANTCNTRL. SUP file.

SERIAL ACCESS.vi is the comport driver that encode the CMD angles and decode the actual angles to be written or read to ACI, respectively,

Notice, although the ANTCNTRL VI calls SETUP VI they are not linked, this implies that both program should not be loaded at the same time in memory (of course, only when SETUP has been called). Observe also, that ANTCNTRL VI, SETUP VI, and H/W CONFIG VI have access to SERVER GLOBAL BANK VI, meanwhile to EXECUTION GLOBAL BANK VI only the sewer and its subroutine (in this case the RS-232 driver) have access to it.

5 Testing and Debugging

The VIs have been tested in modular fashion, including the file management area, server-client interaction tests, error generation, serial port configuration and communication test. At this last point, it was proceeded to develop a kind of Antenna Controller Interface (ACI) simulator in a second computer. It was made in QBASIC due to lack of LabVIEW S/W in this second computer. The first step was to make some simples write/read through a null-modem cable (this particular cable did not include any handshaking lines) between the two computers without any type of protocol involved. It was done using also a break-out box, a box that permits you to see the status of the communication lines and make debugging, At the end, a real emulation of the antenna sending CMD and encoders angles reading was made. When the ACI schematics were analyzed: the DTE was fooled by sending its own request to send (RTS) signal back as clear to send (CTS) return, therefore the DCE did not need any H/W handshaking nor DTE. Therefore, the features of the H/W handshaking embedded in the old QBASIC program (also included in the VI driver designed) were unnecessary, After the emulation phase was completed, the real system was connected. The LabVIEW driver developed could move the antenna at the preset CMD, also it was able to read-back the actual antenna position. And the precision of the angle read/write process is around one tenth of degree as desire.

6 Conclusions and Projections

In summary, the development of this project envisions a new era of the testing procedures: GUI/Internet based system empowering the users to use off-the-shelf H/W in a MMI environment. This is noticed in the intuitive way to control the antenna steering and access customized instruments measurements as G/T among others. Matching perfectly with the idea of increase reliability and repetitiveness of testing procedures[5]. In this particular case to ensure absolute validation under MO&DSD stipulations.

From the programmer point of view, the new language principle called G data-flow programming enable to take the advantage of multitasking capabilities of the systems nowadays available. Also in the specific case of LabVIEW it has the virtual instrument design model along with the iconning or flowchart style of programming more insightful for engineers and developers.

Now the projections will be focused. Details interesting to approach will be to have DDE communication between the Configurable Anal ysis Graphical Environment (CAGE) space systems oriented S/W developed by Stanford Telecommunication, with the created LabVIEW application. This could be useful to animate the TDR satellite link event additionally with the antenna pointing view; considering that CAGE has already libraries emphasized in space systems simulations and analysis. Also additionally to the anticipated features such as network remote access and more GPIB instrument control capabilities; it is for sure to make improvements on actual issues not fully covered:

Correction of numerical errors or imprecision in analysis routines. Improve maximum signal tracking algorithm by reducing the convergence time delay

Correction of H/W glitches in the antenna custom-made controller I/F

Reduce the possibility of antenna damage by wrong usage by including obstruction masking feature

Embed the S/C six elements calculation into the program

7 References

[1] Inter-Range Instrumentation Group(IRIG) procedures, Vol. 1

[2] Space Mission Analysis and Design, Wiley J. Larson & James R. Wertz

[3] STDN S/C RF Compatibility Test Procedures and Data Sheets, STDN no. 408. 1

[4] Space Network (SN) User's Guide, 530-SNUG Rev. 7

98 URC052

NEURAL NETWORKS FOR SUBPIXEL CLASSIFICATION OF MULTISPECTRAL IMAGES[1]

Ricardo R. Figueroa, Shawn D. Hunt
e-mail ricky@exodo.upr.clu.edu
Department of Electrical Engineering
University of Puerto Rico, Mayagüez Campus

ABSTRACT

In this work, the implementation and use of AVHRR images for subpixel analysis is studied. The work consists of two parts, the first is making training data, the second is using the training data to design a neural net subpixel analyzer. Most work on subpixel analysis has been done with images with more spectral bands. AVHRR images were chosen because of their easy acquisition, and because the five spectral bands allow investigation into the development of training data. The first step in subpixel analysis is the development of training data. This consists of the image to be classified, and the classification of each pixel. In order to do the classification, a high spatial resolution image is typically needed in order to manually create a classified image. It is difficult to have both the image of interest, and a high spatial resolution image of the same area taken at the same time. Thus it was studied whether a subsampled image taken from the image of interest could serve as the training data. Statistical work has been done showing the unusefulness of this approach. In the second part of the analysis, a feedforward neural net was trained and used to classify the AVHRR images. Results of these tests, comparing the neural net with typical statistical based schemes are shown.

INTRODUCTION

The world today is one in which the information about our planet is very significant to everyday decisions, made by the military, environmentalists and even ourselves in our everyday life. Satellites give us a broad picture of how our world develops and behaves, and helps us understand and predict many earth processes, that would be difficult to understand without these images. There are many satellites orbiting the earth gathering information and sending it back to different earth receiving stations. Most of these satellites return data gathered from multispectral sensors [1]. Satellites give us different views of the earth, images that vary in resolution, but not always showing us what we want to see.

1. Acknowledgment This work was sponsored by NASA Grant No. NCCN-0088.

An image with a spatial resolution of 1 km only lets the viewer see objects that are larger than 1 km, since each pixel in the image represents lx 1 km of earth [2]. The pixel represents the spectral combination of all the objects included in the specified area.

The objective of this investigation is to obtain information that is contained within one pixel, by this we mean that we are trying to identify objects of interests that are smaller than the spatial resolution of an image pixel. All the information we have from the object of interest are it's spectral characteristics. This is how the pixel represents the object and everything else that is around it. For this task neural networks are presently being used.

IMAGE STATISTICS

As mentioned before, it is difficult to obtain an image of interest and a high spatial resolution image of the same area, taken at the same time. In this study a high resolution image was subsampled to half its original resolution, and it's first and second order statistics were calculated. The results of these measures for the cloud samples for the image are shown below.

High Original Resolution Image
Mean
Ch1=.3912, Ch2=.3378, Ch3=.8821
Ch4=.8211, Ch5=.8038

Covariance Matrix

.0016	***.0012***	***.0005***	***.0006***	***.0005***
.0012	.0010	.0004	.0005	.0004
.0005	.0004	.0003	.0003	.0003
.0006	.0005	.0003	.0004	.0004
.0005	.0004	.0003	.0004	.0003

Standard Deviation
Ch1=.0400 Ch2=.0311 Ch3=.0170
Ch4=.0209 Ch5=.0175

Correlation Coefficients Matrix
1.0000 *.9956 .7887 .7493 .7285*
.9956 1.0000.7973 *.7861 .7673*
.7887 .79731 .0000 .8909 .8750
.7493 .7861 .8909 1.0000.9962
.7285 .7673 .8750 .9962 1.0000

Low Resolution Image Mean
Ch1=.3875, Ch2=.3351, Ch3=.8807
Ch4=.8197, Ch5=.8026

Covariance Matrix
.0020 .0015 .0007 .0008 .0006
.0015 .0012 .0004 .0006 .0005
.0007 .0005 .0003 .0004 .0003
.0008 .0006 .0004 .0005 .0004
.0006 .0005 .0003 .0004 .0003

Standard Deviation
Ch1=.0443 Ch2=.0341 Ch3=.0184
Ch4=.0220 Ch5=.0 183

Correlation Coefficients Matrix
1.0000 *.9965 .8541 .8071 .7880*
.9965 1.0000.8583 .8345 .8173
.8541 .85831 .0000 .9220 .9078
.8071 .8345 .9220 1.0000.9972
.7880 .8173 .9078 .9972 1.0000

Based on these results, it was assumed that the subampling of the high resolution image for obtaining the training data could be done, but after the network classification error did not improve from around a 680/0 classification percentage, a hypothesis testing procedure was carried out to see if the sets of data form the high and the subsampled low resolution images came from the same distribution. The Wilcoxon Rank Sum test was applied to the data sets to see if they had the same mean, giving the result of a 99.5% **certainty** that the data sets did not have the same mean and consequently did not come from the same distribution. This meant that the statistical characteristics of the data were changed by subsampling the high resolution data.

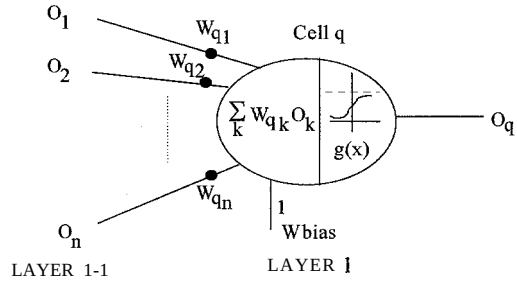
NEURAL NETWORK CONFIGURATION

For this research a feedforward neural network topology was used. The first layer is the input, and each cell in one layer is fully connected to all other cells in the following layer. The output of the last layer is the output of the network. All layers between the input and output layers are called hidden layers.

A linear combination of the outputs of one layer form the input to the next layer after being passed through anon-linear Sigmoid function. Let O_q be the output of a cell q:

$$O_q = g(I_q)$$
$$= g\left(\sum_k w_{qk} O_k\right)$$
$$= g\left(\sum_k w_{qk} g(I_k)\right)$$

where w_{qk} is the weight connection between a cell k in layer l and cell q *in* layer $l+1$, and $O_k = g(I_k)$ is the output of cell k. *[3]* An illustration of this computation is shown in Figure 1.



Neural Network Cell
Figure 1

A *2* hidden layer network was implemented using MATLAB software.

For training and simulation the network received 2x2 blocks of pixels as input, which make up 20 inputs for each pixel classification, since the image is divided into 5 bands.

The configuration for each hidden layer was:
1)20 inputs --4 nodes in the fist hidden layer, and 5 nodes in the second hidden layer. (Variations in the amount of hidden neurons were performed)

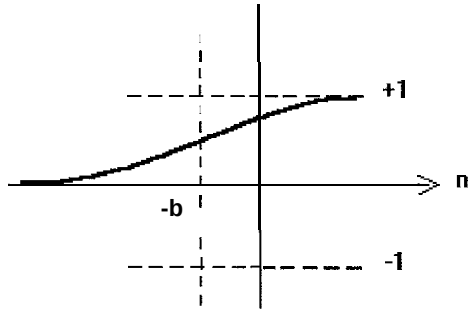
This configuration gave the best results in previous work done with a 3 band infrared image.

TRANSFER FUNCTION

The logistic Sigmoid, or log-Sigmoid, activation transfer function is often used with neurons being trained with the Backpropagation

Algorithm, and it is also applicable to the Levenberg-Marquardt Algorithm. The log-Sigmoid function is used to map a neuron input from the interval (-∞,+∞) into the interval (O,+ 1) [4]. Here is how the log-Sigmoid equation is applied to each element of the input matrix N with a bias vector B, followed by the log-Sigmoid graph:

$$\log sig(n,b) = \frac{1}{1+e^{-(n+b)}}$$



Log-Sigmoid function
Figure 2

CLASSIFICATION

The objective is to identify cloud coverage in AVHRR multispectral images. Each AVHRR image is composed of 5 different images, each representing a band of the wavelength or frequency spectrum. The information in each band is used for the neural network for classification. As mentioned previously, the classification was to be done on a low resolution image, so that the object of interest, clouds, would be smaller that the image's spatial resolution.

This was achieved by taking 2x2 blocks, taking the average of the four pixels, and converting this value into a new pixel. The image was converted to half its original size and the object of interest became smaller than the spatial resolution of the pixel that contained them.

After training the network with the low resolution image as input and the target pixels, already identified from the high resolution image, as outputs, the weights obtained by the network after minimizing the error were used on other similar images for classification,

During training and simulation, it is desired that the network return a value of O as output if the pixel does not contain the unit of interest, otherwise, return a value of 1. The network will change the value of the weights after each iteration, and return the values that minimize the error in some sense. These values will then be used for the simulation of the classifier. During simulation a threshold is performed since the error in the neural network is not O, and the value returned by the network when a pixel is positively classified, will not be exactly 1. The threshold was set to .7, so any output of the network that has a value equal or grater than .7, will be classified as having the unit of interest.

MINIMUM EUCLIDEAN DISTANCE CLASSIFIER

The neural network was compared to the Minimum Euclidean Distance classifier. This classifier assumes covariance equal to 1, and it follows the following equations:

$$g_i(X) = \quad M_i)^T (X \quad M_i)$$
$$g_i(X) \geq h_j(X) \forall j \rightarrow x \in w_i$$
$$M_i = E(X / w_i)$$

where the g_i and h_j *are* called discriminant functions, X is the vector being classified and the M_i is the expected value for class *i.* [5]

RESULTS

The first results obtained were using simulated multispectral images, RGB images, which have only three bands of information. This results produced about a 60% classification accuracy. Currently, AVHRR multispectral images are being used for the research, and improvements have been obtained, since they contain more spectral information.

The best results were obtained using the Levenberg-Marquardt algorithm for training the neural network.

The most recent results showed around a 65-70% classification accuracy using 20 inputs to

classify each AVHRR pixel, but as mentioned before, the statistical characteristics of the data were changed by the subsampling, and the network classification accuracy has not improved with consequent training. The Minimum Euclidean Distance classifier produced a 40 to 50% classification accuracy, which was expected to be low.

A neural network classification example using AVHRR images is shown in Figure 3.

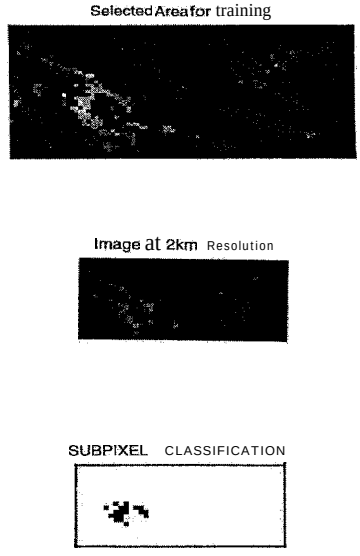


Figure *3*

REFERENCES

[1] Remote Sensing. March 1995. IEEE Spectrum Magazine. 24-31

[2] Subpixel Analysis: Who Needs it?. Robert L. Huguenin. July 1994. Earth Observation Magazine. *37-40.*

[3]http://www.ph.tn.tudelft.nl/Research/neural/feature_extractions/papers/thesis

[4] Neural Network Toolbox Manual. H. Demuth, M. Beale. 9-50.

[5] INEL 6007, Dr. Luis Jiménez Classnotes. UPR Mayagüez, 2nd semester 1996-97.

98 URC053

Estimation of Velocity Field from Oceanographic Image Sequences

Eungchun Cho, S. S. Iyengar
Division of Mathematics and Science, Kentucky State University
Department of Computer Science, Louisiana State University

December 12, 1997

Abstract

Oceanographers desire accurate methods of tracking features in satellite images of the ocean in order to observe and quantify surface layer dynamics. Infrared images of the ocean showing sea surface temperatures are widely used for the studies of this type.

Feature tracking from time series of satellite IR images poses two problems: First, the features of interest have weak edges and constantly evolving shapes from image to image. That means features merge, split, grow, shrink, disappear are created on time scales that are comparable to the sampling interval of the satellite imagery (I 2 hours). In other words, the phenomenology under investigation is Turbulent Fluid flow, not rigid body motion. The second problem which results from the first, is feature motion cannot be defined by parameters that are functions of scale as well as space and time. A simple example of different motions associated with different scales is seen in the ocean "front". Most ocean fronts exhibit shear across the frontal boundary.

In this paper we present some new results in our oceanographic velocity of the image flow and more specifically, we discuss a new approximation method to estimate the velocity field of Gulf stream from a sequence of satellite images. In this method, connected components of the region representing a stream is identified, triangulated and the velocity field on the region is estimated by the affine approximation on each triangle.

1 Introduction

Our primary objective is to get an estimate of the velocity vector field of moving regions in a given sequence of images. In oceanographic images, for example, we are interested in tracking the movement of mesoscale features such as Gulf streams and Eddies. We consider following two subproblems for the problem of estimating the velocity field of moving regions.

Boundary extraction and tracking — The boundaries of a region are extracted by applying segmentation and edge detection operators. The extracted boundary is approximated by a piecewise linear curve, which is a continuous concatenate ion of line segements. A piecewise linear curve is also called a polygonal curve. For example, a circular curve can be approximated by a regular polygonal curve. Though more elaborate higher degree spline approximation techniques are available, the standard linear approximation of the regions is suit able for our method of estimating the velocity vector field. The region itself is triangualted into a simplicial complex based on the piecewise linear boundary of the region. In other words, we triangulate (i.e., construct a piece wise linear approximation of) the region (a two dimensional object) based on the piecewise linear approximation of the boundary curve (a one dimensional object). Such approximation for any dimension is called triangulation. Those constructions can be applied to higher dimensional objects (surfaces, solids, or n-manifolds) recursively.

Motion estimation — Given simplicial complexes approximating the regions in the image sequence and simplicial maps of the complexes of lower dimensions (called skeletons), we construct a simplicial map of the whole complex approximating the motion of the regions by an affine interpolation of the simplicial maps of the skeletons (complexes of lower dimensions) of the regions. In our cases, the skeletons are the piecewise linear curves (together with their vertices) approximating the boundary of the region. The velocity vector field describing the motion of the regions will be estimated from the simplicial maps between the 2-dim simplicial complexes. The piecewise linear boundary curves are l-dim simplicial complexes and the triangulated regions are 2-dim simplicial complexes.

In oceanographic image analysis and understanding, the estimation of velocity field is particularly difficult because of the non-rigidity of the motion of the features. Also, the chaotic nature (turbulence) of fluid dynamics exacerbates the problem.

The MCC (maximum cross correlation) method proposed by Emery et al. computes surface advective velocities by using information from lagged matrices between subareas of image sequences. However, this method is insensitive to rotational components of feature motion.

Sethi and J ain showed that spatio-temporal coherence in extended frame sequences can be used in computing object trajectories throughout an image sequence.

2 Noise removal and edge extraction

To obtain a grey level image with noise removed, we perform an adaptive threshold and median filter operation on the raw image data. We apply further noise removal, segmentation and edge extraction operations on the image. These op-

erations can be done by applying mathematical techniques such as variational method or wavelet method.

Once we have a noise removed, segmented and edge detected images, the boundary curves of the regions are replaced by piecewise linear boundary (edge) curves obtained by approximating the points on the boundary curves with straight line segments. For step by step approximation procedure, we may assume the boundary points are ordered from left to right and from top to bottom.

It is sometimes convenient to use a metric other than the usual euclidean metric on R^2 given by $d_3((x_1,y_1),(x_2,y_2)) = \sqrt{(x_2-x_1)^2+(y_1+y_2)^2}$ We can use a metric given by $d_1((x_1,y_1),(x_2,y_2)) = |x_2-x_1|+|y_1+y_2|$ or by $d_2((x_1,y_1),(x_2,y_2)) = max\{|x_2-x_1|,|y_1+y_2|\}$ where $(x_i,y_i) \in R^2$. These three metrics are all equivalent in the sense that they define the same metric structure and topological structure on the space R^2. The geometries, however, are different. With the euclidean metric d_3, the set of equidistant points from a point is a circle. With the other metrics d_1 and d_2, the set of equidistant points from a point is a square. For example, The set of all points of the unit distance from the origin (O, O), $\{(x,y) : d_1((0,0),(x,y)) = 1\}$ is the (boundary of the) square with vertices at (1, O), (O, 1), (– 1, O) and (O, –l). We may use any of these metric, denoted by d, in this paper. It may be assumed that d is the usual euclidean metric, if necessary,

3 Break points, extreme points and piecewise linear approximate ion

Suppose we already have found the boundary. More specifically, assume we have the (ordered) list of all the boundary points. We want to construct a sublist of the boundary points to serve as the vertices of the piecewise linear approximation of the boundary. We will start with the first three boundary points, say p_1, p_2, p_3. If pi, the i-th boundary point, is within the given threshold distance from the line segment $[p_{i-1}, p_{i+1}]$, then we replace the point p_i by p_{i+1} in the list. However, if the distance is greater than the threshold, p_i is kept and it becomes the initial point of the next approximation step. That is, we check if the point p_{i+1} is within the distance threshold from the line segment $[p_i, p_{i+2}]$. Iterate this procedure until there is no reduction of points in the list. The points in this reduced list is used to construct a polygonal curve (piecewise linear curve) approximating the boundary of the region. The points in the sublist are the vertices of this polygonal curve (the end points of the line segments in the polygonal curve), and they are called the break points on the boundary.

It is obvious that the polygonal curve makes a turn (changes direction) at each breakpoint. The amount of turning at the breakpoint will be measured by

the cosine value of the angle of turning at the break point p_i which is computed by the usual inner product in R^2 $\cos\theta_i = \frac{(p_i - p_{i-1}) \cdot (p_{i+1} - p_i)}{\|(p_i - p_{i-1})\| \cdot \|(p_{i+1} - p_i)\|}$. For example, consider three points $p_{i-1} = (-4,3)$, pi $= (0,0)$, and $p_{i+1} = (4,0)$. The angle of turning at pi is computed by $\cos\theta_i = \frac{(4,-3).(4,0)}{5.4} = \frac{4}{5}$ and θ_i, the angle of turning, is about 37 degrees. The angle of turning and the interior angle at a point add up to 180 degrees, in other words, they angles are supplementary. The cosine value will be close to 1 if the change of the direction is small and will be much less than 1 or even negative if the change of direction insubstantial. If the cosine vain at p_i is less than a predetermined threshold, we will call the point p_i an extreme point. We may consider this definition of extreme point as an analog of the points where the curvature of the curve attain local extrema.

We use the extreme points to estimate the velocity vector field at each stage of the image sequence. In the following example we consider, for simplicity, a pair of extreme points only in the image sequence. Suppose a_1, a_2 are extreme points in the image and b_1, b_2 are the correponding extreme points in the next image in the sequence. The velocity vector at a_1 is estimated by $b_1 - a_1$ and the one at a_2 is estimated by $b_2 - a_2$. The velocity vectors at a_1 and at a_2 are linearly extended to a vector field on the line segment joining al and a_2 as follows. Any point p on the line segement joining al and a_2 can be written as $p = ta_1 + (1-t)a_2$ for some $O \leq t \leq 1$. We estimate the velocity vector at p by $t(b_1 - \text{al}) + (1-t)(b_2 - a_2)$ More specifically, let $a_1 = (O, 1)$, $a_2 = (O, O)$ and $a_3 = (1,0)$ be extreme points and $b_1 = (1,3)$, $b_2 = (2,1)$ and $b_3 = (3,1)$ be corresponding extreme points in the subsequent image. Then the velocity vector at $p = (x, y) = xa_1 + ya_3$ is estimated by $x(b_1 - a_1) + y(b_3 - a_3) = x(1,2) + y(2,1) = (x+2y, 2x+y)$ This computation can be extended to the entire region using the triangulation based upon the piecewise linear approximation of the boundary curve. This procedure is called a simplicial approximation of the velocity vector field.

4 Estimating Velocity vectors from t he approximation of region boundaries

Suppose a region in the image sequence is identified and its boundary curves are approximated by piecewise linear curves. Suppose also extreme points on each piecewise linear approximation of the boundary curves are found. We are assuming the features in the oceanographic images are such that the extreme points of one image move to the extreme points of the next image in the sequence. Under this assumption, we can track extreme points and the boundary curve and eventually every point inside the region by interpolation based on the simplicial approximation of the region.

Let C be a boundary curve in an image and C' be the corresponding boundary curve in the next image of the sequence. Let $[q_0, q_1, \ldots . q_k]$ be a piecewise

linear approximation of C and $[q'_0, q'_1, \ldots q'_k]$ be a piecewise linear approximation of C'. Suppose {cl, ... e_m} is the set of all the extreme points of the curve Canal $\{e'_1, \ldots, e'_m\}$ the set of all extreme points of the corresponding curve C'. Note the sequence $\{e_1, \ldots . e_m\}$ of extreme points (resp. $\{e'_1, \ldots, e'_m\}$) is a subsequence of $\{q., q_1, \ldots . q_k\}$, the sequence of all break points of C (resp. $\{q'_0, q'_1, \ldots, q'_l\}$, the sequence of all break points of C'.)

We are assuming the number of extreme points on two curves are the same. This can be done by adjusting the preset value A (resp. A') which determines the extreme points on C (resp. C'). Since we assume the flow of the image is such that the extreme point e_i on C corresponds to e'_i on C', we approximate the velocity vector at e_i representing the movement of the curve at e_i by taking the differences $e'_i - e_i$. We can extend this approximation of the velocity vector to generate a velocity vector field on the polygonal approximation $[q., q_1, \ldots . q_k]$ of C by affine extension as follows.

Let $p = f(t)$ be a point on the polygonal curve $[q_0, q_1, \ldots . q_k]$ such that p is between the extreme points $e_i = f(s_i)$ and $e_{i+1} = f(s_{i+1})$. Let the corresponding extreme points be $e'_i = f'(s'_i)$ and $e'_{i+1} = f'(s'_{i+1})$. Then we assign the vector $f'(t') - f(t)$ to the point p where $t' = s'_i + \frac{s'_{i+1} - s'_i}{s_{i+1} - s_i}(t - s,)$.

This is an obvious extension to the velocity vector field on the polygonal approximation of the curve C. We will denote the vector field on C by V. If the curve C is closed so that it bounds a region then we can extend V to a velocity field, also denoted by V, on the simplicial approximation of the region by extending V on each simplex $[p_i, p_{i+1}, p_{i+2}]$ in the approximation as $V(f(p)) = tV(f(p_i)) + sV(f(p_{i+1}) + rV(f(p_{i+2}))$ where t,s, r are real numbers $O \leq s, t, r \leq 1$ such that $s + t + r = 1$ and $p = tp_i + sp_{i+1} + rp_{i+2}$.

5 Conclusion

A well established mathematical tool of simplicial approximation of regions on a plane (or on a surface) and simplicial approximation of 'continuous mappings between regions were applied to locate and approximate connected regions in image sequence. Once the regions of interest were approximated by simplices (called triangulation), vertices of the simplices in the approximation were identified and used to evaluate the velocity vector field of the features in the image sequence. We need to identify the corresponding points in each image in the image sequence to be able to estimate the velocity vector field. This is done by locating extreme points on the boundary curves of the region in the image sequence by assuming that extreme points on one image move to extreme points on the following image in the sequence. From these estimates of the velocity vectors, we interpolate the velocity field on the whole region by an affine approximation. This approach can also be applied for computing the velocity field induced by rigid motion of objects in dynamic scenes.

References

1 W. Emery, A. Thomas, M. Collins, W. Crawford and D. Mackas, 'An objective method for computing advective surface velocities from sequential infrared satellite images', Journal of Geophysics Research, vol. 91, no. 12, pp. 12865-12878, 1986.

2 C. Garcia and I. Robinson, 'Sea Surface Velocities in Shallow Seas Extracted from Sequential Coastal Zone Scanner Satellite Data', Journal of Geophysical Research, vol. 94, no. C9, p. 12681, Sept. 1989.

3 R. J. Holyer and S. H. Peckinpaugh, 'Edge detection applied to satellite imagery of the oceans', IEEE Trans. Geosci, Remote Sensing, vol. 27, pp. 45-56, Jan 1989.

4 K. Kelly, 'Comparison of Velocity Estimates From Advanced Very High Resolution Radiometer in the Coastal Transition Zone', Journal of Geophysical Research, vol. 97, no. c6, pp. 9653-9668, 1992

5 S. Krishnamurthy, S. S. Iyengar, R. Holyer and M. Lybanon, 'Histogram-Based Morphological Edge Detection', IEEE Transaction on Geoscience and Remote Sensing, vol. 32, no. 4, July 1994.

6 E. Sethi and R. Jain, 'Finding Trajectories of Feature Points in a Monocular Image Sequence', IEEE Trans. Pattern Analysis, Machine Intelligence, vol. PAMI-9, no. 1, Jan 1987.

7 A. Vastano and S. Borders, 'Sea Surface motion over an anticyclonic eddy on the Oyashio Front', Remote Sens. Environ., vol. 16. pp. 87-90, 1984.

8 A. Vastano and R. Reid, 'Sea Surface Topography Estimation with Infrared Satellite Imagery', Journal of Atmospheric and Oceanic Technology', vol. 2, no. 3, Sept. 1985.

98 URC054

Use of MCNP Monte Carlo Code as a Tool for Collimator Design

Daniel Gutierrez
Ron Grazioso
Michael Partridge
A. Sharif Heger
David Wolfe
NASA Center for Autonomous Control Engineering
The University of New Mexico
Albuquerque, New Mexico

I. Abstract

The use of MCNP Monte Carlo Code as a tool for collimator design was test by comparing actual collimator tests by experimental mean to the simulations generated by of MCNP. Multiple runs of simulated method and the experimental method were performed using a point source. The count detected by each method were plotted individually and then after corrections for total activity were made, they were plotted against each other. The only difference in the two methods was seen at the edges of the hole. This difference could come about because of the inability to make a true point source in the experimental method. Beside the difference at the edges the MCNP Monte Carlo Code result were very highly correlated to the experimental results,

II. Introduction

The purpose of medical imaging is to get an accurate representation of a internal organ on an imaging surface. This process has two main steps. The fist is getting a radioactive source into the desired organ and having it be emitted by the body. The second is being able to detect this emitted radioactive source. Both of these step directly influence the quality of the image. Since as engineers there is little we can do about the radioactive source traveling through the body, we will look at the detection of the emitted radiation, namely photon. The detection system consists of two elements, the collimator and the imaging plain. The collimator work basically like a lens of an optical camera, It focuses the photon on the imaging plain. Since the radiated photons are emitted isotropicly, it is the job of the collimator to single out the photons that are normal to the imaging plain. Ideally we would like to only select the photon with are truly normal to the imaging plain and ignore all others but in reality this is not obtainable. The job of collimator designer is to find the best design for a given situation. This leads to a question of how to find this best design. In this paper we explore the use of NCMP Monte Carlo code as a possible tool to designing collimator. Using a software program to simulate a collimator design will be much quicker and cheaper that building a collimator and then testing its functionality. We suggest that the use of a computer simulation can be used as a tool in designing collimator

III. Objective

The objective is to test the MCNP Monte Carlo code results to that of the actual results acquired by experimentation. This was accomplished by moving a point source from a point over one septa across the hole in small step until it is once again over another septa. This method was used in both by the experimental runs and the simulated runs. The results will then be compared to each other to determine the validity of the use of MCNP Monte Carlo code as a tool for collimator design.

IV. **Simulation**

A square hole collimator was used in both setups. The dimensions of the collimator were take directly of the collimator itself and programmed into code in the MCNP Monte Carlo program. The dimensions of the collimator are shown in figure 1. The collimator was placed directly in front of an imaging surface.

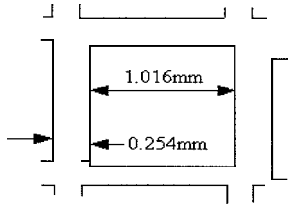


Figure 1
Collimator Dimensions

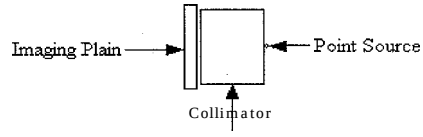
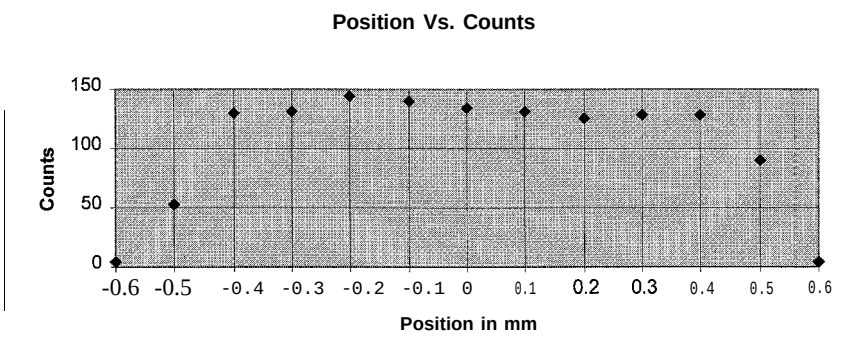


Figure 2
Simulation Setup

A point source was placed in front of the collimator. This setup is shown in figure 2. The hole side for the collimator is 1.0 16mm and the septa walls are 0.254mm. We define the center of the hole to be the origin and the step size be O.1mm, so the first position at which data was taken at was at -0.6mm. After this run was completed the source was moved one step to -0.5mmand another ran was performed. This continued for every position between -0.6mm to +0.6mm, at which point the source is again over a septa. After trying to repeat this process for a second time it was found that the results were exactly the same. To improve the set of data, runs were performed over three holes and averaged to give us the final value at each position. Each run was perform for an activity of 1 million counts. The table below shows the number of counts at each position. The average value was use to plot the graph shown below.

Simulated Results

	Position												
	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6
Hole 1	4	139	130	128	146	140	134	130	126	130	129	127	3
Hole 2	4	16	130	133	145	144	136	132	128	127	129	125	2
Hole 3	3	5	130	133	145	136	133	132	124	130	125	18	5
Average	4	53	130	131	145	140	134	131	126	129	128	90	4



Plot of Simulation Results

V. Experiment

The setup for the experimental runs was set up exactly the same as the simulation with regard to the collimator and the imaging plain. The point source was made by drilling a small hole in a piece of lead and placing the Tc-99m inside the hole with a thin piece of wire. The point source used in the experimental portion of this study is not a true point source but it is as close as we could come to one. The source was then place directly in front of the collimator just like in the simulation. In this case, because we want only the photons passing through the collimator to be counted we needed to install a lead shield to prevent unwanted photons from reaching the detector. Sodium Iodide and Germanium detectors were used for these experiments. The setup for the experiment is shown in Figure 3.

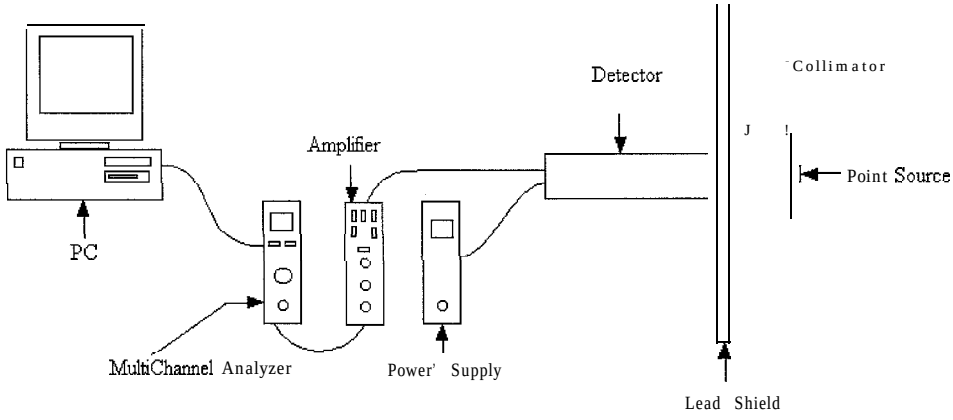
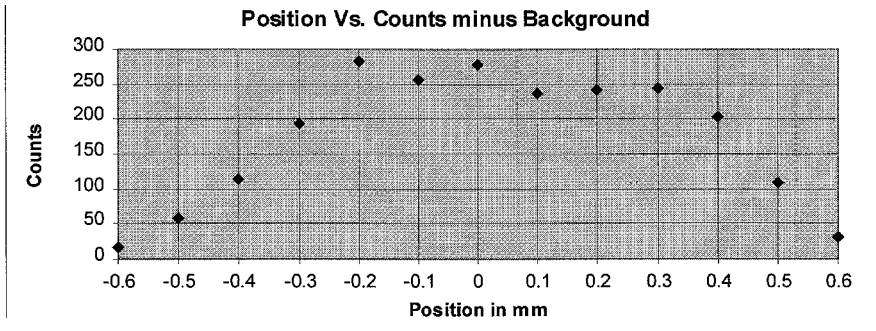


Figure 3
Experimental Setup

The detector was hooked up a setup Bertain Assoc. Model 345 0-5000V power supply which supplied the necessary power for the detector to operate properly. The signal from the detector was the input into the a Ortec 575 amplifier to be amplified. The output of the amplifier was the input to a LeCroy Model 3001 multichannel analyzer. The multichannel analyzer then sent its output signal to our PC via our data acquisition package. The data was gather in the same way we did for the simulation. With the source over one septa, the PC's data acquisition program was run and the data was stored for that position. The source was then moved by one steps of O. 1mm and the program was run again. The run time was for a period of one minute. This process continued just as it did in the simulated portion of this study until once again we were over a septa. This whole process was repeated three times of each detector and the results were averaged. The data was then truncated to just 20 channel which showed the counts due to the Tc-99m's 140KeV signal. The Table below is the result of experimental data.

Experimental Results

	Position												
Ch#	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6
1	91	91	113	112	109	108	93	80	102	104	92	92	91
2	91	102	122	102	119	111	102	102	111	108	104	90	92
3	92	93	124	113	118	103	108	108	120	102	132	92	91
4	94	101	107	102	117	147	117	105	105	116	154	114	95
5	97	111	111	116	114	140	119	127	103	119	180	145	100
6	100	120	102	127	123	173	134	167	124	145	212	151	104
7	73	115	117	154	121	263	154	173	133	180	245	192	107
8	116	134	127	224	166	288	183	223	165	240	302	189	100
9	90	138	157	244	187	284	252	316	184	278	290	208	93
10	108	138	172	284	216	355	286	334	262	342	226	168	121
11	107	157	211	292	292	311	328	315	295	310	213	183	107
12	93	138	210	270	382	273	376	323	341	325	197	117	129
13	82	132	207	239	296	241	345	283	334	279	160	128	112
14	71	114	195	181	264	165	281	275	320	220	107	104	93
15	75	95	181	136	267	115	247	180	286	155	90	71	82
16	75	70	171	89	217	118	160	129	212	123	74	776	85
17	86	86	127	84	182	96	121	114	162	111	65	74	79
18	60	65	98	89	134	81	109	89	124	82	83	65	75
19	69	58	92	60	96	74	85	64	114	68	50	63	86
20	66	69	82	61	79	68	92	64	82	75	53	62	60
Max	116	157	211	292	382	355	376	334	341	342	302	208	129

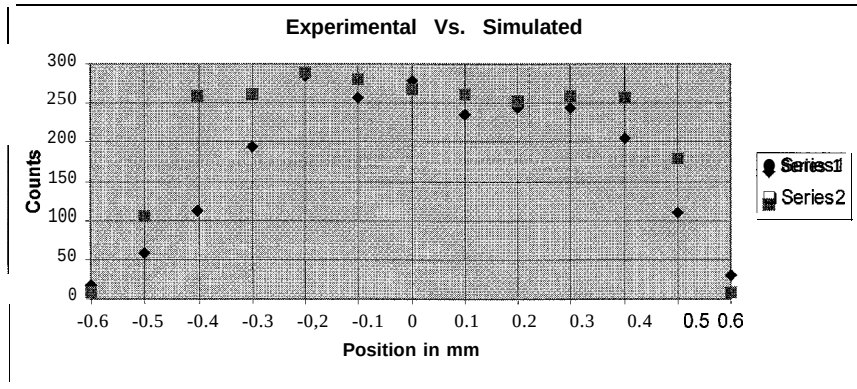


Plot of Experimental Results

A run without a source was performed to establish a background count. The background maximum was then subtracted from the maximum channel count to establish a final value for each position. This final value was then used to generate the graph by plotting the final value against its position. Another run was performed with just the source in front of the detector with out the collimator. The count rate for this run was approximately 1 million counts. Since the source is isotropic, the total activity is roughly double the detected count, so the total activity of just the source was approximately 2 million counts for the total run time.

VI. Conclusion

To compare the results of the simulation and the experiment the data had to be normalized to the same activity of the source. Since the total activity of the experimental data was double that of the simulated data, the simulated data was doubled to give both the experimental and simulated runs the same activity. The plot below is the result of the simulation and experimental data plated together.



Plot of Experimental and Simulated Results

The ends and the middle show a high correlation between the simulated and experimental data. The points just inside the collimator hole have a lot of difference. This difference might be because of the fact that in the experimental runs the source is not truly a point source. Therefore some of the source may have still been covered up by septa at these point which would prevent some of the photons from entering the hole. This is not the case for the simulation because it truly is a point source. This could account for the higher counts seen in the simulation near the septa walls, Beside for this fact, the results show that the simulated and experimental runs provide similar result. This shows that the NCMP Monte Carlo code is a valid tool for the use for designing square hole collimators,

References

1. Anger H. O.; Scintillation camera with multichannel collimators. J Nucl Med 5:515-531, 1964

2. Anger H. O.; Scintillation camera. Rev Sci Inst 29: 27, 1958

3. Beck R. N., Redtung L. D.; Collimator design using ray-tracing techniques, IEEE Trans Nucl Sci NS 32:865-869, 1985

4. Raeside D. E.; Monte Carlo principles and applications. Phys Med Biol. 21:181-197, 1976

Incompressible Turbulent Wing-Body Junction Flow

R. Krishnamurthy, Corey D. Cagle, S. Chandra
NASA CAR, College of Engineering
North Carolina A&T State University
Greensboro, NC 27411

Introduction

The overall objective of this study is to contribute to the optimized design of fan bypass systems in advanced turbofan engines. Increasing the engine bypass ratios have provided a major boost in engine performance improvement over the last fifty years. An engine with high bypass ratio (11-16:1) such as the Advanced Ducted Propulsor (ADP) is being developed and is expected to provide an additional 25% improvement in overall efficiency over the early turbofans. Such significant improvements in overall efficiency would reduce the cost per seat mile, which is a major government and Industry challenge for the 21 st century [1]. The research is part of the Advanced Subsonic Technology (AST) program that involves a NASA, U.S. Industry and FAA partnership with the goal of a safe and highly productive global air transportation system.

The immediate objective of the study is to perform numerical simulation of duct-strut interactions to elucidate the loss mechanisms associated with this configuration that is typical of advanced turbofan engines such as ADP. However, at present experimental data for a duct-strut configuration are not available. Thus, as a fist step a wing-body junction flow would be studied and is the specific objective of the present study. At the outset it is to be recognized that while duct-strut interaction flow is similar to that of wing-body junction flows, there are some differences owing to the presence of a wall at both ends of the strut. Likewise, some differences are due to the sheared inflow (as opposed to a uniform inflow) velocity profile. It is however expected that some features of a wing-body junction flow would persist. Next, some of the salient aspects of the complex flow near a wing-body junction, as revealed by various studies reported in the literature will be reviewed.

One of the principle characteristics of the juncture flow, is the presence of the mean flow components in a plane perpendicular to the direction of the oncoming free-stream flow. The lateral curvature of the wing/strut causes the oncoming turbulent layer to skew about an axis (x-axis) parallel to the plane (xz-plane) of the mean shear. This is the principle mechanism for the generation of secondary flow. As noted by Shabaka and Bradshaw [2], such skew-induced secondary flows are slow to be attenuated by Reynolds stresses. Additional contribution to the generation of secondary flow comes from anisotropies in Reynolds stresses. Upstream of the strut, the mean-vorticity is directed span wise (along the y-direction). The presence of secondary flow in the vicinity of the strut causes the vorticity to stretch around the obstacle in a horse-shoe shape, with each leg having a vorticity of the opposite sense. The blockage effect of the strut imposes a severe adverse pressure gradient on the oncoming turbulent shear layer, causing boundary layer separation ahead of the leading edge, resulting in a vortex that rolls up and flows downstream into the juncture [3] region (fig. 1).

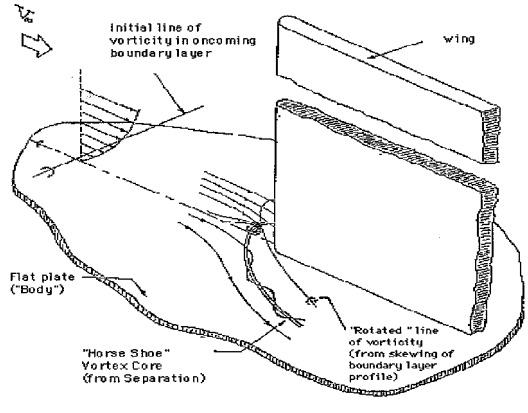


Figure 1. Schematic of three dimensional flow in the juncture

The separation vortices trailing in the wake of the wing can alter the lift or drag characteristics of the surfaces downstream of the wing-body juncture [4]. Likewise, on submarines, the wake flow behind the appendage can

degrade the performance of the propeller located downstream. The complex nature of this flow is caused by the presence of all six components of Reynolds stresses. Devenport and Simpson [5] report that in the vicinity of the horse-shoe vortex there is intense recirculation with turbulent stresses being much larger than those normally observed in turbulent flows. These features contribute to making this flow a challenge to predict numerically.

Some of the past studies provide useful insights into this flow that would guide our numerical efforts. In measurements reported by Shabaka and Bradshaw [2], the eddy viscosity tensor is seen to be non-isotropic and has negative components in certain regions. In an effort to evaluate the closure assumptions of various turbulence models, Devenport and Simpson [6] used their own extensive measurements in juncture flows around the nose of a wing-body junction. Measured values of mean-velocity and /or turbulence kinetic energy was used to predict the magnitude of the shear stress vector. Algebraic stress models performed the best followed by Cebeci-Smith eddy viscosity model. The flow is reported to be dominated by a pressure field produced by the wing and the velocity field generated by the horseshoe vortex that is wrapped around the junction between the wing and wall.

Kubendran et al. [3] conclude from an experimental study that the shape of leading edge of the wing as characterized by its slenderness ratio is a major factor in determining the flow fields in the juncture region. The more thinner the leading edge of the juncture, the weaker the horseshoe vortex is. Also, with a slender leading edge, the secondary flow in the juncture would be due mainly to the cross-stream gradients of Reynolds stresses rather than due to a lateral skewing of the shear layer.

The Experiments

These experiments were performed, in a low speed wind tunnel at Georgia Institute of Technology [7]. The wind tunnel is an open return type with continuously variable flow speeds up to 22.9 m/s (75 ft/s). The size of the test section is 1.07 X 1.09 X 6.10 m (42 X 43 X 240 in). The wing was mounted perpendicular to a flat plate body and aligned with the wind tunnel axis. The wing has a elliptical leading edge and an afterbody with a constant thickness D, of 57.9 mm (2.28 in) (which is used to non-dimensionalize all the spatial coordinates) and a length of 1.22 m (48 in) [7]. The experiments were performed at a nominal free stream velocity of 15 m/s corresponding to a Reynolds number of 984,000/m and a Mach number of 0.04. The flat-plate boundary layer at the wing leading edge location in the absence of the wing was approximately 23 mm thick, corresponding to a ratio of wing thickness to boundary layer thickness of 2.5 [3].

Numerical Simulation

The code NPARC [8] was used in the simulations. It is a general purpose CFD tool that is applicable to a wide variety of aerospace design and analysis problems. Since the flow to be simulated involved a three-dimensional separating flow, the complete Navier Stokes equations were used [9]. A three dimensional multi-bock grid has been used in the computations. Initial computations were performed with the Baldwin Lomax algebraic turbulence model and the results were used to initiate computations with the two-equation k-epsilon model. The Low Reynolds number k-epsilon model of Chien [10] was used in this study. All the computations reported here were performed on the Cary C-90 system at the NAS facility located in NASA Ames Research Center.

Results and Discussion

Initially two dimensional simulations of incompressible, turbulent flow pasta flat plate were performed. The flow conditions chosen were identical to those for the three-dimensional simulation. The specific objective was to ensure that the numerical simulations are able to replicate some of the experimental data that are basically of a two dimensional nature. Also, converged solutions from such a two-dimensional simulation would be useful in generating initial profiles for three-dimensional simulations. The two dimensional results were computed with a grid that had 50 points in the z-direction and 126 points in the x-direction with the length X_{max} of the plate being 52D and the Z_{max} being 3 .5D. Here D (58mm) is the thickness of the wing. As a first step in this validation process, the boundary layer thickness, δ_{99}, was computed at the position where the leading edge of the wing-body would be located. This was 23.5 mm. This value is in agreement with the measurement reported by Kubendran et al.[3].

The two-dimensional mean velocity and turbulent kinetic energy data were compared with experimental results, at a location upstream of the leading edge of the wing-body, where the flow is expected to be predominantly two-dimensional. The two-dimensional mean velocity profiles were also compared with the one-seventh power law. The velocity profile comparison has good agreement with the experimental data and the one-seventh power law. This indicates that the mean flow is essentially two-dimensional at this location. The kinetic energy comparisons are reasonable with the experiment at this location. However the spread of the boundary layer (as implied by the measured turbulence kinetic energy) was not captured by the simulations. All the simulations have been performed at a Mach number of 0.1 which is higher than the one in the experiments, but is still within the incompressible

regime. This was done to enhance the speed of computations. The k-epsilon model is not a self starting model, so reasonable initial profiles had to be provided. Therefore, the computations were initially done with Baldwin-Lomax turbulence model up to 2000 iterations after which the k-epsilon model was initiated.

Comparisons are shown in figures 2(a-d) between the numerically computed and measured mean velocity profiles at various streamwise locations and at various distances in the y-direction from the centerline of the wing-body junction. At 102 mm upstream(Fig. 2a), the flow is still predominantly two-dimensional. Therefore, the results shown are not affected much by the presence of the wing-body juncture. The agreement is good near the edge of the boundary layer, but is reasonable near the wall. The numerical data shows the same trend as the experimental profiles. At locations progressively closer to the junction (figs. 2b , 2c), the effect of the presence of the junction causes the agreement between measurements and computations to deteriorate. At a location 76 mm downstream of the junction, the computations and experiments do not agree with each other. This is primarily due to the fact that at this location the flow is highly separated and the k-epsilon turbulence model used here is not expected to yield accurate results nuder such conditions.

In figures 3a and 3b a comparison between measured and computed turbulence kinetic energy profiles are shown. The peak value is captured by the computations, but not the spread in the direction normal to the body. This matter is currently being investigated farther. In fig. 4 is shown a comparison between measured and computed profiles for the third component of the velocity field, U_y. The flow develops this component when it encounters the junction and this phenomenon is termed ***skewing.*** As can be seen the agreement between measurement and computations is good. In fig. 5 a velocity vector plot is shown illustrating the presence of a vortex in the flow and also a region of recirculating flow indicating the presence of separation. This vortex is termed the horse-shoe vortex and is one of the principal features of this complex flow. Further results and discussion are given in Cagle [9].

Acknowledgment

The support provided by the grant NAG3- 1734 from NASA Lewis Research Center is gratefully acknowledged.

References

[1] Koff. B. F., "Spanning the Globe with Jet Propulsion," AIAA Paper 91-2987, April 1991.

[2] Shabaka, I. M. M. A., Bradshaw, P., "Turbulent Flow Measurements in an idealized Wing/Body Junction", AIAA Journal, Vol. 19, pp. 131-132, No. 2, February 1981.

[3] Kubendran, L. R., McMahon, H. M. , "Turbulent Flow Around a Wing/Fuselage Type Juncture," AIAA Journal, Vol. 24, pp. 1447-1452, No. 9, September 1986.

[4] Hawthorne, W. R., "The Secondary Flow About Struts and Airfoils," Journal of Aeronautical Sciences, pp. 588-608, Sept. 1954.

[5] Devenport, J. W., Simpson, R. L., "Time-dependent and Time-Averaged Turbulence Structure Near the Nose of a Wing-Body Junction," Journal of Fluid Mechanics, Vol. 210, pp. 23-55, 1990.

[6] Devenport, J. W., Simpson, R. L., "Flow Past a Wing-Body Junction-Experimental Evaluation of Turbulence Models," AIAA Journal, Vol, 30, pp. 873-881, No. 4 April 1992.

[7] McMahon, H., Hubbartt, J., and Kubendran, L. R., "Mean Velocities and Reynolds Stresses Upstream of a Simulated Wing-Fuselage Type Juncture," NASA CR 3695, June 1983.

[8] Cooper, G. K., Jones, R. R. III, Power, G. D., Sirbaugh, J. R., Smith, C. F. III, Towne, C. E., NPARC User's Guide, Version 2.0, NPARC Alliance, November 1, 1994.

[9] Cagle, Corey, D., Numerical Simulation of Wing-Body Flows, M.S. Thesis, North Carolina A&T State University, 1997.

[10] Chien, K, "Predictions of Channel and Boundary-Layer Flows with a Low-Reynolds -Number Turbulence Model," AIAA journal Vol. 20. NO. 1 January 1982.

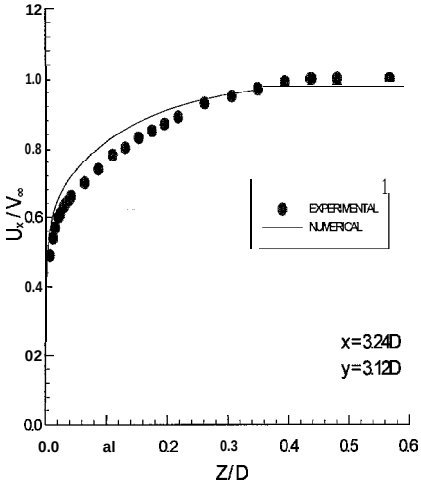


Figure 2a. A comparison of mean velocity profiles 102mm upstream of junction

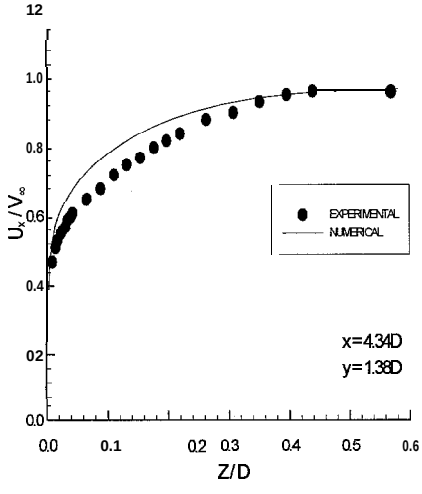


Figure 2b. A comparison of mean velocity profiles 38mm upstream of junction.

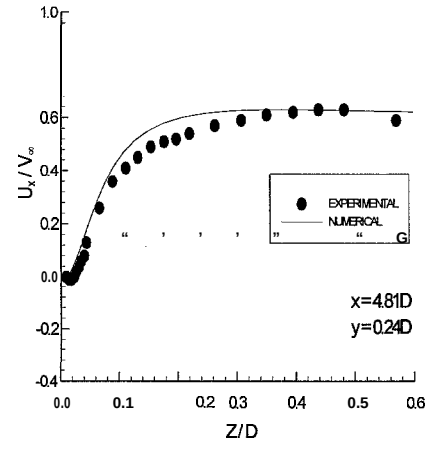


Figure 2c. A comparison of mean velocity profiles 111mm upstream of junction.

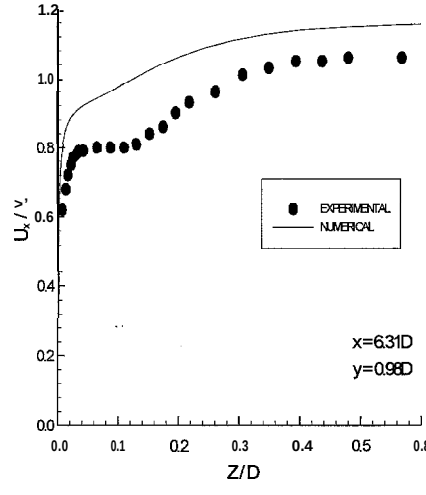


Figure 2d. A comparison of mean velocity profiles 76 mm downstream of junction.

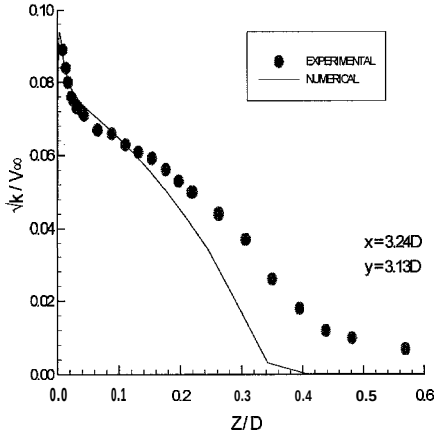


Figure 3a. A comparison of profiles of turbulence kinetic energy 102mm upstream of junction.

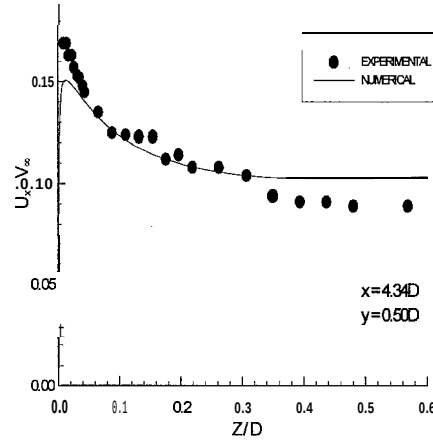


Figure 4. A comparison of profiles of mean Velocity U_y, 38mm upstream of junction.

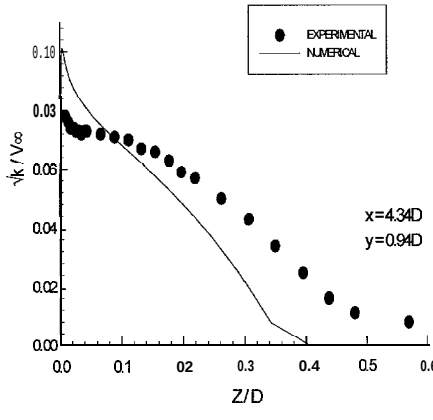


Figure 3b. A comparison of profiles of turbulence kinetic energy 38mm upstream of junction.

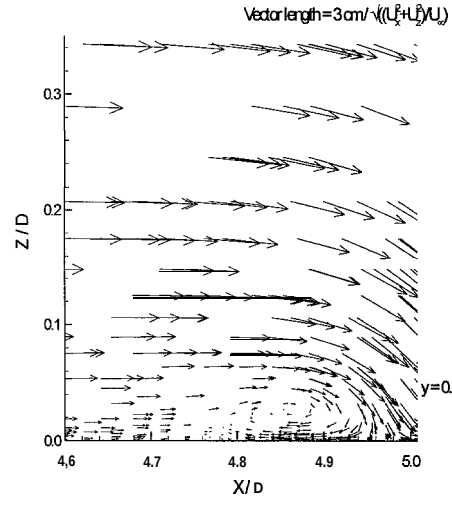


Figure 5. A velocity vector plot showing the presence of a vortex

98 URC056

Near lossless Iterated Block Matching Fractals Image Compression

Hamed Parsiani,
Electrical & Computer Engineering Dept.
University of Puerto Rico
Mayaguez, Puerto Rico, 00681
parsiani@exodo.upr.clu.edu

Paul G. Roman Ortiz
Electrical & Computer Engineering Dept.
University of Puerto Rico
Mayaguez, Puerto Rico, 00681

Abstract- The Iterated Block Matching Fractal encoding process exploits the fractal similarities within an image by coding the image using itself. A best match is obtained for every block to be coded from the pool of blocks generated from the image itself. The matching parameters are transmitted resulting in compression. This method is implemented with its most recent advances. Its encoder/decoder algorithm is complete with Variable Length encoder/decoder and pack-unpack algorithms. Test results were obtained for near lossless compression at high PSNR values.

I. INTRODUCTION

The Fractals as conceived by Barnsley [1] [2], finds the basic features in an image, creates mathematical models of them, then defies the relations between pieces of the image and the basic features. One needs to know well the relationship between "formulas" for transformations and the resultant stretching, twisting, folding and skewing affecting the underlying image. This method proved to be a hardly acceptable approach to real time compression, as it required an expert to thoroughly analyze one image at a time, find the features, determine the respective "formulas", and then proceed to compression.

Arnaud Jacquin's work [3] on the principle of block matching does not rely on modeling of features by formulas. The method of Iterated Block Matching as introduced by Jaquin is briefly explained first. In this paper, this method is referred to as conventional method. The conventional method was modified to produce improved results[4]. The improved encoder/decoder is explained in sections III and IV. The improved encoder/decoder was completed as a stand alone product including both Huffman and pack-unpack algorithms. This product was tested using high altitude images as explained in section V.

The Conventional Method

In the encoding process, the image fractal similarities are exploited by coding the image using itself. In this method, the image is broken down into equal size blocks (2Bx2B) called domains. Then an image block (BxB) to be encoded (called range block) is compared for the best match with each domain block (contracted to size BxB) from the domain pool. The matching parameters obtained are contrast scalings and luminance shifts. To increase the probability of finding a good match, different orientations (or isometrics) of the domain blocks are created which produce a large pool of domain blocks with more varieties.

A. Contrast Scale Factor, Luminance Shift, and Approximate Range Equations

let Im = Image,

D_j = j^{*} domain block, R_j = j^{th} range block

let $\mathrm{Im}|_{D_j}\ j^{th}$ domain block

$\mathrm{Im}|_{SoD_j}\ j^{th}$ contracted domain block

$\mathrm{Im}|_{\hat{R}_i}\ j^{th}$ approximate range block

$\overline{Im}|_{D_j}$ Mean of D_j

$Im|_{\hat{R}_i^{(n)}}$ i^{th} estimated range block at n^{th} iteration

$\overline{Im}|_R$ Mean of R_i

$dr|_R \triangleq Im|_{\hat{R}_i} - \overline{Im}|_R$ dynamic range of R_i

$dr|_{D_j} \triangleq Im|_{D_j} - \overline{Im}|_{D_j}$ dynamic range of D_j

Contrast Scale Factor $\alpha_i \triangleq \frac{dr|_{R_i}}{dr|_{D_j}}$

Luminance Shift $\Delta g_i \triangleq \overline{Im}|_{R_i} - \alpha_i \overline{Im}|_{D_j}$

Approx. Range is $Im|_{\hat{R}} = \alpha_i Im|_{SoD_j} + \Delta g_i$

The coding of an image is done block by block in a raster scan fashion. Each block (a range) is coded by searching the pool of domains for the best match. The search time could be reduced by first classifying the image into blocks of the same type. Three block types are defined as: shade blocks , midrange blocks, and edge blocks. The domain pool is generated by sliding a window of size DxD (D=2B) across the original image in steps of B/2. Then the blocks are classified accordingly.

During the encoding procedure the image is scanned one block (range) at a time and the range block R_i to be encoded is compared against all domain blocks of the same type and the best match D_j is obtained. From the best match the desired parameters α_i and Δg_i are calculated. The criteria for the best match is a distortion measure (RMS value) calculated for each $Im|_{\hat{R}_i}$ and $Im|_{R_i}$, where

$$Im|_{\hat{R}} = \alpha_i Im|_{SoD_j} + \Delta g_i .$$

B. Parameters Transmitted:

Shade Blocks:

Average gray level, $\overline{Im}|_{R_i}$, and the block's location (x,y) are transmitted.

Midrange and Edge Blocks:

In the process of finding the best match, the variables saved are the type of the block, the location of the best match domain block, (x,y), the contrast scaling factor, α_i, the luminance shift, Δg_i, and the type of the best found isometry, Ii An isometry is defined by any of the seven possible rotations and flips (about vertical, horizontal or diagonal axis of image).

C. ***Decoding Procedure***

1) Initialize the entire image plane to some value or an unrelated image.

The variables received are used for the decoding purposes. If the block type variable indicates a shade block then the uniform value of $\overline{Im}|_{R_i}$ fills a block at the (x,y) address. If the block type is different then part (2) is implemented.

2) Locate the matched block using the domain (x, y) address. Apply the affine transformation on this domain block to calculate range block. (i.e. 1 't iteration decoded Range block= $\alpha_i Im|_{S \circ D_j} + \Delta g_i$) Do this operation on all of the image blocks.

3) Repeat this process and then measure the change in the image from the previous iteration. If this change is zero or within a negligible threshold value, then stop, otherwise continue the iteration.

II. LIMITATIONS OF *a*, *IN* THE CONVENTIONAL ENCODER

The conventional encoding of image originated by Jacquin and continued by subsequent researchers requires the values of contrast scale factors α_i to be less than 1. This requirement is necessary since it is possible that the individual transformation of some domain blocks would not be contractile (convergent) at the decoder for some $\alpha_i \geq 1$. This can be simply shown by actually doing several iterations mathematically.

III. IMPROVED ENCODER

In this work[4], two improvements were made at the encoder.

1) The number of bits to be transmitted by the encoder is substantially reduced by transmitting the mean of each range block instead of Δg_i. The location coordinates for the shades are not transmitted, and a code of only one bit is transmitted to distinguish between two classes of midrange and edge blocks.

The encoding of midrange blocks will only require the transmission of α_i, and the location of the matched domain (x,y). The encoding of edge blocks is the same as the midrange blocks with an additional coding of the type of isometry.

2) All domains are candidates for the block matching algorithm. This allows a higher quality image encoding.

IV. IMPROVED DECODER

The transmission of range block-mean $\overline{\text{Im}}|_{R_i}$, instead of Luminance shift, Δg_i, permits several important improvements at the decoder

l) Contrast factor α_i is free to assume any values at the encoder with a guarantee of block convergence after n iterations at the decoder, where n is any value. It has been shown mathematically [4] that the estimated range block after n iterations, $\text{Im}|_{\hat{R}_i^{(n)}}$, remains convergent for all values of α_i and n.

2) At the first iteration a low pass version of the image is available. And, image converges at less number of iterations than the conventional method.

V. TEST OF IMPROVED ENCODER/DECODER AS A PRODUCT

The improved Encoder/Decoder was modified by including a Huffman encoder/decoder and a pack-unpack algorithm. The specifications assumed for encoder/decoder are: Image size = 512x512, and block size = 4x4. The number of bits used for parameters are given in Table 1.

TABLE 1
PARAMETER CODING

Parameters	Bits
$\overline{\text{Im}}\|_{R_i}$	5
type	1
α	6
(x,y)	12
Ii	3

This stand alone product was tested, Table 2, using high altitude images of Mayaguez, Añasco Puerto Rico and the image of Lenna.

TABLE 2
TEST OF COMPRESSION PRODUCT

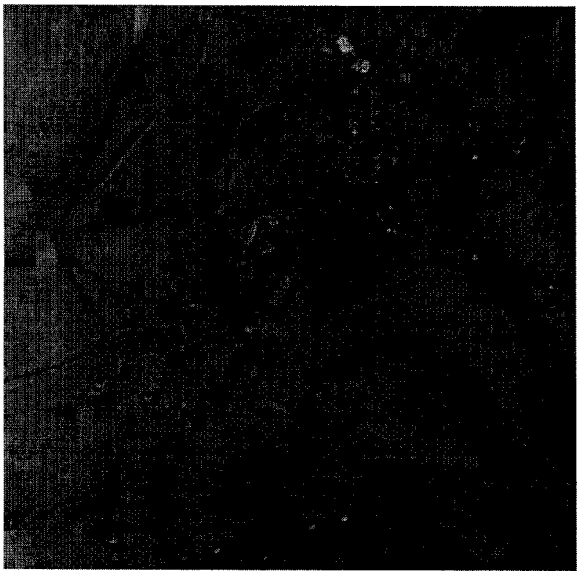
	Previous Algorithm [4]	Previous Algorithm [4] with Pack-Unpack	Present Algorithm, with Huffman encoder/decoder and Pack-Unpack		
Image	PSNR	Compression Ratio	Compression Ratio,	PSNR	SEE FIGURES
Añasco, P.R.	37.6	7.06	9.6	34.8	1(a,b)
Mayaguez, P.R.	34.9	6.26	6.7	33.1	2(a,b)
Lenna	31.1	10.88	11.7	30.2	3(a,b)

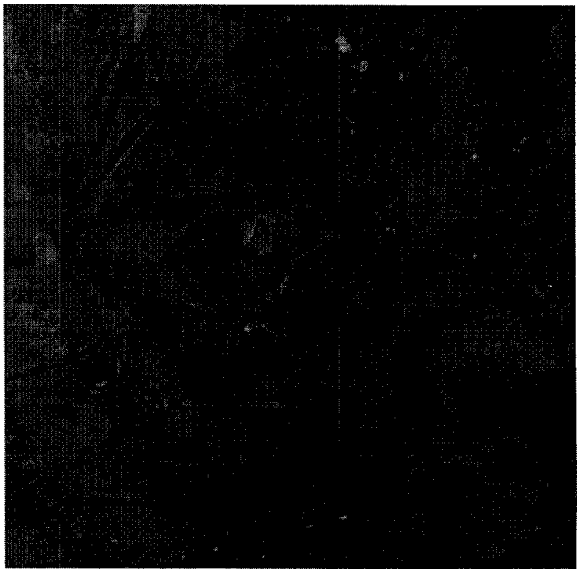
VI. CONCLUSION

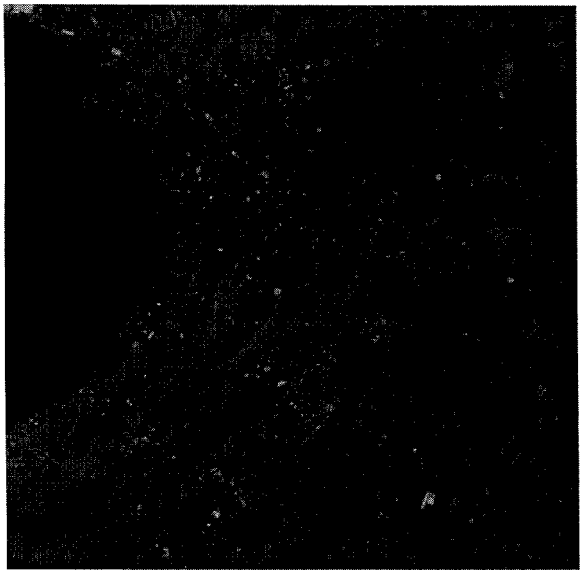
The improved version of Iterated Block Matching Fractals was complemented with Huffman encoder/decoder and pack-unpack algorithms to make it a stand alone product. This product showed a slight improvement in compression ratios over the previous algorithm[4], as shown in table 2. The decompressed images of Mayaguez and Añasco were submitted to an image analysis software (Khoros) which produced the same result as when the original images of the same cities were analyzed by Khoros. The psychovisual analysis of the image also shows a near lossless compression result.

REFERENCES

[1] M. F. Barnsley and S. Demko, "Iterated function systems, and global construction of fractals," Proc. Roy. Sot. London, vol. A399, pp. 243-275, 1985.
[2] M. F. Barnsley, "Fractals Everywhere". New York, Academic Press, 1988.
[3] A. E. Jacquin, "Fractal Image Coding: A Review," Proceedings of the IEEE, vol. 81, No. 10, Oct. 1993.
[4] H. Parsiani and R. Garcia, "State of the Art Iterated Block Matching Fractals in Image Compression", Proceedings of MWSCAS'97, Aug. 3-6, 1997.







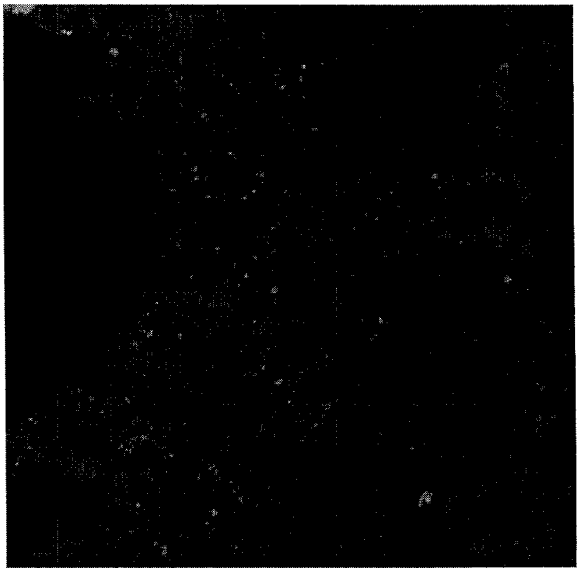


Fig. 3(a) Lenna Original

Fig. 3(b) Lenna 4x4
Improved Encoder/Decoder 3rd Iteration, PSNR=30.2

98 URC057

SUFFICIENT CONDITIONS FOR THE STABILITY OF A LUR'E SYSTEM WITH TWO NONZERO INPUTS BASED ON LYAPUNOVS DIRECT METHOD

Olawale Adetona*, Sivapragasm Sathananthan*,
Ephrahim Garcia+ and Lee H. Keel*.

*NASA Center for Automated Space Science
Tennessee State University
Nashville, TN 37203-3401.
+Mechanical Engineering Department
Vanderbilt University
Nashville, TN 37235

ABSTRACT

The Lur'e problem is a famous problem in nonlinear control theory. The analysis of this system is usually restricted to Lur'e systems in which the external input is zero. One form of this problem involves the stability analysis of the Lur'e system with two nonzero inputs. The stability analysis of this type of system is traditionally performed based on L stability theory. The objective of this paper is to present sufficient conditions for the stability of this system based on Lyapunov's direct method. This is a conceptually simpler approach to the existing analysis based on L stability theory and it provides some additional information about the stability of the system.

KEYWORDS: Lur'e, Popov, Lyapunov, Stability, Nonlinear System.

INTRODUCTION

Lur'e and Postnikov proposed the study of the stability of the type of systems shown in Figure 1. The operator G(s) is assumed to be a stable linear time invariant transfer function all poles of which are in the left half of the *s* plane. The function $f(\sigma, t)$ is generally a real, continuos, time varying, single-valued scalar function contained in an angle $0 \le \frac{f(\sigma,t)}{\sigma} \le k$ (i.e. sector *k* of Figure 2). This function is assumed to lie within the first and third quadrants of the σ, $f(\sigma, t)$ plane. Given the nonlinear gain $f(\sigma,t)$, the objective of the analysis is to determine the conditions on the stable transfer function G(s) and the rate of time variation of the nonlinear function $\dot{f}(\sigma,t)$ that are sufficient to guarantee the absolute stability of the system. In our analysis $\dot{f}(\sigma,t)$ is assumed to be identically zero.

The Popov Criterion

In 1959 V.M Popov proposed a frequency domain criterion for absolute stability of a class of nonlinear systems. The Popov criterion has a geometric interpretation which

makes it easier to use than the direct method of Lyapunov. The absolute stability criterion due to Popov is stated below.

Theorem 1: If the closed loop system of Figure 1 consists of a stable linear part with a transfer function G(s), all poles of which are in the left half of the s plane and a single valued, time invariant, non-linear function $f(a)$, where $\frac{f(\sigma)}{\sigma} \in [0,/\%]$ for $\sigma \in (-\infty,+\infty)$. Given that $\alpha \equiv$ O, a sufficient condition for stability of the closed loop system is:

$$\mathrm{Re}\left[(1 + j\delta\omega)\cdot G(j\omega)\right] + \frac{1}{k} > 0, \text{ for } \omega \in (\mathrm{O}, \mathrm{CO}) \quad (1)$$

where δ is an arbitrary real number, $k < co$, and $\lim_{\omega\to\infty} G(j\omega) =$ O.

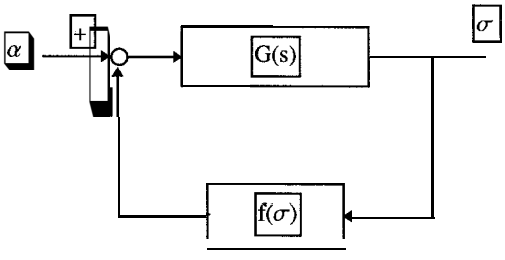


Figure 1: Lur'e System

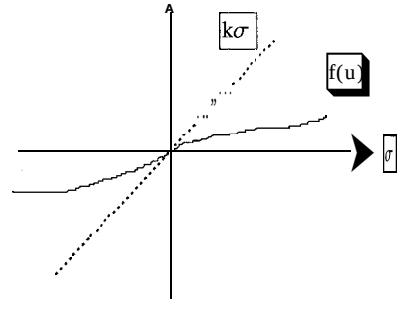


Figure 2: Sector Bounded Nonlinear Gain

Geometric Interpretation Of The Popov Criterion

Let $G(j\omega) \equiv u(\omega) + jv(\omega)$. (2)

Therefore, from Equation (1):

$$\mathrm{Re}\left[(1 + j\delta\omega)\cdot(u(\omega) + jv(\omega))\right] + \frac{1}{k} > 0 \quad (3)$$

implies

$$u(\omega) - \delta\omega v(\omega) + \frac{1}{k} > 0. \quad (4)$$

Consider the following transformed frequency response $G_t(\omega)$:

$$G_t(\omega) := u(\omega) + jv_t(\omega), \tag{5}$$

$$v_t(\omega) := \omega \cdot v(\omega). \tag{6}$$

Also consider the straight line $r(\omega)$ in the Popov plane (u(a), $j\omega \cdot v(\omega)$ plane):

$$r(\omega) \equiv u(\omega) + jv_o(\omega), \tag{7}$$

$$v_o(\omega) := \frac{1}{\delta} u(\omega) + \frac{1}{k} \tag{8}$$

Therefore, the Popov criterion requires that the Locus $G_t(\omega)$ must be to the right of the straight line $r(\omega)$. Note that $r(\omega)$ has an intercept of $\frac{1}{k}$ on the $u(\omega)$ axis and a gradient of $\frac{1}{\delta}$. Finally, the absolute stability criterion can not be used directly if the linear part is neutral or unstable. If the linear part is unstable, the system will be unstable with any non-linear element.

In the standard form, the Popov criterion can only be used to prove the stability of systems where the nonlinear function (gain) is in the first and third quadrants. These nonlinear functions have one horizontal axis intercept at the origin. This work is presented to broaden the class of nonlinear gains $f(o)$ to include nonlinear functions that pass through the second or fourth quadrants in addition to the first and third quadrants. This class of nonlinear functions have a unique horizontal axis intercept at $\sigma = -\alpha_2$, where $\alpha_2 \in (-\infty, +\infty)$.

PRELIMINARIES

Asymptotic Stability

The concept of stability introduced by A.M Lyapunov can be defined as follows:

Stability: The origin is a *stable* equilibrium point of the state x of a dynamic system, if for every real $\varepsilon > 0$ and initial time t_0 there exists a real $\delta(\varepsilon, t_0) > 0$ such that $\|\mathbf{x}(t_0)\| \leq \delta$ implies $\|\mathbf{x}(t)\| \leq \varepsilon$ for all $t \geq t_0$.

Attractivity: The origin is an *attractive* point of a dynamic system, if for every real $\eta > 0$, some real $p > 0$ and initial time t_0, there exists a number $T(\eta, \mathbf{x}(t_0), t_0)$ such that for all $\|\mathbf{x}(t_0)\| \leq p$, $t \geq to + T$ implies $\|\mathbf{x}(t)\| \leq q$.

Asymptotic Stability: The origin is an *asymptotically stable* equilibrium point of a dynamic system if it is both stable and attractive, i.e. for all real ε >0, initial time t_0 and time $t \geq t_0$ there exists a constant $_{\varepsilon} > 0$ such that $\|\mathbf{x}(t)\| \leq \varepsilon$ for all $\mathbf{x}(t_0)\| \leq \delta$ and $\lim_{\mathbf{x}(t)\to\infty}$ x = O.

Note that an attractive system can be unstable because the concepts of stability and attractivity are independent.

Absolute Stability

A system is said to be absolutely stable if it is asymptotically stable regardless of the size of the initial region δ. The concept of absolute stability is based on the direct method of Lyapunov. The beauty of the direct method of Lyapunov lies in the fact that it enables the stability of a system to be determined from the equations that define the system without having to explicitly solve these equations. The direct method of Lyapunov is used to determine the stability of motion of a nonlinear system by constructing an "energy-like" function v for the system and studying the total time derivative of this function.

EXTENDED POPOV CRITERION

Theorem 2: If the closed loop system of Figure 1 consists of a stable linear part with a transfer function G(s), all poles of which are in the left half of the *s* plane and a single valued, time invariant, non-linear function $f(\sigma)$ where: $f(0) = \alpha_1$ and $\frac{f(\sigma) - a,}{\sigma} \in [0, k]$ for $\sigma \in (-\infty, +\infty)$. In addition, ~(t)= α_0 and $f(\alpha_2(t)) = O$ respectively, for all $t \geq t_0$ where t_0 and α_0 are an arbitrary pair of finite, real constant values. Given:

x(t) = $A\mathbf{x} + \mathbf{b}\tau$, $\quad \mathbf{x} \in \Re^{n\times 1}$, $A \in \Re^{n\times n}$, $\mathbf{b} \in \Re^{n\times 1}$, $\tau \in \Re^{1\times 1}$

$\sigma = \mathbf{h}^T\mathbf{x} + \rho\tau$, $\quad \sigma \in \Re^{1\times 1}$, $\mathbf{h}^T \in \Re^{1\times n}$, $\mathbf{p} \bullet \Re^{1\times 1}$

$\tau = \alpha_1 - f(\sigma)$, $\quad f : \Re^1 \to \Re^1$, $f(\sigma)$ =0 if and only if $\sigma(t) = -\alpha_2(t)$,

and the transfer function G(s) is defined as

$$G(s) := \frac{L[\sigma]}{L[\tau]} = \rho + \mathbf{h}^T[sI - A]^{-1}\mathbf{b}, \qquad \textbf{(9)}$$

where $\mathbf{h}^T, A, \mathbf{b}$ are of the phase variable canonical form, a sufficient condition for stability of the closed loop system is:

$$\text{Re}[(1 + \text{j\&n}).\ G(j\omega)] + \frac{1}{k} > 0,\ \omega \bullet (\text{O, CO}), \qquad \text{(lo)}$$

α_i(t) < co, α_2(t) <∞ and $\dot{\alpha}_2(t) < \infty$ for all $t \geq 0$, respectively, where δ is an arbitrary real number, $k \in (O, \infty)$ and $\lim_{\omega\to\infty} G(j\omega)$ = O (i.e. $p = O$).

Note that when al(t) ≡ O and ~(t)= O for all $t \geq 0$, respectively, we have the standard Popov criterion for the standard Lur'e system.

Summary of the Proof

For a specified transfer function G(s) and specified signals al(t) and $\alpha_2(t)$ respectively, the system described in Theorem 2 can be defined in the form of the system shown in Figure 3. This can be achieved if ~(d):= $f(\sigma)$ where $\hat{f}(\hat{\sigma}) \in [0, k]$, $\hat{\sigma}(t) := \sigma(t) + \alpha_2(t)$ and $\hat{f}(\hat{\sigma}) = O$ if and only if $\hat{\sigma} = O$.

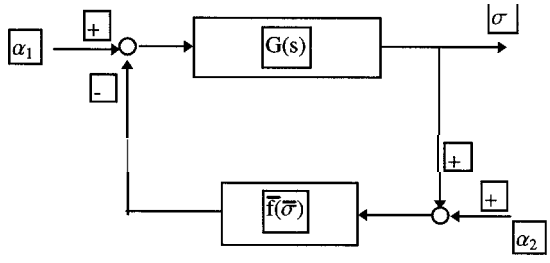


Figure 3: Extended Lur'e System

Note that:

1) For all $t \geq 0$, $f(\sigma) = O$ if and only if $\sigma(t) = -\alpha_2(t)$ implies ~(d)= O if and only if $\hat{\sigma} = 0$.
2) For all $t \geq 0$, $f(0) = \alpha_1(t)$ implies $\hat{f}(\alpha_2(t)) = \alpha_1(t)$. This implies ~(t)= α_0 for all $t \geq t_0$, where $t_0 < \infty$ and $\alpha_1(t) < \infty$ for all $t \geq 0$. Also, $\hat{f}(-\alpha_2(t)) = 0$ for all $t \geq t_0$, where $t_0 < \infty$.
3) Note that, $\alpha_2(t) < \infty$ and $\dot{\alpha}_2(t) < \infty$, respectively for all $t \geq 0$.

Sufficient conditions for bounded-input-bounded-output stability of the system of Figure 3 have been obtained based on L_2 stability theory and are similar to the conditions specified in Theorem 2 (Vidyasagar 1993, p. 354).

CONCLUDING REMARKS

Sufficient conditions for absolute stability of Lur'e systems that contain nonlinear functions that are not restricted to the first and third quadrants were presented. This type of system can be reformulated as a Lur' e system with two nonzero inputs. It was observed that the sufficient conditions for the absolute stability of this type of systems are similar

to known sufficient conditions for bounded-input-bounded-output (BIBO) stability of these systems. The asymptotic stability results that were presented compliment the known BIBO stability results by providing additional information about the qualitative behavior of the system if the conditions specified in Theorem 2 are satisfied. Furthermore, the approach presented in this paper is conceptually simpler than the existing results based on L_2 stability theory.

ACKNOWLEDGEMENTS

This research is supported in part by NASA Grant NCC5-228. This support is greatly appreciated.

REFERENCES

1) K.S. Narendra, and J.H. Taylor, Frequency Domain Criteria for Absolute Stability, Academic Press, 1973.
2) A. Netushill, Theory of Automatic Control, Mir Publishers, 1978.
3) V.M. Popov, Hyperstability of Control Systems, Springer-Verlag, 1973.
4) M. Vidyasagar, Nonlinear Systems Analysis, Prentice-Hall, 1993.

Acceleration of Logarithmic Convergence

J. G. Gaskin, SUNY College at Oswego,
and W. F. Ford, NASA, Lewis Research Center.

Introduction. In this paper, we shall give a characterization of all monotonically decreasing sequence of positive terms, whose sum converge and then introduce a Transformation which can be used to accelerate the convergence of a large class of logarithmically convergent series.

Preliminaries. Throughout this note, f shall denote a continuous, positive valued and monotonically decreasing function on $[1,\infty)$ satisfying $\lim_{x\to\infty}\frac{f(x+1)}{f(x)}$ and $\int_1^\infty f(x)dx < \infty$. If μ is a positive real number ≥ 1, then we define the μ^{-th} partial sum $S(\mu)$ of the series $\sum_{n=1}^{\infty} a_n$ by

$$S(\mu) = a_1+a_2+\ldots+a_{[\mu]}+(\mu-[\mu])a_{[\mu]+1},$$

where $[\mu]$ is the greatest integer less than μ. If $\varphi(x)$ is a positive valued function and $\rho\in(0,1)$, we define the $T_{\varphi,\rho}$ Transformation by

$$T_{\varphi,\rho}(S(n)) = \frac{S(n)-(1/\rho)S(\varphi(n))}{1-(1/\rho)}.$$

Finaly, we say that g is logarithmic if g is monotonic on $[\alpha,\infty)$, for some $\alpha\in\Re$, and

$$\lim_{x\to\infty}\frac{g(x)}{g(x+1)} = 1.$$

Theorem 1. Let f be continuous, positive valued and monotonically decreasing on $[1,\infty)$, and let $\rho\in(0,1)$. Then $\sum_{n=1}^{\infty} f(n)$ converges if and only if there exists a differentiable function φ satisfying

$$\varphi(x) > x \text{ and } \varphi'(x)f(\varphi(x)) = \rho f(x).$$

Proof. Let F(x) be the antiderivative of f satisfying $\lim_{x\to\infty} F(x) = 0$, and let $\varphi(x) = F^{-1}(\rho F(x))$. Differentiating $\varphi(x)$, we get $\varphi'(x)f(\varphi(x)) = \rho f(x)$. Since F(x) is increasing on $[1,\infty)$ and $\lim_{x\to\infty} F(x) = 0$, we have that $\lim_{n\to\infty}\varphi(x)=\infty$ and $\varphi(x) > x$.

Next, suppose that $\varphi(x) > x$ and $\varphi'(x)f(\varphi(x)) = \rho f(x)$. Then,

$\frac{\varphi'(x)f(\varphi(x))}{f(x)} = p < 1$. Therefore by [2] (see page 44), $\sum_{n=1}^{\infty} f(n)$ converges.

For the rest of this note, $\varphi(x)$ will be defined as in the proof of Theorem 1.

Theorem 2. Let $S = \sum_{n=1}^{\infty} f(n)$ be a logarithmic series of positive terms and suppose that φ' is logarithmic. Then $T_{\varphi,\rho}(S(n))$ converges more rapidly than $S(\varphi(n))$.

Proof. First, we shall show that $\lim_{n\to\infty} \frac{S(\varphi(n))-S(\varphi(n-1))}{f(n)} = \rho$ To this end, note that

$$\frac{S(\varphi(n))-S(\varphi(n-1))}{f(n)} \leq \frac{(\varphi(n)-\varphi(n-1))f(\varphi(n-1))}{f(n)}$$

$$\leq \left(\frac{\varphi'(\xi_n)f(\varphi(n-1))}{f(n)}\right)$$

$$= \left(\frac{\varphi'(n-1)f(\varphi(n-1))}{f(n-1)}\right)\left(\frac{f(n-1)}{f(n)}\right)\left(\frac{\varphi'(\xi_n)}{\varphi'(n-1)}\right), \text{ where}$$

$\xi_n \in [n-1,n]$. Thus,

$$\overline{\lim}\frac{S(\varphi(n))-S(\varphi(n-1))}{f(n)} \leq \rho$$

Similarly,

$$\frac{S(\varphi(n))-S(\varphi(n-1))}{f(n)} \geq \frac{(\varphi(n)-\varphi(n-1))f(\varphi(n))}{f(n)}$$

$$\geq \left(\frac{\varphi'(\eta_n)f(\varphi(n))}{f(n)}\right) \geq \left(\frac{\varphi'(n)f(\varphi(n))}{f(n)}\right)\left(\frac{\varphi'(\eta_n)}{\varphi'(n)}\right), \text{ where}$$

$\eta_n \in [n-1,n]$. Therefore,

$$\lim\frac{S(\varphi(n))-S(\varphi(n-1))}{f(n)} > \rho \text{ and hence, } \lim_{n\to\infty}\frac{S(\varphi(n))-S(\varphi(n-1))}{f(n)} = \rho$$

Now, $S(\varphi(n))-S(n)\to 0$ and $S(\varphi(n))-S\to 0$ monotonically. Therefore, by [1] (see page 413), and above. we have that:

$$\lim_{n\to\infty}\left(\frac{S(\varphi(n))-S(n)}{S(\varphi(n))-S}\right) = 1 - \lim_{n\to\infty}\left(\frac{f(n)}{S(\varphi(n)-S(\varphi(n-1))}\right) = \left(\frac{\rho-1}{\rho}\right). \text{ Hence,}$$

$$\lim_{n\to\infty}\left(\frac{T_{\varphi,\rho}(S(n))-S}{S(\varphi(n))-S}\right) = 1 - \left(\frac{1}{1-(1/\rho)}\right)\lim_{n\to\infty}\left(\frac{S(\varphi(n))-S(n)}{S(\varphi(n))-S}\right)$$

$$= 1 - \left(\frac{\rho}{\rho-1}\right)\left(\frac{\rho-1}{\rho}\right) = 0.$$

We shall conclude this note with some examples showing how to determine $\varphi(x)$ when $f(x)$ and $\rho \in (O, 1)$ are known.

Example 1. Let $f(x) = \frac{1}{x^2}$. Then, $F(x) = \frac{-1}{x}$ and $F\text{-}1(x) = \frac{-1}{x}$. Therefore,

$\varphi(x) = \frac{x}{\rho}$. If $p = \frac{1}{m}$, where m is a positive integer, then $T_{\varphi,\rho}(S(n)) = \frac{S(n) - mS(mn)}{1-m}$ reduces to the $T_{.m}$ Transformation by Clark and Gray (see [2], page 268, definition 4. 1).

Example 2. Let $f(x) = \frac{1}{x(\ln x)^2}$. Then, $F(x) = \frac{-1}{\ln(x)}$ and $F\text{-}1(x) = e^{\frac{-1}{x}}$. Thus,

$\varphi(x) = x^{\frac{1}{\rho}}$. Since $\varphi'(x)$ is increasing on $[1,\infty)$ and $\lim_{x\to\infty} \frac{\varphi'(x+1)}{\varphi'(x)} = 1$, φ' is logarithmic.

Therefore, $T_{\varphi,\rho}(S(n)) = \frac{S(n) - \frac{1}{\rho}S(n^{\frac{1}{\rho}})}{1-\frac{1}{\rho}}$ converges more rapidly than $S(n^{\frac{1}{\rho}})$.

Example 3. Let $f(x) = \frac{1}{x \ln x (\ln(\ln x))^2}$. Then, $F(x) = \frac{-1}{\ln(\ln x)}$ and $F^{-1}(x) = e^{e^{-x}}$

Hence, $\varphi(x) = e^{(\ln x)^{\frac{1}{p}}}$ and $\varphi'(x) = \frac{(\ln x)^{\left(\frac{1}{\rho_p}-1\right)}\varphi(x)}{\rho x}$. It is a straightforward exercise to show that φ' is logarithmic. Thus, $T_{\varphi,\rho}(S(n))$ converges more rapidly than $S(\varphi(n))$.

References

[1] Bromwich, T. J., An Introduction to the Theory of Infinite Series, Macmillan & Co. Ltd. New York, Second Edition, 1926.

[2] H.L. Gray, and W.D. Clark, "On a Class of Nonlinear Transformation and Their Applications to the Evaluation of Infinite Series," Journal of Research of the National Bureau of Standards-B. Mathematical Sciences, Vol. 73B, No. 3, July–September, 1969.

A Conceptual Approach to Assimilating Remote Sensing Data to Improve Soil Moisture Profile Estimates in a Surface Flux/Hydrology Model
Part 1: Overview

William L. Crosson[a], Charles A. Laymon[a], Ramarao Inguva[b], Marius Schamschula[c], John Caulfield[d]

[a]Institute for Global Change Research and Education, Global Hydrology and Climate Center, Huntsville, AL

[b]Science and Technology Consultants, Huntsville, AL

department of Physics and Center for Hydrology, Soil Climatology, and Remote Sensing, Alabama A&M University; Normal, AL

'Department of Physics, Fisk University, Nashville, TN

I. Introduction

Knowledge of the amount of water in the soil is of great importance to many earth science disciplines. Soil moisture is a key variable in controlling the exchange of water and energy between the land surface and the atmosphere. Thus, soil moisture information is valuable in a wide range of applications including weather and climate, runoff potential and flood control, early warning of droughts, irrigation, crop yield forecasting, soil erosion, reservoir management, geotechnical engineering, and water quality. Despite the importance of soil moisture information, widespread and continuous measurements of soil moisture are not possible today. Although many earth surface conditions can be measured from satellites, we still cannot adequately measure soil moisture from space.

Because of its spatial nature and temporal continuity, remote sensing holds great promise for hydrology. Today, it is used to characterize topography, land cover, precipitation, snow cover, surface temperature, evapotranspiration, and other variables. The availability of remote sensing data has facilitated a move from lumped hydrologic models to more spatially-distributed models. As remote sensing has developed, we have learned to make measurements of hydrologic variables not possible with traditional techniques. Despite the potential, the hydrologic community has been slow to embrace remote sensing because traditional modeling variables are not available via remote sensing, or the resolution of the remote sensing data sets are inconsistent with the modeling scale Or the scale of the hydrologic processes of interest.

There exists a scale disparity between the resolution of remote sensing data, the scale of hydrologic processes which we are attempting to measure with the remote sensor, and the grid scale of the hydrologic model. When designing an instrument, trade-offs are made among the potential applications of the data and development costs. The resolution of remote sensing data is a function of sensor design, the energy wavelengths which the instrument is designed to detect, scanning or viewing characteristics of the instrument, and data processing procedures. Usually the temporal frequency of the measurement is inconsistent with temporal scale of a hydrologic process. The spatial scale of hydrologic processes is somewhat more ambiguous. In nature, the process scale is the minimum scale at which the variability of a process has a negligible impact on the phenomenon of interest, and is clearly dependent on the natural spatial heterogeneity of related properties. The model grid scale is the length scale at which processes are simulated and is controlled more by resolution of available input data and computational efficiency than by the scale of the hydrological processes. Problems associated with scale disparities are exacerbated as the resolution of the remote sensing data decreases. For technological reasons, the longer the

sensed wavelength, the coarser the resolution of the resulting data set. At microwave wavelengths, data resolution is on the order of 10 to 100 km!

Research in soil moisture remote sensing began in the mid 1970's shortly after the surge in satellite development (Eagleman and Lin, 1976; Njoku and Kong, 1977; Schmugge et al., 1977). Recent advances in remote sensing have shown that soil moisture can be measured, at least qualitatively, by several methods. Quantitative measurements of moisture in the soil surface layer have been most successful using both passive and active microwave remote sensing, although complications arise from surface roughness and vegetation type and density. Early attempts to measure soil moisture from space-borne microwave instruments were hindered by what is now considered sub-optimal wavelengths (shorter than 5 cm) and the coarse spatial resolution of the measurements. L-band frequencies between 1 and 3 GHz (10-30 cm) have been deemed optimal for detection of soil moisture in the upper few centimeters of soil. The Electronically Steered Thinned Array Radiometer (ESTAR), an aircraft-based instrument operating a 1,4 GHz, has shown great promise for soil moisture determination. Initiatives are underway to develop a similar instrument for space. Existing space-borne synthetic aperture radars (SARs) operating at C- and L-band have also shown some potential to detect surface wetness. The advantage of radar is its much higher resolution than passive microwave systems, but it is currently hampered by surface roughness effects and the lack of a good algorithm based on a single frequency and single polarization. In addition, its repeat frequency is generally low (about 40 days).

In the meantime, two new radiometers offer some hope for remote sensing of soil moisture from space. The Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), launched in November 1997, possesses a 10.65 GHz channel and the Advanced Microwave Scanning Radiometer (AMSR) on both the ADEOS-II and Earth Observing System AM-1 platforms to be launched in 1999 possesses a 6.9 GHz channel. Aside from issues about interference from vegetation, the coarse resolution of these data will provide considerable challenges pertaining to their application. The resolution of TMI is about 45 km and that of AMSR is about 70 km. These resolutions are grossly inconsistent with the scale of soil moisture processes and the spatial variability of factors that control soil moisture. Scale disparities such as these are forcing us to rethink how we assimilate data of various scales in hydrologic models. Of particular interest is how to assimilate soil moisture data by reconciling the scale disparity between what we can expect from present and future remote sensing measurements of soil moisture and modeling soil moisture processes.

It is because of this disparity between the resolution of space-based sensors and the scale of data needed for capturing the spatial variability of soil moisture and related properties that remote sensing of soil moisture has not met with more widespread success. Within a single footprint of current sensors at the wavelengths optimal for this application, in most cases there is enormous heterogeneity in soil moisture created by differences in landcover, soils and topography, as well as variability in antecedent precipitation. It is difficult to interpret the meaning of 'mean' soil moisture under such conditions and even more difficult to apply such a value. Because of the non-linear relationships between near-surface soil moisture and other variables of interest, such as surface energy fluxes and runoff, mean soil moisture has little applicability at such large scales. It is for these reasons that the use of remote sensing in conjunction with a hydrologic model appears to be of benefit in capturing the complete spatial and temporal structure of soil moisture.

This paper is Part I of a four-part series describing a method for intermittently assimilating remotely-sensed soil moisture information to improve performance of a distributed land surface hydrology model. The method, summarized in section II, involves the following components, each of which is detailed in the indicated section of this paper or subsequent papers in this series:

. Forward radiative transfer model methods (section II and Part IV)

- Use of a Kalman filter to assimilate remotely-sensed soil moisture estimates with the model profile (section II and Part IV)
- Application of a soil hydrology model to capture the continuous evolution of the soil moisture profile within and below the root zone (section III)
- Statistical aggregation techniques (section IV and Part 11)
- Disaggregation techniques using a neural network approach (section IV and Part III)
- Maximum likelihood and Bayesian algorithms for inversely solving for the soil moisture profile in the upper few cm (Part IV)

II. Framework

The method presented here for estimating the three-dimensional soil moisture distribution is based, in part, on the premise that the appropriate time scales for soil hydrologic processes are depth-dependent. Soil moisture in the upper layers responds rapidly to precipitation, infiltration, evapotranspiration and vertical fluxes. Uncertainties associated with modeling near-surface soil moisture are greater than for the deep soil due to imperfect forcing variables (primarily precipitation); it is in this domain only that remote sensing is of potential benefit. Deep soil moisture functions on a much longer time scale and thus has different modeling requirements. Tailoring our methodology to accommodate these time scales helps to minimize computational burden. In this section, we present an overview of the method; summaries of each component are given in the following sections.

The proposed framework for estimating the spatial nature of the soil moisture profile is represented schematically in Figure 1. Time is represented in this schematic by the vertical dimension, and the computational procedure advances in a counter-clockwise direction. Models or algorithms are represented by cubical elements, and outputs from these are shown as two-dimensional grids. It is assumed here that there exist only two spatial scales - a 'fine' or high-resolution scale at which the surface model operates, and a 'coarse' or low-resolution scale commensurate with microwave remote sensing footprints. High-resolution grids are shown here as two-dimensional fields with grid lines, low-resolution fields without.

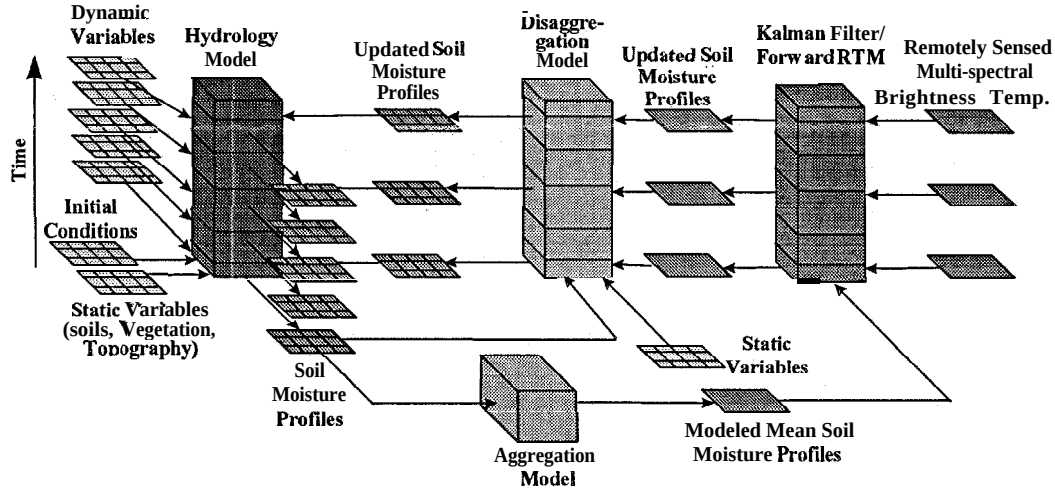


Fig. 1. Schematic depiction of the method for using remote sensing to update soil moisture profiles in a land surface hydrology model.

The land surface hydrology-runoff modeling system SHEELS-DRUM (Simulator for Hydrology and Energy Exchange at the Land Surface-Distributed Runoff Model, described in the following section) is initialized with soil moisture profiles based on measurements or climatology and static surface properties representing soils, vegetation and topography. Meteorological inputs are required at each time step. During times of precipitation and subsequent runoff, SHEELS-DRUM operates as a filly distributed modeling system, estimating the lateral flow of surface water between grid cells. Between precipitation events, SHEELS is run without DRUM in a gridded mode with no connectivity between grid points. In both cases, the model produces spatial grids of moisture for each soil layer.

At times when remotely-sensed brightness temperatures (T_B) are available, model soil moisture profiles to a specified depth (5-10 cm) are aggregated to produce moisture profiles at the coarser scale of remotely-sensed observations. Using aggregated modeled soil moisture and observed T_B for a set of microwave frequencies, a Kalman filter is applied. A forward radiative transfer model built into the Kalman filter converts soil moisture profiles to T_B at each frequency. The output is an updated or 'optimal' soil moisture profile at satellite resolution. The Kalman filter serves to nudge the model estimates toward the measurements and thus is used as a consistency check in which the remote observations constrain the model output. This technique is similar to that proposed by Entekhabi et al. (1994). The low-resolution soil moisture profiles are then disaggregated back to model scale. The resulting high-resolution spatial fields of soil moisture at each soil layer are then used to update the model, which continues to run until the next time remote sensing data are available. Because of the limited emitting depth of microwave remote sensing, only the upper part of the profile is adjusted by direct substitution of the remotely-sensed moisture estimate. The remainder of the profile is then adjusted in keeping with the new surface value by a blending technique.

III. Land surface hydrology model

A. General description

The distributed land surface hydrology modeling system that will be used in this project comprises the models SHEELS and DRUM. In the coupled model system, SHEELS is run continuously to estimate surface energy fluxes and sub-surface soil water fluxes, while DRUM operates only during and following precipitation events to predict surface runoff and streamflow. The lateral re-distribution of surface water determined by DRUM is passed to SHEELS, which adjusts soil water contents throughout the profile. Although SHEELS-DRUM is designed for watershed studies, SHEELS can be applied at larger scales in which the sub-grid scale variability of vegetation, soil and topography properties is represented statistically using scaling algorithms.

The physics of SHEELS are based on those of the Biosphere-Atmosphere Transfer Scheme (BATS) (Dickinson et al., 1993). SHEELS has retained the physical treatment of vegetation properties and the surface flux parameterizations of BATS. However, the sub-surface processes in SHEELS differ significantly. The nested three-layer approach of BATS has been converted to a discrete layer configuration in SHEELS, in which the number and depth of layers is flexible. This affords better vertical resolution, especially near the surface where temperature and moisture gradients are large, and is compatible with remotely-sensed moisture data. Also, the soil water dynamics algorithms have been modified following the Darcy flow equation approach of the Soil Hydrology Model (SHM; Capehart and Carlson, 1994). This module estimates diffusive exchange and drainage between each of the soil layers.

DRUM is a distributed, storm-event surface runoff model using a finite element solution of the kinematic wave equation (Vieux et al., 1990; Vieux and Gauer, 1994). Explicit patterns of land use, soils and topography are required, as are initial surface soil moisture conditions. Slope and

aspect at each grid cell derived from a Digital Elevation Model control the lateral movement of surface water. During a storm event, DRUM is driven by rainfall and diagnoses surface runoff and routing of water into a channel network. The model continues to run after rainfall ceases until base flow is approached. Typical temporal and spatial resolution are 1 minute and 1 km.

B. Model inputs and outputs

SHEELS is driven by seven atmospheric forcing variables – air temperature, relative humidity, atmospheric pressure, incident solar and infrared radiative fluxes, wind speed and precipitation. Meteorological input variables may be spatially distributed or uniform, depending on the scale of the model domain. Because the spatial variability of precipitation is large relative to the other variables and because soil moisture variability is controlled to a large degree by the distribution of precipitation, it will be input in a spatially-distributed manner. SHEELS also requires spatially distributed soil properties, including saturated hydraulic conductivity, saturated matric potential, depths of layers, porosity, and wilting point. Required vegetation properties include canopy height, fractional cover, minimum stomata] resistance, leaf area index, and reflectance properties. These are specified as functions of landcover class.

The output of SHEELS-DRUM of primary interest in this context is the soil moisture profile. SHEELS estimates both gravimetric and volumetric water contents at each time step for each soil layer. The layer depths are model parameters and can be established to meet the needs of the application. For remote sensing studies in which the upper soil moisture profile is being updated using microwave data, the upper 5-10 cm soil depth is partitioned into layers of approximately 1 cm each. In determining soil moisture, infiltration, evapotranspiration, runoff (surface and sub-surface) and ponded water are modeled. During periods of precipitation and runoff, DRUM estimates infiltration and runoff at each grid point, routing the runoff to the downslope grid cell.

C. Soil water dynamics equations

The methods for modeling the movement of water into and within the soil medium are presented in this section. Formulations of other model variables such as evapotranspiration, sensible heat flux and temperatures are similar to those in an earlier version of the model (Smith et al., 1993). A description of the overland flow and streamflow calculations in DRUM are given in Vieux et al. 1990 and Vieux and Gauer, 1994.

The temporal change in soil moisture content in each of the N layers, as well as surface runoff and ponded water, are determined by considering the contributions of precipitation, infiltration, evaporation, transpiration, diffusion, and gravitational drainage. First, infiltration (I) is calculated using the Green-Ampt equation based on the amount of precipitation reaching the soil surface directly or through the vegetation canopy. Surface runoff is then calculated based on the local slope angle and precipitation in excess of infiltration. The amount of water reaching the soil surface which does not run off or infiltrate is assumed to pond on the surface. Given infiltration, evaporation (E) and transpiration (T), the change in depth of water in soil layer i can be expressed in terms of the amounts of infiltration, evaporation, and transpiration attributed to the layer. These terms are proportions of the total quantities I, E, and T, and are determined by a weighting function which decreases to a value of zero at specified depths. After water depths are adjusted to account for infiltration, evaporation, and transpiration, volumetric water contents are updated. Finally, the effect of vertical soil water fluxes (diffusion and drainage) on soil moisture in each layer is determined using Darcy's equation, a version of which is solved in matrix form for the N soil layers using the Crank-Nicholson numerical scheme.

IV. Aggregation and disaggregation methods

Model-estimated soil moisture is aggregated from the high resolution of the model to produce spatial fields at the coarser grid size of the remotely-sensed data. The simplest means by which this is performed is arithmetic averaging. However, linear aggregation may not produce the effective moisture content as measured by a remote sensor at the larger scale. This is because of two effects. First, there is a non-linear relationship between soil moisture and brightness temperature. Second, the sensitivity of a microwave sensor is a function of vegetation density which is likely to vary within the sensor footprint. The aggregation procedure is described in more detail in Part II.

Disaggregation is necessary to convert the updated soil moisture profiles at satellite scale (the output of the Kalman filter) to information that can be used to update the hydrology model. The disaggregation problem is one of 'creating more from less', that is, increasing the information content of the input data. However, there are valuable pieces of information that can be utilized in this algorithm, particularly the previous spatial pattern of soil moisture as well as static surface properties (landcover, soils, and topography). The disaggregation procedure utilizes these inputs and neural network techniques to obtain the updated soil moisture at model resolution. This technique is discussed in Part III.

References

Capehart, W.J. and T.N. Carlson, 1994. Estimating near-surface soil moisture availability using a meteorologically driven soil-water profile model. *J. Hydrology*, 160, 1-20.

Dickinson, R. E., A. Henderson-Sellers, P.J. Kennedy and F. Giorgi, 1993. Biosphere Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Climate Model. NCAR/TN-387+STR, 72 pp.

Eagleman, J.R. and W.C. Lin, 1976. Remote sensing of soil moisture by a 21-cm passive radiometer. *J. Geophys. Res.*, 81, 3660-3666.

Entekhabi, D., H. Nakamura and E.G. Njoku, 1994. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. ***IEEE Trans. Geosci. Rem. Sens., 32, 438-448.***

Njoku, E.G. and N.-A. Kong, 1977. Theory for passive microwave remote sensing of near-surface soil moisture. *J. Geophys. Res., 82, 3108-3118.*

Schmugge, T. J., J.M. Menennly, A. Rango and R. Neff, 1977. Satellite microwave observations of soil moisture variations. ***Water Resour. Bull.,*** 13, 265-281.

Smith, E. A., W.L. Crosson, H.J. Cooper, and H.-Y. Weng, 1993. Estimation of surface heat and moisture fluxes over a prairie grassland. Part III: Design of a hybrid physical/remote sensing biosphere model. J. *Geophys. Res., 98,4951-4978.*

Vieux, B.E., V.F. Bralts, L.J. Segerlind and R.B. Wallace, 1990. Finite element watershed modeling: One-dimensional elements. *J. Water Res. Planning and Mngt.,* 116, 803-819.

Vieux, B.E. and N. Gauer, 1994. Finite element modeling of storm water runoff using GRASS GIS. ***Microcomputers in Civ. Eng.,*** 9,263-270.

98 URC060

A Conceptual Approach to Assimilating Remote Sensing Data to Improve Soil Moisture Profile Estimates in a Surface Flux/Hydrology Model. 2. Aggregation.

Marius Schamschula[c], William L. Crosson[a], Ramarao Inguva[b], Thomas Yates[d], Charles A. Laymen', John Caulfield[e]

'Institute for Global Change Research and Education, Global Hydrology and Climate Center, Huntsville, AL

[b]Science and Technology Consultants, Huntsville, AL, and Department of Physics , University of Wyoming, Laramie, WY

'Department of Physics and Center for Hydrology, Soil Climatology, and Remote Sensing, Alabama A&M University, Normal, AL

[d]Oakwood College, Huntsville, AL

'Department of Physics, Fisk University, Nashville, TN

Abstract

This is a follow up on the preceding presentation by Crosson. The grid size for remote microwave measurements is much coarser than the hydrological model computatational grids. To validate the hydrological models with measurements we propose mechanisms to aggregate the hydrological model outputs for soil moisture to allow comparison with measurements. Weighted neighborhood averaging methods are proposed to facilitate the comparison. We willalso discuss such complications as misalignment, rotation and other distortions introduced by a generalized sensor image.

Z. Introduction

Given two disparate scales for sensor and model data, how can we use sensors to verify the validity of the hydrological model? A computer model is only limited by the speed and memory available on a given processor. Thus it makes good sense to make a tight mesh of grid points. The sensor measurement, either air or space based, is limited by the resolution of the sensor. The ground "footprint" of a sensor pixel is usually much larger than the grid spacing of the hydrological model.

We also have to remember that the model grid is freed, whereas the sensor is moving along some trajectory, giving a generalized ground "footprint. " Given the state vector of the sensor, it is possible to transform the model data to match the image content of the sensor.

II. The Required Transformation

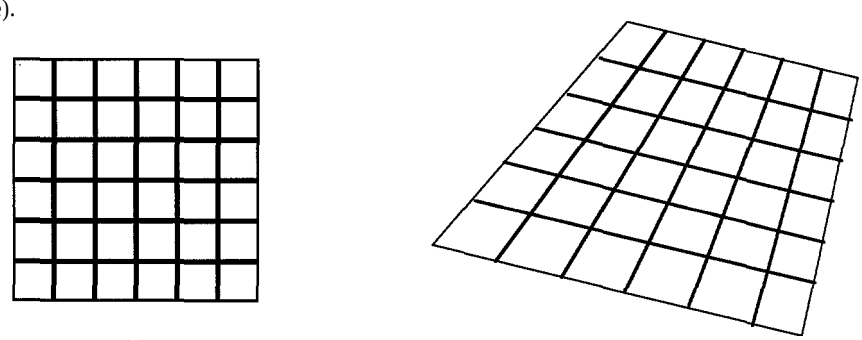
The most general ground footprint of an image from a rectangular sensor array is a trapezoid. This means that the pixels closest to the sensor show a smaller area than the ones farthest away. Given the state vector of the sensor platform, be it an airplane or satelite, and the diretion the sensor is pointing, we can calculate the ground footprint of the sensor for a given image.

In the best case scenario, the grid will be aligned, and the image is taken from directly above. In this case we just need to synthezise the information equivalent to the sensor data from the model by some averaging algorithm, with a constant number of model grid points contributing to each "sensor" pixel.

This alignment and sensor orientation, however, is almost never the case for real data. Usually there is a rotation between the sensor and model grids. This can be

easily remieded with a simple matrix rotation. The more general problem is the distortion introduced if the sensor is not pointing directly downwards. This problem can also be removed by a set of mathematical transformations.

A typical example for an airborne sesnor is shown in Figure 1. Note, however, that more complex transformations are involved when using spaceborne sensors (Earth curvature).



(a) (b)

Figure 1. a) model grid, and b) sensor grid. Note that the scale difference is usually much greater.

Once we have determined what transformations need to be applied to go from the sensor data to the model data, we are ready to do the inverse: transform model data into the same image space as the sensor data. This transformation, however, is not the whole story. There are a number of model grid points within the foofprint of each sensor pixel. This number is rarely constant, due to the distortions and rotation introduced by the transformation. Often model grid points contribute to multiple sensor pixels. From the transformation we can determine the required ratios.

Once we have determined how much of each model grid point contributes to each sensor pixel we are ready for the main task Aggregation.

III. Aggregation

In order to obtain a meaningful output from the aggregation algorithm, we have to choose the proper variable to aggregate. Given the nature of the hydrological model, we choose the brightness temperature T_B of the soil. It is" rather straight forward to extract this information from the various layers of the model. This parameter corresponds to the sensor output, once it has been calibrated and processed.

The brightness temperature includes weighted information about the layers of soil down to the penetration depth of the sensor. More sophisticated variants of this calculation will include such factors as soil type, and vegetative ground cover. In some cases, this ground is not covered by soil, but is a body of water, or some man-made structure such as a road or building. In either of these cases, there is no need to apply the model to extract the T_B. Each situation is one extreme, complete moisture for a body of water, or complete lack of moisture for a building (excluding puddles, snow or ice packs on the roof).

The most simple approach to aggregation is a simple mean. If we require data from N model grid points, we could set the brightness temperature to

$$T_{BM} = \Sigma (T_{BM})_{xy}/N.$$

This is a rather crude approximation. It is useful to weight the membership of each model data point. The weight function has two primary geometrical factors:

1. The model grid point is only partially within the sensor pixel, and

2. the distortion of each sensor pixel's footprint near model grid ements contribute more than those further away.

Even though such factors as soil type and ground cover can be included in the calculation of the T_B for each model grid cell, this may not be sufficent. We also need to include a weighting factors for:

1. soil type, and
2. ground cover,

in the calculation of the mean brightness temperature.

AU these geometrical and ***a priori*** factors are included in a weight function w_{xy}. Any fractional pixels have to be accounted for in the calculation of the mean. Thus we obtain

$$T_{BM} = \Sigma w_{xy} \{(T_{BM})_{xy}\}/N',$$

the adjusted number of model grid points N.

Note that the weighting function need not be linear. It can assume any functional from such as a power law.

IV. The Comparison and Model Adjustment

The comparison of T_{BM} to the sensor T_B leads to a verification of the current state of the model. How do we adjust the parameters in the hydrological model? We are faced with the fact that each sensor pixel can represent any number of grid points in the hydrological model, and each grid point has soil moisture information for several depths. We will use the method of disaggregation to take the sensor information and split it up into the proper number of model grid points. Once T_B values for each grid point have been obtained, a inverse radiative transfer model has to be invoked to obtain the soil moisture for various depths. Our approach to disaggregation is topic of another paper in this set.

98 URC061

**A Conceptual Approach to Assimilating Remote Sensing Data to Improve Soil Moisture Profile Estimates in a Surface Flux/Hydrology Model. 3. Disaggregation.

John Caulfield[d], William L. Crosson[a], Ramarao Inguva[b], Charles A. Laymon[a], Marius Schamschula[c]

[a]Institute for Global Change Research and Education, Global Hydrology and Climate Center, Huntsville, AL

[b]Science and Technology Consultants, Huntsville, AL, and Department of Physics , University of Wyoming, Laramie, WY

[c]Department of Physics and Center for Hydrology, Soil Climatology, and Remote Sensing, Alabama A&M University, Normal, AL

[d]Department of Physics, Fisk University, Nashville, TN

Abstract

This is a followup on the preceding presentation by Crosson and Schamschula. The grid size for remote microwave measurements is much coarser than the hydrological model computatational grids. To validate the hydrological models with measurements we propose mechanisms to disaggregate the microwave measurements to allow comparison with outputs from the hydrological models. Weighted interpolation and Bayesian methods are proposed to facilitate the comparison. While remote measurements occur at a large scale, they reflect underlying small-scale features. We can give continuing estimates of the small scale features by correcting the simple 0^{th}-order ?, starting with each small-scale model with each large-scale measurement using a straightforward method based on Kalman filtering.

Z. Introduction

This paper is part of a set of papers jointly describing how an on-going model of the soil moisture in 4D [three spatial coordinates plus time) can be constructed and maintained. This model can be run on "fast forward to predict future states as well. No attempt will be made to make this paper self standing by reviewing the others.

The problem to be addressed is that of scale. Each measurement may have its own spatio-temporal "footprint" and ground penetration capability. Almost never will those correspond to the needs of

- computer modellers (who must do their work well in order to make the on-going model accurate),
- forecasters (who want to use both the current state of the model and the models of its future state to predict floods, harvests, response to irrigation, etc.), or
- local users (who care about their field, not their neighbors').

If we can use the on-going model to characterize "small" regions, then we can aggregate (described in another paper from this set) to

- satisfy local users, and
- give a "predicted value to compare with the observed value for any measurement footprint.

How we go from the large footprints which may cover only part of the models area and may also cover regions outside that area is the problem of disaggregation.

II. Disaggregation Problems

Of all the problems which must be solved in making the global, on-going model accurate; this is the most difficult. The reason for that difficulty is easy to identify: disaggregation is wildly under-determined. There are infinitely many disaggregations consistent with aggregate measurements. How can we choose one? And, how confident can we be about that choice?

Readers of the previous paper in this set will recognize this as a classic inverse problem and know that the optimum approach is through Bayesian logic.

Unfortunately, doing it is much harder than invoking it. The real problem, therefore, is how to approximate a Bayesian disaggregation.

III. Bayesian Disaggregation

Let the detailed (at the finest resolution of the needed disaggregation) description of the true soil moisture pattern be m (x, y, z, t). This leads to a radiation pattern r (x, y, z, t). As there is a known, invertible relationship between m and r, we choose to disaggregate in terms of r. We can always invert r to get m - the desired property.

Fortunately we have some *a priori* information. Here are some of the knowables and guessable:

- We have a map of the region being modeled to the r-resolution level,
- Within that map/region, we can identify certain broad features such as
  1, open fields of known porosity, etc.
  2. streams, rivers, lakes, ponds, and other m = 1 objects,
  3. roads, buildings of various types, and other human artifacts for which we might expect (except for snow/ice cover) m = O,
  4. crops of known nature and state of development,
  5. woods of various types and densities, and
  6. other namable constituents.

All of these entered into our ongoing model as described in earlier parts of this multi-part paper.

Let us assign each pixel of r to one of those classes. Our detailed models predict r and its variance for each class at any moment. This gives a O^{th}-order disaggregation model.

But that disaggregated model has not been corrected by the aggregated measurements.

Suppose, over some large footprint we measure R_m but our aggregated 0^{th} -order model predicts $R_0 \neq R_m$. How can we "correct" our disaggregated model?

As noted in other papers in this series,

$$\Delta R = R_m - R_o$$

is called the innovation in R Using Kalman filtering (itself derivable form Bayes' theorem), we find the innovation in r is

$$\Delta r = k\, AR$$

where k is the Kalman gain and the best estimate of r is

$$r = r_{predicted} + \Delta r$$

This implies a rather interesting and simple disaggregation updating scheme in view of AR. All pixels of r share the burden of accounting for AR equally! Regions where r is high require relatively low adjustment. Regions where r is small require relatively large adjustments.

Let us now look at a given pixel of r. Suppose there are N variable r pixels in R (leaving the m = O and m = 1 regions freed). Then each variable pixel of r must account 1/N of AR. This is a simple renormailization reminiscent of Dempster-Shafer theory.

IV. Conclusions

A straightforward Kalman update of the predicted disaggregation state in view of an aggregated measurement and taking into account "fixed pixels" can be justified on a Bayesian basis.

98 URC062

Estimation of Soil Moisture Profile using a Simple Hydrology Model and Passive Microwave Remote Sensing

Vishwas V. Soman[1], William L. Crosson[1], Charles Laymon[1], and Teferi Tsegaye[2]

[1]Institute for Global Change Research and Education, Huntsville, AL
[2]Center for Hydrology, Soil Climatology, and Remote Sensing, Alabama A&M University, Normal, AL

Abstract

Soil moisture is an important component of analysis in many Earth science disciplines. Soil moisture information can be obtained either by using microwave remote sensing or by using a hydrologic model. In this study, we combined these two approaches to increase the accuracy of profile soil moisture estimation. A hydrologic model was used to analyze the errors in the estimation of soil moisture using the data collected during Huntsville '96 microwave remote sensing experiment in Huntsville, Alabama. Root mean square errors (RMSE) in soil moisture estimation increase by 22% with increase in the model input interval from 6 hr to 12 hr for the grass-covered plot. RMSEs were reduced for given model time step by 20-50% when model soil moisture estimates were updated using remotely-sensed data. This methodology has a potential to be employed in soil moisture estimation using rainfall data collected by a space-borne sensor, such as the Tropical Rainfall Measuring Mission (TRMM) satellite, if remotely-sensed data are available to update the model estimates.

1. Introduction

Soil moisture is an important component of the hydrologic cycle which influences runoff, subsurface runoff, infiltration, evaporation and transpiration. In agronomy, it is considered important due to its influence on crop yield. Soil moisture is an important input to global circulation models. In these and other fields, there exists an increasing need for soil moisture information on regional and global scales. Field measurements of soil moisture over these scales is impossible due to inherently high temporal and spatial variability. Microwave remote sensing can be employed to measure surface soil moisture conditions (Jackson, 1993; Engman and Chauhan, 1995). Remote sensing is spatial in nature and therefore can be used to make spatially comprehensive measurements. In addition, microwave remote sensing could be used from a space platform to make global soil moisture measurements. Applicability of microwave remote sensing for estimating soil moisture is limited due to (1) measurements restricted to a thin top layer of soil (-5 -10 cm) and (2) low frequency of the measurements (1 observation every few days). Although surface soil moisture measurements have direct applicability in some areas, a much wider range of applications in hydrology and agronomy can benefit from the soil moisture information for the entire profile.

There are several approaches to estimate the profile soil moisture information using surface soil moisture measurements. The simplest of these approaches is to use a regression equation to predict the profile soil moisture using surface layer measurements. This approach has been used with some degree of success by several researchers (Biswas and Dasgupta, 1979; Blanchard, 1979; Smith and Newton, 1983; Arya et al., 1983). This approach can only work, however, if the relationship between the profile soil moisture and the surface measurements is linear, a condition that is rarely satisfied. Furthermore, this approach is site-specific and, thus, cannot be used as a general solution, especially over large areas (e.g. watershed scale).

Another approach to address this issue is to use a model with surface soil moisture as the driving component. Bernard et al. (1981) used this approach for modeling soil water and estimating evapotranspiration under bare soil conditions. The authors employed Richards' equation to model one-dimensional isothermal soil water movement and were able to estimate evaporation with reasonable accuracy. Smith and Newton (1983) developed a physically-based soil water simulation model that utilizes remotely-sensed data. The authors used this model to predict the soil moisture profile based on hourly observations of surface moisture.

Soil moisture information can also be obtained using a hydrologic model alone. A sophisticated hydrologic model could be used to estimate the soil moisture profile to a desired depth and for any desired time step, if the required input data for the hydrologic model are available. Though this approach seems to have an advantage over other approaches, it also has its own limitations. These limitations are imposed by the accuracy and frequency of the available input data as well as uncertainty in soil and vegetation properties. The model soil moisture estimates are especially sensitive to precipitation data. In some real-world situations the rainfall data might be available at a low frequency. For example, the sampling interval of the space-borne Precipitation Radar on the Tropical Rainfall Measuring Mission (TRMM) satellite is 11.75 hr. Such a long interval may miss the entire rainfall event resulting in errors in soil moisture estimates beyond acceptable limits. Also, the errors due to inaccurate inputs could continue to increase due to the temporal correlation between successive soil moisture estimates by the model.

Thus, from the above discussion we can see that these two sources of soil moisture information (microwave remote sensing and hydrologic modeling) have their own set of limitations when used alone. However, these two sources seem to complement each other when used in combination. Therefore, we hypothesize that hydrologic modeling in combination with periodic remotely-sensed surface soil moisture can produce better estimates of profile soil moisture than either techniques alone. Our objectives are to investigate (1) the behavior of errors in the estimation of soil moisture with decrease in temporal frequency of the model input, and (2) the use of remotely-sensed surface soil moisture data to reduce these errors. We have used a hydrologic model to calculate profile soil moisture. Modeled soil moisture values are "corrected at times when remotely-sensed observations are available. This approach is different from that of Bernard et al. (1981) and Smith and Newton (1 983) in that we are not using remotely-sensed data to drive the hydrologic model but to correct the modeled estimates.

2. **Field Experiment Data**

Data were collected during the Huntsville' 96 microwave remote sensing of soil moisture field experiment conducted in Huntsville, Alabama, from July 1-14, 1996 at Alabama A&M University. A research site was established at the University's Agricultural Research Station located about 20 km north of Huntsville. The experiment site contained four plots each about 50 x 60 m. Two plots were vegetated and other two were bare. Laymen et al. (1997) provides a detailed description of the experimental design and objectives, as well as some preliminary results regarding remote sensing of soil moisture.

The hydrologic model requires meteorological and soil parameter related data. Meteorological, soil moisture profile data were continuously measured by automated systems. Meteorological data collected include air temperature, solar and longwave radiation, relative humidity, wind speed and direction, and rainfall. All measurements were made at 10-second intervals and averaged over 20-minute periods. The soil moisture profile data collected using Water Content Reflectometers were used to initialize the hydrologic model. Surface soil moisture measurements were made with the S (2.65 GHz, 11.3 cm) and L (1.4 13 GHz, 21.2 cm) band Microwave Radiometer (SLMR). Measurements were made over all the plots on nearly an hourly basis during the day throughout the experiment. The radiometers were put in autocollect mode over one of the plots during the night.

3. **Simple Hydrology Model**

The Simple Hydrology Model (SHM) used in this study is a modified version of the Soil Hydrology Model by Capehart and Carlson (1994). This model is a one-dimensional simple water budget model developed to determine the vertical profile of soil moisture. The model is driven by conventional meteorological and land use data. Evapotranspiration is estimated using the Penman-Monteith equation and infiltration is estimated using the Green-Ampt technique. (See Capehart and Carlson, 1994 for details).

The Soil Hydrology Model was modified to meet the requirements of this study. Site-specific parameters were replaced with variables. Modifications were made with regards to handling the meteorological data and setting vegetation parameters. The most important changes are the addition of two methods to update the modeled moisture profile using remotely-sensed data. The first method replaces the modeled estimates of soil moisture for the uppermost layer with remotely-sensed observations whenever such observations are available. The second method updates the modeled soil moisture profile to a specified depth by blending remotely-sensed information with modeled soil moisture profile. In this study, this depth was set to 5 cm as this value produced the most realistic vertical soil moisture gradients.

4. Analysis

4.1 Establishing a baseline

To compare the soil moisture values estimated in each model run, a baseline soil moisture time series was established. Several factors that were considered include availability of representative soil moisture values, availability of measurements, and the overall objective of the study. The SHM output using the data with finest temporal resolution (i.e. sampling interval of 20 rein) and remotely-sensed updates using profile blending was chosen as the baseline. This option was considered the best estimate because it uses all the available data (i.e. meteorological and remotely-sensed data) to obtain the soil moisture profile.

4.2 Simulations

In this study, simulations were conducted for the 14-day period from July 2-July 16 using Huntsville '96 data for the grass-covered plot. Input data required for the model runs includes air temperature, dew point temperature, radiation and precipitation. These data were available at 20 min intervals. Model runs made using input intervals of 6 hr and 12 hr were compared to the baseline run with 20-min input data to investigate the effect of decreasing the temporal frequency on soil moisture estimation. SHM was run at the 6 and 12 hr time steps using input values from the 20-min data set. Meteorological conditions, including rain rate, were assumed to be constant over the model time step. For precipitation, the observed 20-min. rate was scaled to produce total rainfall for the given model time step. The soil moisture profile was initialized using surface gravimetric measurements and WCR observations. Remotely-sensed soil moisture observations were used (1) to directly replace the model estimates of soil moisture for the upper-most soil layer (method 1) and (2) to define a revised soil moisture profile for the upper 5 cm by the assimilation scheme described earlier (method 2).

5. Results and Discussion

The root mean square errors (RMSE) from the baseline for average soil moisture through the upper 5 cm layer are shown in Table 1 for model time steps of 6 hr and 12 hr. Figures 1 and 2 show the soil moisture estimates for the baseline and by the following three methods: (1) Model run with no remote sensing updates, (2) Model run with method 1 updating and (3) Model run with method 2 updating for model time steps of 6 hr and 12 hr respectively. The model with 6 hr time step missed the first rainfall event completely, but captured the second. A considerable improvement in the soil moisture estimation could be noticed after the second rainfall event. Remotely-sensed updates helped in reducing the effects of missing the first rainfall by increasing the soil moisture estimates (Figure 1). The model with 12 hr time step (Figure 2), however, missed both the rainfall peaks (although could capture some rainfall from the second event) which resulted in large errors in the soil moisture estimation.

Remote sensing updates by method 1 reduced the RMSEs by 22% and 14% for time steps of 6 hr and 12 hr respectively. These reductions are smaller compared to the reductions in RMSEs due to remote sensing updates by method 2. Remote sensing updates method 2 reduced RMSEs by 49% for 6 hr and by 41'%. for 12 hr time steps. These reductions are higher than those obtained for bare plots. For bare plots, the reduction in RMSEs due to remote sensing updates by method 2 were 210/. for 6 hr and 160/0 for 12 hr time steps (Soman et al., 1997). Figures 1 and 2 clearly indicate the errors in soil moisture estimation that occur when the rainfall information is inadequate. However, this situation can be ameliorated with efficient use of remotely-sensed observations and an assimilation scheme.

6. Conclusions

Remotely-sensed data and a hydrologic model can be combined to produce better soil moisture profile than either of these methods used alone. This information about profile soil moisture could be used in many hydrologic and agronomic applications. In this study, we used a hydrologic model with remotely-sensed data and two different schemes to update the modeled estimates. RMSEs increased by 22°/0 as the modeling time step was increased from 6 hr to 12 hr. For a given modeling time step, RMSEs reduced when the modeled soil moisture estimates were periodically updated. RMSEs were reduced by 22% and 49% with method 1 and method 2 updating respectively, for 6 hr time step. Similarly, RMSEs were reduced by 14% and 41% with method 1 and method 2, respectively, for 12 hr time step. These results indicate the need to adjust the model-estimated soil moisture profile to a certain

depth by using remotely-sensed data. However, further studies are needed to obtain the optimal depth to which the blending should be done.

Acknowledgments

This work was supported by Universities Space Research Association (USRA) SUB95-164 under NASA Grant NCCW-0084 with Alabama A&M University's HSCaRS Center. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government. The Institute for Global Change Research and Education is jointly operated by the University of Alabama in Huntsville and Universities Space Research Association. We acknowledge Teferi Tsegaye, Frank Archer, William Belisle, Johnny Boggs, Ron Dupree, Ahmed Fahsi, Garland Robertson, Zack Senwo, Derrick Taylor, Wubishet Tadesse and the HSCaRS support staff for invaluable assistance during the field experiment.

References

Arya, L. M., Richter, J. C., and Paris, J. F., 1983. Estimating profile water storage from surface zone soil moisture measurements under bare field conditions. Water Resour. Res., 19:403-412.

Bernard, R., Vauclin, M., and Vidal-Madjar, D., 1981. Possible use of active microwave remote sensing data for prediction of regional evaporation by numerical simulation of soil water movement in the unsaturated zone. Water Resour. Res., 17:1603-1610.

Biswas, B. C., and Dasgupta, S. K., 1979. Estimation of soil moisture at deeper depth from surface layer data, Mausam. 30:511-516.

Blanchard, B. J., 1979. Correlation of spacecraft passive microwave system data with soil moisture indices (API). Remote Sensing Center, Texas A&M Univ., College Station, TX, RCS-3622-2.

Capehart, W. J., and Carlson, T. N., 1994. Estimating near-surface soil moisture availability using a meteorologically driven soil-water profile model. J. Hydrol., 160:1-20.

Engman, E. T., and Chauhan, N., 1995. Status of microwave soil moisture measurements with remote sensing. Remote Sensing Env., 51:189-198.

Jackson, T. J., 1993. Measuring surface soil moisture using passive microwave remote sensing. Hydrolog. Processes, 7:139-152.

Laymen, C., Belisle, W., Coleman, T., Crosson, W., Fahsi, A., Hood, R., Jackson, T., Luvall, J., Manu, A., O'Neill, P., Senwo, Z., Tsegaye, T., and Wu, S., 1997. Huntsville '96 Microwave Remote Sensing of Soil Moisture Experiment: Overview and Preliminary Results. Preprints of the AMS Conf. On Hydrology Long Beach, CA, 2-7 Feb. 1997. American Meteorological Society, Boston, MA, pp. 184-187.

Smith, M. R., and Newton, R. W., 1983. The prediction of root zone soil moisture with a water balance – microwave emission model. AgRISTARS Rep. SM-T3-04425.

Soman, V. V., Crosson, W. L., and Laymen, C, A., 1997. Modeling soil moisture using passive remote sensing. Conference on Remote Sensing for Agriculture, Ecosystems, and Hydrology at Aerospace Remote Sensing, Institute of Electrical Engineers (IEE), London, U.K.

Table 1: Comparison of root mean square errors from baseline for different sampling intervals.

Simulation	RMSE (6 hr)	RMSE (12 hr)
1. SHM run	0.1550	0.1893
2. SHM run+ remote sensing updates	0.1203	0.1636
3. SHM run+ remote sensing updates with blending	0.0788	0.1125

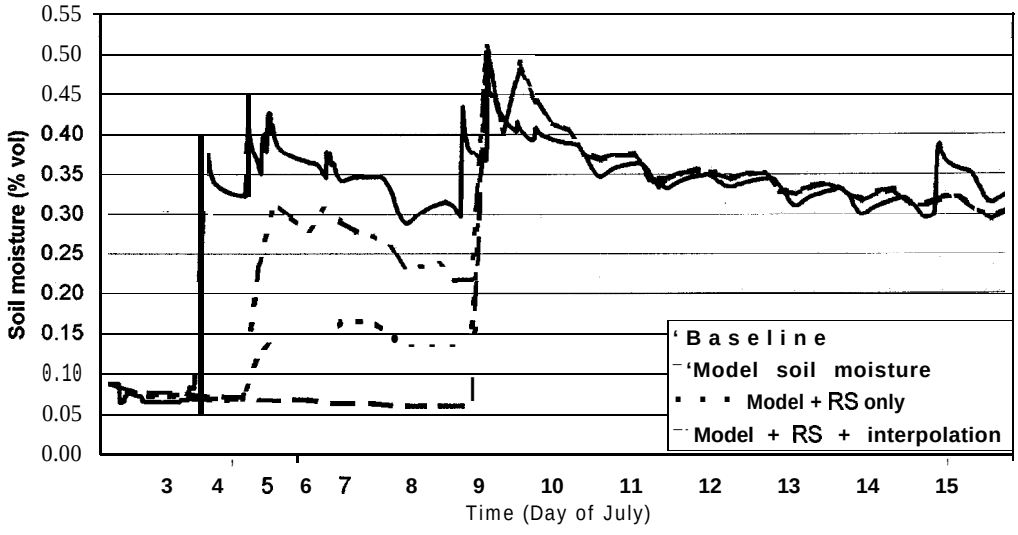


Figure 1: Soil moisture estimation for input interval of 6 hr using (a) model only, (b) model with remote sensing, and (c) model and remote sensing with interpolation.

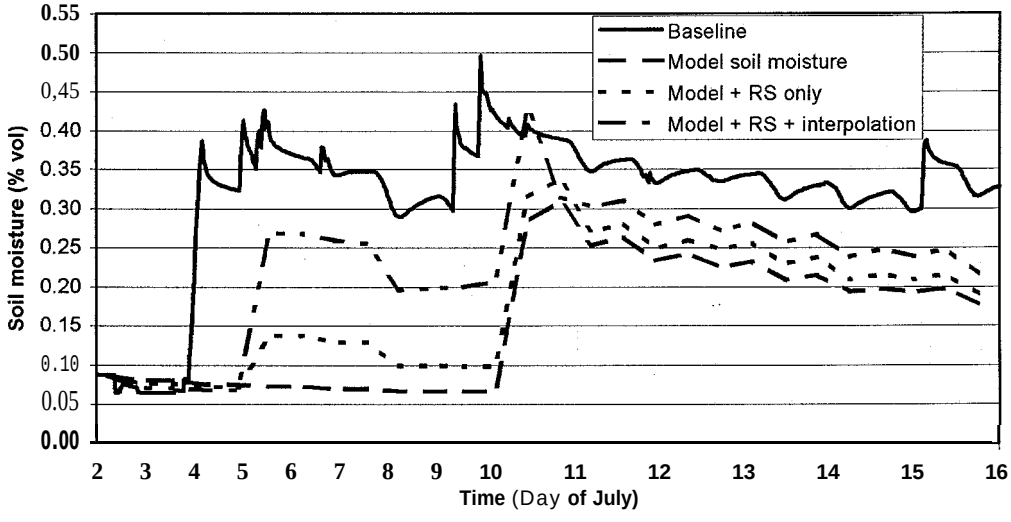


Figure 2: Soil moisture estimation for input internal of 12 hr using (a) model only, (b) model with remote sensing, and (c) model and remote sensing with interpolation.

98 URC063

AN EXAMPLE OF MIXED PROCESSING OF DELTA-SIGMA MODULATED PULSE STREAM

Marcelina Gallegos, Djuro Zrilic, J. Ramirez-Angulo
NASA Center for Autonomous Control Engineering
New Mexico Highlands University
Department of Engineering
E-mail: mm_gallegos@venus.nmhu.edu
Fax (505) 454-3306
Las Vegas, NM 87701

Abstract: This paper is about mixed processing of a delta- sigma modulated pulse stream with an analog signal. To show that mixed analog/digital processing 'is possible, we implemented a system for frequency doubling. Hardware design and relevant signals are presented.

KeyWords: Delta modulation, mixed analog/digital signal processing, frequency doubling

1. INTRODUCTION

Delta-sigma modulation (DSM) was developed over three decades ago, but commercial implementation was achieved in the late eighties and early nineties. Commercially available, VLSI chips utilize over-sampling and digital filtering to achieve high signal to noise ratio at lower costs. The advent of commercial delta-sigma converters is due, in most part, to recent advances in mixed analog/digital VLSI technology. Precision analog circuitry can now be integrated on the same chip with powerful digital filters [1].

Delta-sigma converters available on the commercial market today consist of analog modulators and digital filters, Figure 1.

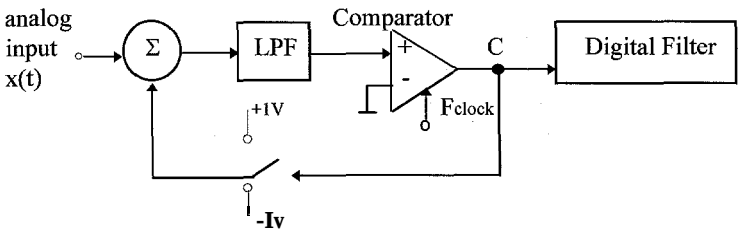


Figure 1: Delta -Sigma ADC with decimation filter

The modulator loop over-samples and processes the analog input signal at a rate much higher than the bandwidth of interest. The modulator's output consists of a very high rate, one bit, pulse density stream. This pulse density stream is not suitable for processing with ordinary digital signal processing (DSP) hardware in direct form and thus reduction in sampling rate is needed. The reason that decimation is employed in Figure 1, is to drop the modulator's output to a more manageable system sampling rate.

In this paper we will employ a new approach of direct processing of DSM pulse density stream. Using the dual nature of DSM pulse density stream, we will show that doubling of the input frequency, of the analog input signal, can be achieved.

2. HARDWARE IMPLEMENTATION

It is a well known fact that demodulation of DSM pulse density sequence is done by a simple collecting of digital samples and averaging them to produce a demodulated output. This means if the length of the averager is long enough, an analog signal will be reconstructed with a certain accuracy. This idea can be employed for mixed analog/digital signal processing proposed in Figure 2.

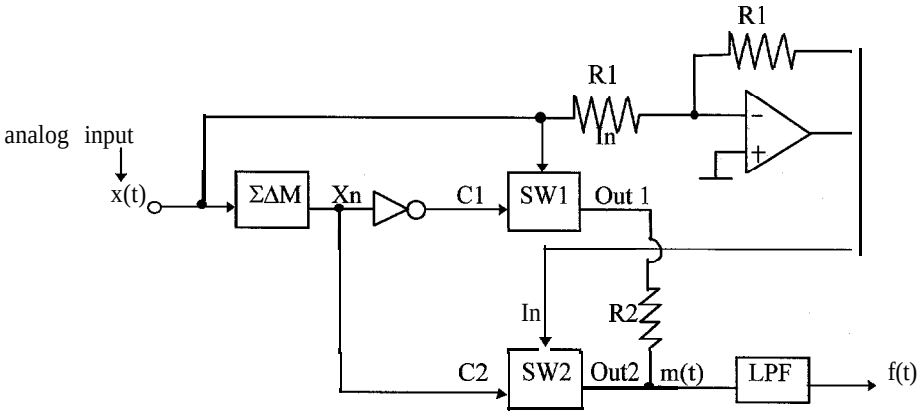


Figure 2: Mixed analog/digital signal processor for frequency doubling

Suppose that the sinusoidal signal x(t) = sinωt, figure 3a, has to be ADC by use of DSM into sequence of binary bits $Xn \in \{-1, +1\}$. This sequence is fed as a control signal of opposite polarity to switches SW1 and SW2. An analog signal of opposite polarity is fed to the inputs of switches SW1 and SW2 as well. In our case a bilateral switch, CD4066, was used in actual hardware design. As a result, the amplitude modulated signal, m(t), is obtained, Figure 3b and after a low pass filtering signal f(t) of double frequency of x(t) is obtained, Figure 3c.

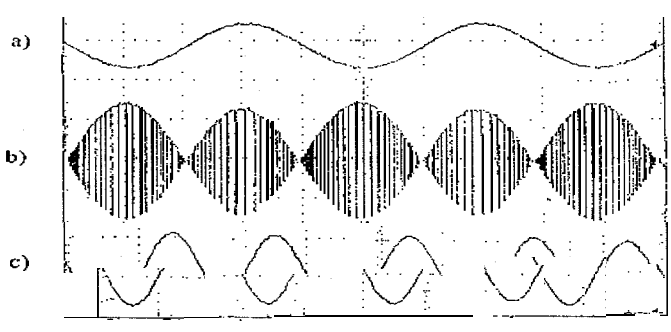


Figure 3: Relevant waveforms from Figure 2. (a) input signal (b) AM signal (c) filtered output

It is easy to show that this holds for sinusoidal inputs because after mixing (multiplying) x(t) with its dual equivalent (DSM pulse stream, Xn), $x(t)*Xn = x^2(t) = \sin^2\omega t = (1-\cos 2\omega t)/2$. In addition to the signal of double frequency, a DC component is present which is easy to remove by a capacitor. Our experimental results in Figure 3 show good agreements between theory and practice [2]. It is easy to envision a cascade of switches to realize a signal processor as Figure 4. A signal of double frequency will appear after the odd mixers outputs. Signal f(t) will have a double frequency of x(t), signal f3(t) will

have a double frequency of $f_1(t)$, etc. This means, after mixers $M_1, M_3, \ldots M_{n+1}$, frequency doubling will occur.

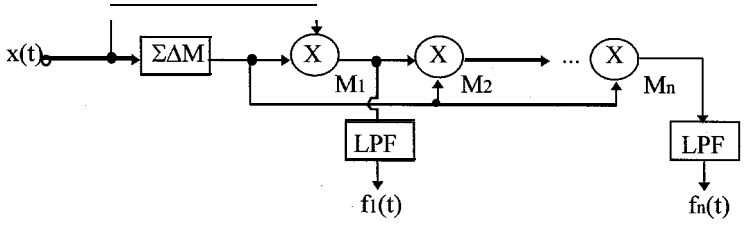


Figure 4: Cascade connection of mixers from Figure 2, for frequency doubling

One can show that analog input of the mixer (doubler) can be any analog signal, y(t). It is possible to process the analog signal, y(t), first (filter, amplify, rectify, etc.), and then mix it with a DSM pulse modulated stream to get the desired signal.

Processing of DSM pulse stream is possible as well [2]. Figure 5 shows a general block diagram of a functional processor.

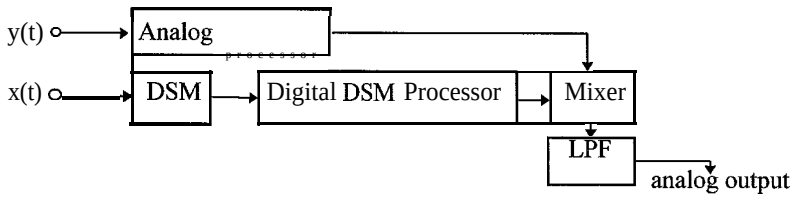


Figure 5: Block diagram of DSM fictional processor

3. CONCLUSION

We have shown that a mixed analog/digital processing of a DSM stream in direct form is possible. Instead of using decimated DSM pulse stream, we proposed a circuit for direct processing. One can easily envision that the number of functions can be realized according to the block diagram in Figure 5.

4. REFERENCES

[1] "Analog/Digital Conversion IC's" - Catalog Crystal Semiconductor Corporation. P.O. Box 17847, Austin, TX 78744.

[2] Zrilic, D; "Mixed Analog/Digital Signal Processing Using Sigma-Delta Modulation" paper submitted to IEEE Transaction on Circuits and Systems for possible publication.

98 URC064

REAL – TIME ACTIVE VIBRATION CONTROL USING PIEZOELECTRIC ACTUATORS IN PLATE STRUCTURES

Yaxi Shen, Abdollah Homaifar, Marwan Bikdash, Ahmed A. Naser

yaxi@ncat.edu, homaifar@ncat.edu, bikdash@ncat.edu, ahmada@ncat.edu

NASA-CAR, Center of Aerospace

Department of Electrical Engineering

North Carolina A&T State University

Greensboro, NC 27411

ABSTRACT: This work presents a control system testbed which is used for active vibration control. Software simulation and real-time experiment are realized at the same time using different control methodologies through dSPACE environment. Localized control with hybrid-fuzzy PD controller and rate-feedback controller are shown to attenuate vibration at resonance frequencies. The result confirmed that these two kinds of control methods are reliable in active damping of the vibration,

KEYWORDS: PZT patches, Modeling, Active vibration control, Rate-feedback, Fuzzy logic

1. INTRODUCTION

Launch vehicles often experience serious vibration problems during the ascent phase due to intense acoustic fields and mechanically transmitted disturbances. These disturbances are characterized by a large number of resonances over a wide frequency band. The payload inside the launch vehicle must be designed to withstand these structural vibrations. Passive isolators are used to isolate the payloads from vibrating, but this kind of isolation has no effect on the acoustic excitation. This study focuses on methods for achieving active damping on plate structures by means of piezoelectric actuators (PZTs). We conduct an experiment using PZT actuators and accelerometers mounted on a rectangular plate with an active damping controller implemented on a PC machine, appropriate closed-loop controller (dSPACE controller) are designed and the quantitative amount of active attenuation is measured. Preliminary test results indicated that this technique can lower the loads on payloads and permit a reduction of the size, weight and cost of the structure. A typical spacecraft or launch vehicle structure consists of a series of thin panels that are clamped around their individual boundaries by a continuous metal frame. Avionic equipment mounted onto these panels can be exposed to excessive vibration due to large acoustic emanations during launch. The plate that is used in our experiment resembles some of the panels that cover a space vehicle during launch. Electrically actuatPZT elements mounted on the plates can create mechanical force and moments to attenuate the acoustic vibration and protect the space payload during launch. A voltage applied across a PZT can create a plate deformation. The application of an electric field along the polarization direction of thePZT forces it to expand along the directions perpendicular to the electrical field, producing a local strain on the plate surface, on the other hand, if the electric field is opposite of the polarization direction, the PZT will constract in the direction perpendicular to the electrical field. This expansion or constriction of a PZT element bonded on a plate generates local moments on the plate surface. If the applied voltage is kept at a low level (well below the depoling level), the relationship between the local moment andPZT voltage is linear.

In our experiment, there are fivePZT patch pairs which are bonded on the plate. Each pair consists of a top patch which is bonded on the top side of the plate, and an identical patch bonded on the bottom side of the plate, directly below the top patch, The two patches sandwich the plate and constitute one actuators. If we use the proper orientation of PZT polarization axis and input voltage, a doubling of force can occur (Figure 1)(Falangas, Dworak and Koshigoe, 1994 and Crawley and de Luis, 1986). “

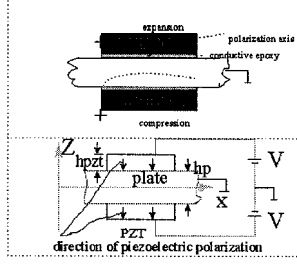


Figure 1: Principle of PZT operation

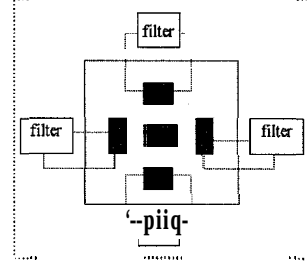


Figure 2: Localized control

2. PLATE MODEL

The structure was modeled from experimental data as well as analytically from first principles. The complete derivation of equations of motion of the plate can be found in (Nagle, Homaifar, Abdelnaser and Bikdash, 1996). We only present a synopsis. According to the force and moment analysis, we can get the equation of motion for the plate with multiple PZT patches from equilibrium equations of force and moments (if we assume the the density of the plate is approximately the same as that of the PZT, and the thickness of the PZT is negligible compared with that of the plate) :

$$D\{\frac{\partial^4 w}{\partial x^4} + 2 - + \frac{\partial^4 w}{\partial y^4} + \alpha\psi \sum_{p=1}^{**...} [\frac{\partial^2}{\partial x^2}\left(\chi_p\left(\frac{\partial^2 w}{\partial x^2} + v\frac{\partial^2 w}{\partial y^2}\right)\right) + 2(1-v)\frac{\partial^2}{\partial x \partial y}\left(\chi_p \frac{\partial^2 w}{\partial x \partial y}\right) + \frac{\partial^2}{\partial y^2}\left(\chi_p \left(\frac{\partial^2 w}{\partial y^2} + \nu\frac{\partial^2 w}{\partial x^2}\right)\right)]\} + \rho_p h_p \ddot{w} = \sum_{q=1}^{N} F_q + \frac{Y_{pz}}{1-\nu} \sum_{p=1}^{M} h_t d_{31}\left(\frac{\partial^2}{\partial x^2} + \nu\frac{\partial^2}{\partial y^2}\right)(\chi_p V_p) \tag{1}$$

where $D = Y_p h_p^3/12(1-\nu_p^2), \alpha = 8\xi^3 + 12\xi^2 + 6\xi, \xi = h_{pz}/h_p, h_{pz}, h_p$ is the thickness of PZT and plate, h_t is total thickness of the plate and the PZT patches, w is displacement of vibration for the plate, @ $Y_{pz}/Y_p, Y_{pz}, Y_p$ is Young's modulus of the PZT patches and the plate, ν is the Poisson's ratio, χ_p = 1 or O, when the PZT pan' p is present or not, the subscript p denotes the patch pair number, M is the total number of patch pairs, F_q is the qth lateral force applied to plate, ρ_p is the density of the plate, d_{31} is the charge coefficient.

Consider the free vibration problem (Vp=0 and Fq=0). Assumed W(x,y) to be separable in the x and y directions (Elishakoff, 1977):

$$w(x,y,t) = W(x,y)e^{jwt}, W(x,y) = X(x)Y(y) \tag{2}$$

Assume that the plate is homogeneous such that the elastic effect of PZT patches is distributed over the entire plate, an approximate solution for the eigenvalues and eignfunctions of equation with the appropriate boundary conditions are given by

$$\omega = \left(\frac{D}{\rho_p}\right)^{1/2}\left[\left(\frac{\bar{r}\pi}{a}\right)^2 + \left(\frac{\bar{s}\pi}{b}\right)^2\right] \tag{3}$$

$$X_{rs}(x) = \cosh\frac{\theta\pi x}{a} - \cos\frac{r\pi x}{a} - \frac{C_1 - c_1}{S_1 - s_1}\left(\frac{\theta\pi x}{\theta/\bar{r}a} - \frac{\theta}{\bar{r}}\sin\frac{\bar{r}\pi x}{a}\right) \tag{4}$$

$$Y_{rs}(y) = \cosh\frac{\alpha\pi x}{b} - \cos\frac{\bar{s}\pi x}{b} - \frac{C_2 - c_2}{S_2 - s_2}\left(\frac{\alpha\pi x}{b} - \frac{a}{\bar{s}}\sin\frac{\bar{s}\pi x}{b}\right) \tag{5}$$

$$\Psi_{rs}(x,y) = X_{rs}(x)Y_{rs}(y) \tag{6}$$

where a and b are the plate length along the x-axis and y-axis. Cl, $c_1, C_2, c_2, S_1, s_1, S_2, s_2, \theta$, et, δ are defied as:

$$\begin{aligned} &c_1 = \cos\bar{r}\pi, s_1 = \sin\bar{r}\pi, C_1 = \cosh\theta\pi, S_1 = \sinh\theta\pi \\ &c_2 = \cos\bar{s}\pi, s_2 = \sin\bar{s}\pi, C_2 = \cosh\text{ cm}, S_2 = \sinh\alpha\pi \\ &\theta = \sqrt{\frac{2\bar{s}^2}{\delta^2} + \bar{r}^2}, \alpha = \sqrt{\frac{2\bar{r}^2}{\delta^2} + \bar{s}^2}, \delta = \frac{a}{b} \end{aligned} \tag{7}$$

For a plate clamped all around $\bar{r}$ and $\bar{s}$ which correspond to mode r in the x-direction and modes in the y direction are the root of the following transcendental equations :

$$1 - c_1C_1 + \frac{\bar{s}\delta}{\bar{r}}\left[2 + \frac{\bar{r}^2}{\bar{s}^2\delta^2}\right]^{1/2} s_1S_1 = 0 \tag{8}$$

$$1 - c_2C_2 + \frac{r}{\bar{s}\delta}\left[2 + \frac{\bar{s}^2\delta^2}{\bar{r}^2}\right]^{1/2} s_2S_2 = 0 \tag{9}$$

For a simply-supported plate , $\bar{r} = r, \bar{s}$ = s. After applying orthogonality conditions and expressing the solution as

$$w(x,y,t) = \sum_{j=1}^{\infty} \Psi_j(x,y)\eta_j(t) \tag{lo}$$

where Ψ_j is the eigenfunction of free vibration, η is the generalized displacement, we obtain the following infinite-dimensional system of equations for forced vibration:

$$\ddot{\eta}_j + 2\zeta_j\dot{\eta}_j + \omega_j^2\eta_j = \sum_{q=1}^{N} F_{jq} + \sum_{p=1}^{M} Q_{jp}V_p \tag{11}$$

Table 1: Comparison of experimental and theoretical natural frequencies

Mode #	Experiment	Theory	Diff. %
1(1,1)	26.73	26.73	1.982
2(2, 1)	47.11	47.11	1.995
3(1,2)	70.37	70.73	2.003
4(3,1)	81.92	82.15	2.296
5(2,2)	101.23	89.96	-9.349
6(4,1)	129.91	131.2	3.017
7(3,2)	137.37	123.7	-8.129
8(1,3)	140,15	138,0	0,439

where F_{jq} *is* force applied at position q, V_p is the voltage applied to PZT p, j is the mode number. If the first eight modes are selected for experiment, it is given:

$$\tilde{M}_{8\times8}\ddot{\eta}_{8\times1} + \tilde{C}_{8\times8}\dot{\eta} + \tilde{K}_{8\times8}\eta = \tilde{F}_{8\times1}F(t) + \tilde{D}_{8\times5}V_{5\times1} \tag{12}$$

where $\tilde{M}$ is mass matrix, $\tilde{C}$ is damping ratio matrix, and $\tilde{K}$ is spring stiffness matrix. According to the above derivation, we can represent the system using the state-space equations:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & I \\ -\tilde{K} & -\tilde{C} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{D} \end{bmatrix}V + \begin{bmatrix} 0 \\ \tilde{F} \end{bmatrix}F(t) \tag{13}$$

Our analysis mainly concentrated on the fist eight modeshape. The reasons are:

- Lower-frequency modes have much higher vibration amplitudes than the higher-frequency modes. they are then the main contributor to vibration (Falangas, Dworak and Koshigoe, 1994).
- The actuator/sensor collocation guarantees that the system is positive real and stable regardless of the gain value at the lower-frequency mode. This guarantee does not apply to higher-frequency modes because the collocation principle doesn't apply to a mode whose wavelength is comparable to the size of the PZT actuator. In addition, phase lags and computation delays can drive some higher frequency modes unstable.

The natural frequency ω_j of the system is determined experimentally using an impulse hammer test. It can also be calculated using formula (Table 1). It was noticed that the theoretical natural frequencies did reasonably and closely match those obtained experimentally, This is because in a real life situation the boundary conditions on a panel can never perfectly clamped nor simply supported.. Instead we modeled the boundary conditions as a combination of the clamped and simply-supported boundary conditions, we use a weight function for each mode to account for the contribution of the constrained rotation on the boundaries (Nagle, Homaifar, Abdelnaser and Bikdash, 1996).

3. CONTROL DESIGN

3.1 Control objectives, configurations and design methods

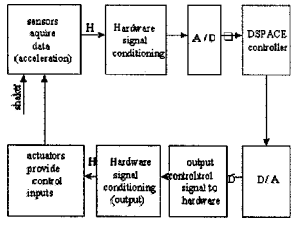


Figure 3: Rate-feedback control system

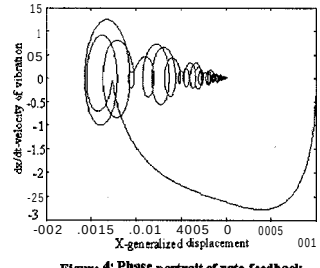


Figure 4: Phase portrait of rate-feedback control system

We want (1) to reduce vibration at a given frequency or a set of frequencies, (2) to control vibration without

"spillover" into uncontrolled modes, (3) to be able to change the frequencies to be controlled without redesign, and (4) at same time remain robust in face of model uncertainty. We prefer to use localized control for system configuration (Figure 2)(Dworak, Falangas, Koshigoe and Tsheng, 1992). An accelerometer is used as feedback to a collocated PZT actuator. This configuration a simple concept because no coordinate transformation to decouple modes is utilized. As such, It is more suitable for application where there is not sufficient knowledge about the structural model. We try two design methods as follows:

Rate feedback : The rate feedback design approach is to integrate the accelerometer signal that sense the motion of the plate, amplify it, and send it to the appropriate PZT for actuation. The integration process shifts the phase by 90 degree in order to provide a rate feedback. This effect increase the damping in all controllable/observable modes (Figure 3).

Stability analysis of rate feedback Our control methods are based on the idea of symmetric output feedback. Now we will give the stability of closed-loop, consider the equation of motion(Junkins and Kim, 1993).

$$\tilde{M}_{8\times8}\ddot{\eta}_{8\times1} + \tilde{C}_{8\times8}\dot{\eta} + \tilde{K}_{8\times8}\eta = \tilde{F}_{8\times1}F(t) + \tilde{D}_{8\times5}V_{5\times1} \tag{14}$$

where V is control vector, M is n x n positive definite symmetric mass matrix, C is n x n positive semidefinite symmetric structural damping matrix, K is n x n positive semidefinite symmetric stiffness matrix, and D is n x m control influence matrix. We introduce the following symmetric output feedback form of the control law (collocated sensors and actuators):

$$\mathrm{V} = -(G_1 y + G_2 \dot{y}), y = D^T\eta \Rightarrow V = -(G_1 D^T\eta + G_2 D^T\dot{\eta}) \tag{15}$$

G_1, G_2 are fully populated, m x m positive definite symmetric gain matrices. with the assumption that the sensors and actuators are perfectly linear and instantaneou,the closed-loop equations can be written as

$$M\ddot{\eta} + (C + DG_2D^T)\dot{\eta} + (K + C + DG_1D^T)\eta = O \tag{16}$$

Since G_i *are* positive definite, then the control induced stiffness and damping perturbations DG_iD^T symmetric positive semidefinite. We have shown that for all members of this special class of output feedback controllers, asymptotic stability of closed-loop system is guaranteed. An important consequence of the stability analysis is that the stability of a controllable symmetric system (having symmetric positive definite output feedback) is maintained

- Regardless the number of modes retained in the reduced order model
- Regardless of in accuracies in the parameter values used in the design

This result is attributable to the fact that since positive definiG1 and G2 are chosen, M,$K+DG_1D^T$,$C+DG_2D^T$ *are* always positive definite regardless of full order or reduced order modal are used, Thus, by using the symmetric output feedback controller. The potential instability problems due to spillover can be completely avoided and the parameters do not have to be known accurately to guarantee stability. This closed-loop stability analysis leads us to the conclusion that if the system is controllable, the the closed loop system is at least asymptotically stable so long as the gain matrices are chosen as positive definite matrices. Phase portrait is shown in Figure 4.

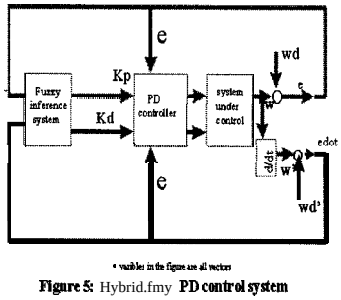


Figure 5: Hybrid.fmy **PD control system**

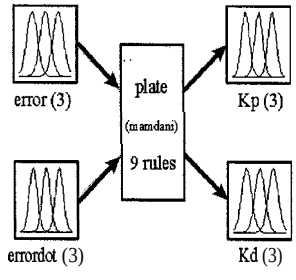


Figure 6: Fuzzy inference system

Fuzzy **PD** controller : We design a hybrid fuzzy-PD controller based on regular PD controller and simulate it in software (preparation for our real-time experiment). We select the displacement error and its time derivative as input and individual proportional, and derivative coefficients of the PD controller as outputs for the fuzzy inference system. Appropriate membership functions associated with fuzzy sets are constructed to ensure a smooth transition between neighboring fuzzy sets, and a set of fuzzy rules is developed (Figure 5). Figure 6 reflects the names and characterizing parameters (i.e. e, edot and Kp, Kd) of each fuzzy set (Homaifar, Sayyarrodsari and Snyder, 1994). To design the rule base we utilized our knowledge about the system under control, as well as the existing relationship between P and D coefficients and the performance of the plate system. Table 2 reflects the developed rule base for our simulation where L, M and S stand for large, medium and small respectively(Homaifar, et al., 1994).

Table 2: Fuzzy rule table

$\frac{dE}{dt}$ \ ERROR	N	ZE	P
N	L	M	s
ZE	M	s	s
P	s	M	M

Kp

$\frac{dE}{dt}$ \ ERROR	N	ZE	P
N	L	M	s
ZE	M	s	s
P	s	M	M

Kd

4. Results and conclusions

The rectangular plate we use is .05" thick aluminum with edges of 28 " and 21". The frame of the structure clamps the plate between two 1" square aluminum members. The frame is supported at the corners with legs of angle aluminum which are anchored to a benchop, preventing any rigid body motion. We use accelerometer(PCB SN 1722) to detect the signal. The signal then flows to a band-pass filter which is implemented using Simulink. All control algorithms are realized through dSPACE environment. The signal finally passes through a PCB Piezotronics 790 series power amplifier before arriving at the actuators. The result of real-time vibration control using rate feedback controller is shown in Figure 6. The simulation of vibration control using hybrid fuzzy-PD controller is shown in Figure 7. From the results of the experiment, we can see that the control design with rate feedback is relatively simple, and the controller spreads the control authority to all modes over a wide frequency range. This is helpful when your controlled object is uncertain.

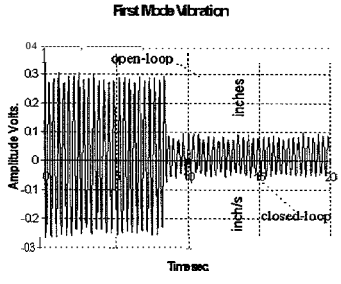


Figure 7: Real-time vibration control using rate-feedback controller

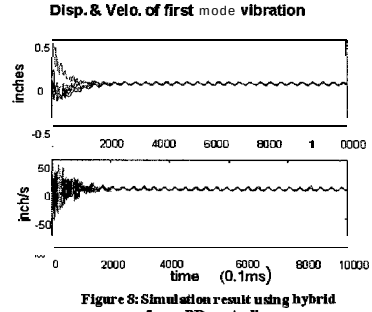


Figure 8: Simulation result using hybrid fuzzy-PD controller

This paper has described the modeling of a controlled plate to be used for the comparison of various control methodologies. Localized control configuration using Rate-feedback controller and Hybrid-fuzzy PD controller were shown to attenuate vibration at several natural frequencies. Stability analysis of rate feedback using this plate model predict superior real-time performance of rate feedback controller for a broad band disturbance. An simulation result confirmed that Hybrid fuzzy-PD controller is reliable in active damping the vibration.. The real time control using Hybrid fuzzy-PD controller will be implemented in the future, ultimately we will compare the performance of these controller when faced with model parameter uncertainty

ACKNOWLEDGMENT

We would like to thank Eric A. Wheater, Ph.D candidate of the Structures Group of NASA Core at North Carolina A&T State University for his assistance and involvement. This work is partially funded by grants from the NASA Center of Aerospace at North Carolina A&T State University under grant number NAGW-2924.

REFERENCE

1. J. L. Junkins, and Y. Kim, Introduction to Dynamics and Control of Flexible Structures, AIAA, Education Series, Washington, DC, 1993.
2. E. T. Falangas, J. A. Dworak, and S. Koshigoe, "Controlling Plate Vibrations Using Piezoelectric Actuators," IEEE Control System, 1994, pp. 34-41.
3. J. A. Dworak, E. T. Falangas, S. Koshigoe, and G. T. Tsheng, "Vibro-acoustic Control Using PZT Actuators," presented at *the 3rd Int. conf. Adaptive structures in* Nov.92, *San* Diego, CA, pp. 529-545.
4. J. C. Nagle, A. Homaifar, A. Abdelnaser, and M. Bikdash, "Modeling and Control of a Thin Plate," *Proceedings of the 1st National student conference, The* National Alliance of NASA University Research Centers at Minority Institutions, Editors : E. O. Daso, and S., Mebanee, TRI Press, March 31, April 2, 1996, pp. 362-365.

5. A. Homaifar, B. Sayyarrodsari, and W. E. Snyder, "A Theoretical Justification for Nonlinear Controlling Property of a class of Fuzzy Logic Controllers, " *Third IEEE International Conference on Fuzzy Systems,* 1994. pp. 382-386.
6. A. Homaifar, B. Sayyarrodsari, J. E. Hogans,'< Fuzzy Controller for Robot Arm Trajectory", *Journal of Information Sciences, 1994.*
7. E. F. Crawley and J. de Luis, "Experimental Verification of Distributed Piezoelectric Actuators for Use in Precision Space Structures," *Proc. 27th AIAA Struct. Dyn. Conf., 1986,* pp. 116-124.
8. I. Elishakoff, "Random Vibrations of Orthopic Plates Clamped or Simply Supported All Round," *Acts Mechanica, 1977,* Vol. 28, pp. 165-176.

98 URC065

EVALUATION OF $Ba_{.65}Sr_{.35}TiO_3$ MIS STRUCTURE FOR SEMICONDUCTOR MEMORY APPLICATIONS

K. Washington, F. Brown, T.N. Fogarty, H. McWhinney, F.C. Wang, R. Wilkins
PVAMU-CARR

As device sizes shrink, new demands are made on device materials. A major factor in the reliability and radiation tolerance of MOS devices is the performance of the insulating "oxide" layer. High dielectric constant materials such as ferroelectrics are being investigated as a replacement for SiO_2 in these devices. The advantages are expected to be increased charge storage capacity, low operating voltages, long term endurance and greater radiation tolerance. Some possible applications are as capacitive storage elements in DRAM cells or as voltage tunable capacitors in filters, oscillators and other circuits. This paper examines and explores the potential for such materials with emphasis on their applications as storage elements in DRAM cells and by examining the effects of ionizing radiation on these materials. The samples used in this study were composed of $Ba_{.65}Sr_{.35}TiO_3$ and laid on bare silicon using a metal organic decomposition technique. Capacitance-Voltage measurements were made to characterize the MIS defects as a function of ionizing radiation. These measurements indicated the presence of deep insulator traps, change in the interface structure, possible diffusion between the silicon and barium strontium titanate layers and the presence of mobile ions. A surface depth profile was performed to aid in the verification of these results.

OBJECTIVES

Due to the increasing role that barium strontium titanate will play in the development and production of future devices, it is the purpose of this study to investigate and analyze the properties of this material. This was done by performing C-V analysis on an aluminum dot metal insulator semiconductor (MIS) sample. The insulator was BST and the semiconductor was n-doped silicon. Previous studies have utilized ferroelectric thin film BST with a barrier layer (usually platinum) between the BST and the silicon wafer. Our study differs from these studies because our sample has paraelectric BST laid on bare silicon without a barrier layer.

The objective of this study was to electrically characterize the MIS sample and to examine the effects of ionizing radiation on it. The following tasks were performed in order to achieve this objective: 1) high frequency C-V measurements on the MIS sample. 2) low frequency C-V "measurements on the MIS sample. 3) the examination of the flat baud voltage, V_{FB}, characteristics. 4) the examination of the density of interface states, D_{IT}, characteristics. 5) the investigation of the effects after the exposure to radiation of the MIS sample. 6) mobile ion tests on the MIS sample and 7) a surface depth profile.

DRAM AND MOS CAPACITOR OPERATION

A dynamic RAM cell uses dynamic charge storage nodes to hold data. Atypical DRAM cell consists of a capacitor for charge storage and a MOSFET for charge switching. This configuration is called 1 Transistor -1 Capacitor (1T - 1C) and is very popular due to its simplicity and small size. The cell is accessed by turning on the MOSFET via the word line. Charge is then transferred into/out of the capacitor. The information stored in the cell is read destructively and subsequently restored by a sense amplifier.

For a given gate voltage, the charge decreases with decreasing area. Higher density memories have smaller-area memory cells, and various techniques are employed to increase the total charge per cell. One of these techniques is the use of double ion-implanted regions in the high-capacity (Hi-C) DRAM cell [1]. Ion implantation is a method of introducing dopant atoms of very precisely controlled depth and concentration into a semiconductor. Acceptor atoms are implanted into the storage region of the Hi-C DRAM cell. The resulting higher doping concentration increases the space charge region capacitance, but

it does not increase the charge-storage capacity of the cell, because the surface potential difference between an empty and a full potential well decreases. The surface potential change due to the donors, in conjunction with the increased space charge region capacitance due to the acceptors, results in an enhancement of charge-storage capacity of typically 50°/0 over a conventional cell.

The Hi-C DRAM achieves its higher charge-storage capacity by virtue of a clever dopant scheme. The trench or corrugated capacitor relies on a geometrical factor. A hole is etched into the semi-conductor substrate prior to cell fabrication. Next, an oxide is grown on the surface in the hole and subsequently the hole is filled with a polysilicon storage electrode.

Capacitance offered by a cell of effective area A and effective dielectric thickness t, is given by the formula

$$c = \epsilon_0 \epsilon_r (A/t) \qquad (1)$$

where ϵ_0 is the permittivity of free space and ϵ_r is the relative permittivity or the dielectric constant of the material. SiO_2 has a dielectric constant of 3.9. Currently used material is a combination of SiO_2 and Si_3N_4. This layer is referred to as Oxy-Nitride (ON) or Oxy-Nitride-Oxide (ONO) and has a dielectric constant of about 6. It is estimated that the minimum capacitance that the cell should offer for the satisfactory storage of charge and operation of the DRAM cell is 30 $fF/\mu m^2$. This value is called the Critical Capacitance C_{crit}. Barium Strontium Titanate with a dielectric constant in excess of 200 is projected to be a replacement for ON and ONO layers.

The need for denser memories has thus placed a greater constraint on the materials employed. Barium Strontium Titanate ($Ba_{1-x}Sr_xTiO_3$) has been projected to be an effective replacement and a material of choice for future generations of DRAM. The other leading possible materials include strontium titanate and lead ziconium titanate.

Lets examine the MOS capacitor. An oxide layer is grown on top of a p-type semiconductor and a metal/polysilicon contact is placed on the oxide. In general, the oxide could be any large bandgap insulator. The main purpose of the oxide layer is to provide an isolation between the metal/polysilicon and the semiconductor so that there is essentially no current flow between the top metal contact and the metal contact at the bottom of the semiconductor.

When the capacitor is formed, the Fermi levels align themselves so that there is no gradient. The conduction band of the oxide is at a higher position than the conduction band of the semiconductor. There is a fairly high potential barrier between the metal/polysilicon and the semiconductor. At zero applied bias, the bandbending at the semi-conductor is determined by the difference in the work functions of the metal/polysilicon and the semiconductor. If one applies an external bias, the bandbending can be removed by applying a compensating bias V_{fb} given by

$$V_{fb} = \phi_m - \chi_s - E_c + E_F$$

$$= \phi_m - \phi_s = \phi_{ms} \qquad (2)$$

where ϕ_m and ϕ_s are the metal/polysilicon and semiconductor work functions, ϕ_{ms} is their difference, χ_s is the semiconductor electron affinity, E_c the semiconductor conduction bandedge, and E_F the Fermi level. Starting from the flat band position, there are three important regimes of biasing in the MOS capacitor. These are accumulation, depletion, and inversion [2].

SYNTHESIS AND PREPARATION OF SAMPLE

The samples in this study were initially prepared by Pradeep Jana, a graduate student at Texas A&M University under the supervision of Dr. R.K. Pandey. Three 2-inch n-doped silicon wafers were purchased and a layer of BST was placed on the bare silicon using a metal organic decomposition technique. In order to complete the construction of the MIS devices, metal contacts had to be placed upon the wafers. This was performed at the University of Houston. Aluminum metal contacts were placed on two of the three samples. A mask was used to evaporate aluminum dots .025 inches in diameter on the top BST layer. In addition, a solid aluminum back contact was placed on the silicon substrate. During this preparation, extreme care was taken with the samples to avoid contamination.

The metal organic decomposition process is a technique for producing inorganic films without processing in vacuum or going through a powder step. A metal-organic compound has a metal atom bonded to an organic radical through a hetero-atom like oxygen, nitrogen, sulfur or phosphorous. The individual organometallic compounds are mixed in cation stoichiometry to form a true solution. This solution is then deposited on a substrate by any suitable method [3]. Our selected composition was $Ba_{.35}Sr_{.65}TiO_3$. The substrate was then heated to pyrolyze the formulation thereby getting rid of the solvent and then to decompose the organometallic compound to produce an inorganic film. A further heating step apropos the annealing was required to control the physical properties including stoichimometry. Multiple processing was performed in order to achieve a greater thickness, prior to annealing.

EXPERIMENTAL PLAN

Once the BST sample was prepared, the Keithley Model 82-DOS Simultaneous C-V System in Prairie View's Center For Applied Radiation Research, (CARR), Lab was used to perform the required measurements. Initial baseline measurements were performed on the sample. The BST sample was then exposed to 7.5 krads of total dose gamma radiation at Texas A&M's Center for Veterinary Medicine. Measurements were taken 1 hour after radiation exposure, and every 30 minutes until 4 hours after radiation exposure. Measurements were taken once again 3 days after radiation exposure. A mobile ion test was performed on the sample and a surface depth profile was performed.

DISCUSSION OF RESULTS

Following exposure to ionizing radiation and after the radiation-generated holes have had time to complete their transport through the oxide, MOS strictures typically exhibit a negative voltage shift component, ΔV_{OT}, in their electrical characteristics (say VT or V_{FB}) that is not sensitive to silicon surface potential and that persists for hours to years. This long-lived radiation effect component is the most commonly observed form of radiation damage in MOS devices and is attributed to the long-term trapping of net positive charge in the insulating layer, i.e. deep insulator traps. This effect generally dominates other radiation damage processes in MOS structures, including negative charge (electron) trapping and interface trap buildup effects [4].

The C-V measurements made on our MIS sample before and after radiation exposure indicate shifts in the high frequency C-V curve and flatband voltage. These shifts are consistent with the changes in electrical characteristics created by deep insulator traps and suggests the presence of these traps in our sample, (see Figure 1).

When silicon is thermally oxidized, the interface between the amorphous oxide and the crystalline silicon is generally deficient of oxygen (or abundant in Si), giving rise to strained as well as uncompleted, or "dangling" silicon bonds. These dangling bonds act as interface traps with energy levels within the forbidden bandgap at the SiO_2/Si interface. When MOS samples are exposed to ionizing radiation, additional interface traps can be generated.

Interface traps seriously affect the performance of MOS devices. When interface traps are generated by irradiation, the shape of the current-voltage characteristics are affected. Before irradiation, the areal density of interface traps in good modern devices is $< 10^{10}$ traps/cm^2, a density such that interface traps are normally not much of a problem. Upon irradiation, however, interface traps can buildup to a significant level resulting indiscernible and often detrimental effects in devices[4].

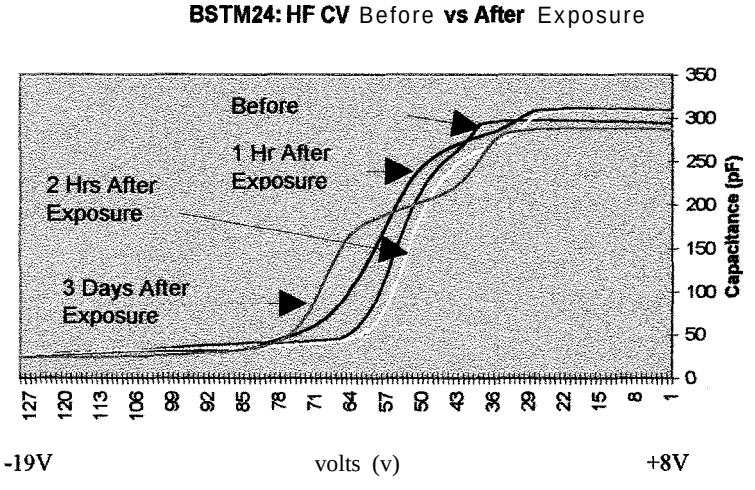


Figure 1

In our MIS device, the density of interface traps before radiation is already at a high level, 10^{13} to 10^{14} traps/cm^{2}eV. This is due in part to the way the BST matches up with silicon at the interface and due in part to the thickness of the BST layer. After radiation exposure, the level of the density of interface states did not significantly change, but a shift occurred in the density of interface traps versus energy from mid-gap. The shift in the D_{IT} versus energy from mid-gap while the level of density of interface traps remains relatively unchanged suggests a change in the interface structure and possibly capture cross-section.

A standard mobile ion test was performed on our MIS sample. The results of this test confirmed the existence of mobile ions in the sample. It is possible that the mobile ions. present could be sodium ions, however, it should be noted that several previous MOS samples have been handled in our lab with no evidence of sodium contamination. Another possibility is that of diffusion of barium and strontium at the insulator/semiconductor interface. This is one of the reasons that most researchers in this area have used samples with a barrier layer such as platinum between the BST and silicon. This prevents contamination of the BST with silicon by interdiffusion. Also, most investigators have used a proven stable ferroelectric phase of BST rather than the paraelectric phase studied here. The disadvantage to this, however, is that now you have a metal insulator metal, (MIM) capacitor in which you will not seethe C-V characteristics of accumulation and depletion, but only a constant value of capacitance.

In order to better understand the composition of our sample, to identify the type of mobile ions present and to verify the existence of two phase material a surface depth profile (SDP) was performed, The SDP revealed the presence of silicon through to the surface, (see Figure 2). It also indicated that the atomic concentration of barium and strontium decreased, but did not go to zero as one goes deeper from the interface into the sample, (see Figure 3). These results meant that our sample did not posess a distinct interface and verified the two phase behavior observed from C-V measurements.

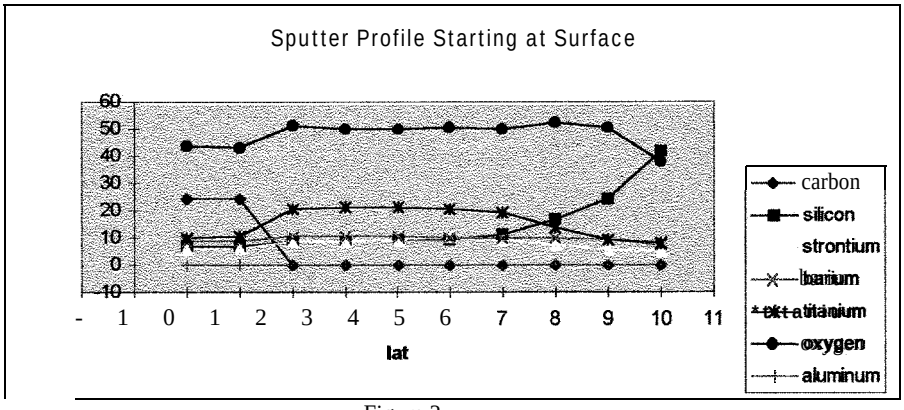


Figure 2

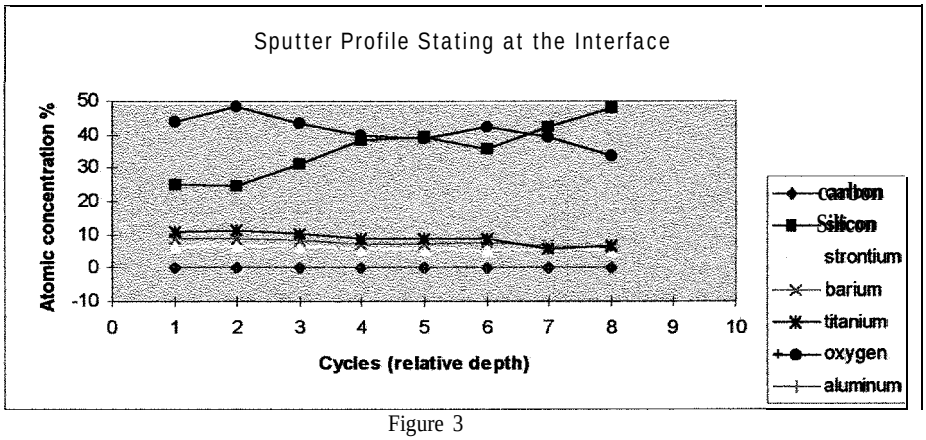


Figure 3

CONCLUSIONS

A MIS sample with paraelectric BST laid on bare silicon without a barrier layer was electrically characterized by performing high frequency and quasi-static/low frequency C--V measurements. The flatband voltage and density of interface states characteristics were examined. The MIS sample was exposed to gamma radiation from a Cobalt 60 source and the effects of this radiation exposure were investigated. A mobile ion test and a surface depth profile were performed on the sample.

The following conclusions were made based on the results of the experiments:

1. Comparison of parameter measurements made before radiation exposure with the parameter measurements made 3 days after radiation exposure indicate a shift in the high frequency C-V plot. A shift in the voltage value where the quasi-static/low frequency minimum occurs takes place. On average this was a shift of +1.24V. A shift in the flatband voltage occurs. On average it has shifted -1.39V. The shifts in these parameters suggest the presence of deep insulator traps.

2. Measurements taken on some of the aluminum dot capacitors show the presence of two inflection points on the high frequency C-V curve. This suggested the behavior of a dual dielectric insulator.

3. In examining the D_{IT} versus energy from midgap, a shift occurred after radiation exposure. On average the shift is a negative 22.8 meV. Interface states are already at a high level, 10^{13} to 10^{14} traps/cm^{2}eV, however, the small change observed in D_{IT} in this range maybe due to experimental error. The change in energy structure from midgap indicates change in the silicon dangling bond structure at the interface and possibly a change in capture cross-section as reported by Fogarty and Mau [5].

4. Examination of the C-V curves as taken at zero bias, after +10V bias, and after -10V bias indicate that mobile ions (probably barium or strontium) are present in the sample. The behavior of the high frequency curve shows a relative negative shift at 1 hour after radiation exposure, but at 2 hours after it has rebounded positive with respect to the before exposure C-V curve. One might believe that this is due to rebound phenomenon, however the fact that there is little change of interface states with radiation suggests that mobile ion movement or silicon BST interdiffusion is a more probable cause.

5. The results of the surface depth profile measurements confirm the lack of a distinct interface. It gives substantial evidence of silicon through to the surface. There is also evidence of barium movement into the silicon. This confirms the presence of two phase material indicated by C-V measurements.

REFERENCES

1. D.K. Schoder, "Advanced MOS Devices", Modular Series on Solid State Devices, Gerold W. Neudeck, Robert F. Pierret, Editors, Addison-Wesley Publishing Company, 1987.

2. J. Singh, "Semiconductor Devices An Introduction", McGraw-Hill, Inc. 1994.

3. P. Jana, R.K. Pandey, "Barium Stontium Titanate Thin Films For DRAM Applications", Conference Proceedings, NASA University Research Centers Technical Advances in Education, Aeronautics, Space, Autonomy, Earth and Environment, Vol. 1, 1997.

4. T.P. Ma and Paul V. Dressendorfer, "Ionizing Radiation Effects In Mos Devices and Circuits", by John Wiley& Sons, Inc., 1989.

5. T.N. Fogarty, "Process and Radiation Induced Defects in Electronic Materials and Devices", Invited paper ACerS May, 1995, published in "Crystal Growth of Novel Electronic Materials", Ceramic Transactions, Volume 60.

98 URC066

Analysis of Information Content in High-Spectral Resolution Sounders using Subset Selection Analysis

Miguel Vélez-Reyes[1]

[1]Tropical Center for Earth and Space Studies

Electrical and Computer Engineering Department

University of Puerto Rico, Mayagüez, PR 00681-9042

Tel. 787-832-4040 Ext. 3086

FAX 787-831-7564

e-mail: mvelez@exodo.upr.clu.edu

Joanna Joiner[2]

[2]Data Assimilation Office

Laboratory for Atmospheres

NASA Goddard Space Flight Center

Code 910.3, Greenbelt, MD 20771

Phone: 301-805-8442

e-mail:joiner@dao.gsfc.nasa.gov

Keywords: Atmospheric Remote Sensing, Retrievals, Data Assimilation

1 Introduction

Observations from high spectral resolution sounders sets having thousands of narrow spectral bands such as Atmospheric Infrared Sounder (AIRS) and the Infrared Atmospheric Sounding Interferometer (IASI) will lead to high dimensional data. The computational cost of direct assimilation of such huge data will be prohibitive. Even the computational cost of indirect assimilation through retrievals might not be feasible either. Therefore it is of interest to investigate the use of data compression techniques that will lead to a reduction in the number of measurements to assimilate with minimum information loss.

In this paper, we summarize the results of the sensitivity analysis and data reduction carried out to determine the information content of AIRS and IASI channels. The analysis and data reduction was based on the use of subset selection techniques developed in the linear algebra and statistical community to study linear dependencies in high dimensional data sets. We applied the subset selection method to study dependency among channels by studying the dependency among their weighting functions. Also, we applied the technique to study the information provided by the different levels in which the atmosphere is discretized for retrievals and analysis. Results from the method correlate well with intuition in many respects and point out to possible modifications for band selection in sensor design and number and location of levels in the analysis process.

2 Subset Selection

Subset selection, in our context, is a technique used to study linear dependencies in data sets [7]. In the problem at hand, we will study the near linear dependent structure of the Jacobian for HIRS and IASI caused by the high-spectral-resolution of the sensors. The basic idea of the procedure is to determine the independent columns (or rows) of the Jacobian. These independent columns (or rows) will represent the most independent pieces of information provided by the sensor. The Jacobian rows correspond to sensor weighting function so the most independent rows will correspond to the most independent channels. The columns of the Jacobian correspond to the vertical levels in which the atmosphere has been discretized so the most independent columns should correspond to the most informative levels.

Since there are more channels than levels, the Jacobian for the sounder will be full rank but extremely ill-conditioned which is related to near rank deficiency. To determine the dependency structure, we need first to compute the singular value decomposition (SVD) of the Jacobian **J** which is given by

$$\mathbf{J} = \mathbf{U}\Sigma\mathbf{V}^T \tag{1}$$

where F is the n **x** m **Jacobian (n-channels and** m-levels), **U** and **V are** orthonormal **matrices of left and right singular vectors, and** $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots \sigma_n\}$ **is the diagonal matrix of singular values (typically singular values are arranged in descending order). Due to near rank deficiency, in many cases**

$$\sigma_1 \geq \sigma_2 \geq \ldots \sigma_p >> \sigma_{p+1} \geq \sigma_n$$

The large gap between σ_p and σ_{p+1} indicates that **J** is of numeric rank p. In many data reduction applications, dimentionality reduction is achieved by using the first p left singular vectors (i.e. first p columns of **U) as** in principal component analysis [6]. The first principal components summarize most of the information content available in the data. However, principal components are linear transformation of the original data which in our application will lead to meaningless information. However, it is shown in [2, 4, 5], that a good approximation to the principal components can be achieved by selecting a subset of sufficiently independent variables. This subset of variables will have direct physical meaning while providing the same information content as the principal components.

One way to state the subset selection problem is to find a permutation matrix Π such that the p-most independent columns become the first p columns of the matrix

$$\bar{\mathbf{J}} = \mathbf{J}\,\Pi = [\mathbf{B}_1\ \mathbf{B}_2] \tag{2}$$

here $\mathbf{B}_1$ is the submatrix with the p-most independent columns. Many subset selection criteria try to select columns such that $\mathbf{B}_1$ is well conditioned. This is a combinatorial optimization problem. Several methods have been proposed in the literature which will give you a reasonable $\mathbf{B}_1$. Methods for subset selection based on the singular value decomposition are presented in [4, 5] and on QR rank revealing decompositions on [2, 3]. Notice that it is easy to formulate the subset selection problem for the rows of **J.**

3 Numerical Results

The sensitivity analysis was carried out by studying the near linear dependencies in the the Jacobian for both sensors. Such dependencies can be used to identify, using subset selection, the most independent weighting functions (pointing to most informative channels) and the most independent columns of the Jacobian (pointing to most informative pressure levels). The AIRS temperature Jacobian was computed using a program developed a NASA GSFC Data Assimilation Office. For IASI, we used a precomputed Jacobian provided by Dr. Carmine Serio to the IASI workteam.

3.1 Analysis of AIRS temperature Jacobian

The subset selection method was applied to the columns of the Jacobian to determine the most independent levels. The levels consisted of 60 levels uniformly space in a logarithmic scale from 1 to 1013 mb. The singular values of the Jacobian are shown in Figure 1 note that there is a gap around the 45th singular value. Otherwise this is a slowly decaying SV spectra which hints into an ill-posed problem where the near rank degeneracy is not very obvious and more difficult to deal with. The levels selected by the subset selection method are shown in Table 1.

SVD and QR refer to the methods used to determine what columns of the Jacobian were the most independent. These are compared to the GEOS analysis levels. The results show that most selected level are quite close to the analysis levels however the subset selection method adds an additional level on the troposphere. We definitely need to see what the inclusion of the water vapor and ozone Jacobian components do to the results but nevertheless these results are very interesting.

The subset selection procedure was applied to the rows of the temperature Jacobian to study the information content of channels by looking at the independence of their weighting function divided by the channel noise covariance. The ten most independent channels are shown in Table 2.

A plot showing the 60 most independent channels is given in Figure 2. In this case an important feature is the apparent preference of the algorithm to select channels around the 650 cm^{-1} CO_2 band.

3.2 Numerical Results for IASI the Jacobian

Subset selection technique were applied to the IASI Jacobian computed by Carmine Serio. The information provided included a temperature Jacobian for 17 levels, the ozone Jacobian for 6 levels, and the water vapor Jacobian for 8 levels. The values for IASI levels are presented in [1]. We did two types of analysis, one for each individual Jacobian and one for the full Jacobian. The first subset analysis was carried on each Jacobian individually to determine the best channels for estimation of temperature, ozone, water vapor individually. The second subset analysis was used to determine overall performance. No noise weighting was done. The singular values for the IASI Jacobian for each individual and full case are shown in Figures 3 and 5. In this case, we do not have the desired gap property for near rank deficiency however, the pivoting matrix

SVD	QR	STD LEV
1.7787	1.4607	**0.4000**
3.2424	**2.6525**	**1.0000**
5.8979	7.1806	2.0000
10.5789	10.5789	5.0000
22.3186	18.5903	10.0000
37.9854	37.9854	30.0000
62.8795	62.8795	50.0000
101.2448	101.2448	**70.0000**
158.8833	**183.5883**	**100.0000**
243.0514	**243.0514**	**150.0000**
316.3118	**316.3118**	**200.0000**
400.8599	**446.3258**	**250.0000**
541.3188	589.7161	300.0000
637.7965	684.8111	400.0000
729.9740	729.9740	500.0000
772.5025	**811.6727**	**700.0000**
905.3409	**928.5793**	**850.0000**
977.2046	**977.2046**	**1000.0000**

Table 1: Pressure levels (in mb) selected by the Subset Selection Method using QR and SVD methods and GEOS analysis levels.

produced by the RRQR can be used to determine the most independent channels with p=number of levels instead of the numerical rank.

The results for channel selection are shown in Figure 4 and 6 In contrast with AIRS, the subset selection picked channels in the neighborhood on 2300 cm^{-1}. Also, several temperature channels were picked in the water vapor absorption band. For the ozone Jacobian, four of the six channels were picked in the band near 2000 cm^{-1} and NOT in the one near 1000 cm^{-1} as one might have expected. Subset selection analysis of the full Jacobian picked the channels shown in Figure 6.

When comparing the results of the subset selection for the full and individual Jacobians for temperature, ozone, and water vapor, it is interesting to note that the same ozone and water-vapor channels were also picked during the full analysis. Four of the channels in the water vapor band picked during the analysis of the temperature Jacobian were not picked in the full analysis. The full analysis picked two additional channels near 2300 cm^{-1} (2375 cm^{-1} and 2388 cm^{-1}) which are better for temperature sounding, one at 1450 cm^{-1}, and another at 1734 cm^{-1} both in the water vapor band.

The IASI results need to be further validated since there is no information on how the Jacobian was computed. In is of interest to find out whether solar radiation was considered in the computation which would explain some of the results around 2000 cm^{-1}.

4 Conclusions

The results presented in this paper show the potential of using subset selection techniques to assess the information content of high-resolution-sounders. The algorithms used for the analysis are based on standard matrix decompositions available in packages such as MATLAB. Computation time needed was quite short and the procedure is a lot simpler than that presented in [8].

It is difficult to make a fair comparison with the results presented in [8] since we did not have access to the full AIRS Jacobian. However, the results shown here suggest that further work should be pursued to develop a better understanding of sensitivity and information content in high-spectral-resolution sounders. The use of the subset selection approach is very attractive given the simplicity of the proposed approach.

5 Acknowledgements

The research presented here was sponsored in part by NASA URC Program under grant NCCW-0088 and the 1997 NASA Summer Faculty Fellowship Program.

Num.	Freq. (cm^{-1})
1	666.2800
2	666.8360
3	667.3920
4	667.6700
5	668.5050
6	680.0230
7	691.1640
8	741.2800
9	2,387.4
10	2,392.4

Table 2: Most independent channels selected by the subset selection method.

References

[1] U. Amato and V. Cuomo and C. Serio. "Assesing the impact of radiometric noise on IASI performance." In *Int. J.* Remote Sensing, Vol. 16 no. 15, 2927-2938, 1995.

[2] T.F. Chan and P.C. Hansen. "Some Applications of the Rank Revealing QR Factorization." In SIAM Journal on *Scientific* and Statistical Computing, Vol. 13, No. 3, pp. 727-741, 1992.

[3] S. Chandrasekaran and I.C.F. Ipsen. "On Rank-Revealing Factorization." In *SIAM* Journal on Matrix Analysis and Applications, Vol. 15, No. 2, pp. 592-622, 1994.

[4] G.H. Golub, V. Klema, and G.W. Stewart. Rank Degeneracy and *Least* Squares Problems. Technical Report TR-456, Department of Computer Science, University of Maryland, College Park, 1976.

[5] G.H. Golub and C.F. Van Loan. *Matrix* Computations. John Hopkins, 3rd edition, 1997.

[6] I.T. Jollife. Principal *Component* Analysis. Springer-Verlag, 1986.

[7] V.E. Kane, R.C. Ward, and G.J. Davis. "Assessment of Linear Dependencies in Multivariate Data." *SIAM* Journal on *Scientific* and *Statistical* Computing, Vol. 6, No. 4, 1985.

[8] C.D. Rodgers. Information *content of high spectral* resolution measurements. Unpublished.

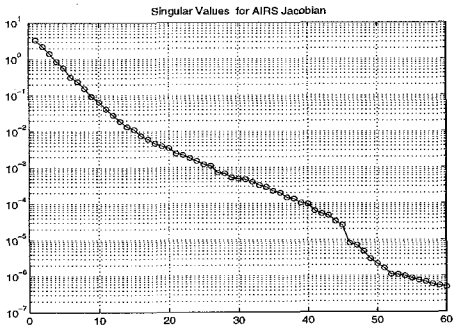


Figure 1: Singular Value Plot for the AIRS Temperature Jacobian.

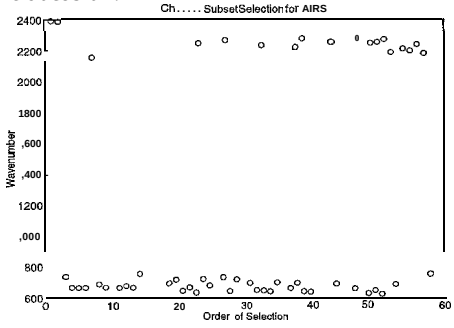


Figure 2: Results of Subset Selection for AIRS Channels.

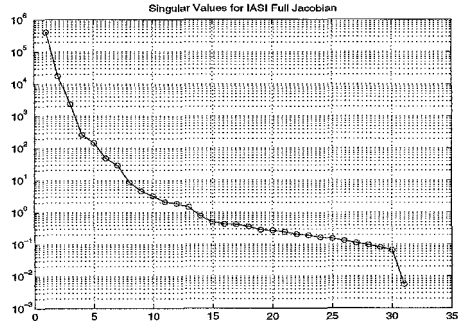


Figure 3: Singular Value Plots for IASI Full Jacobian.

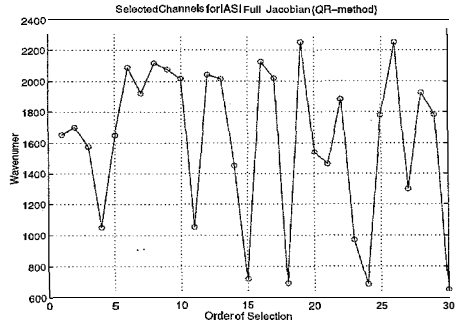
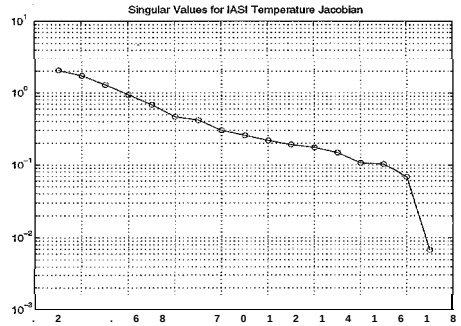
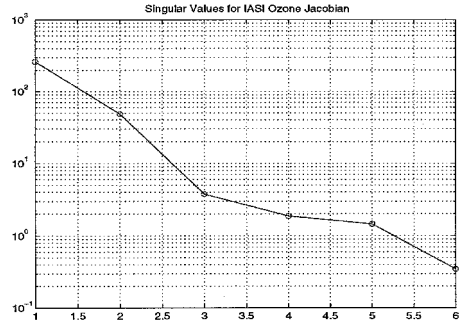


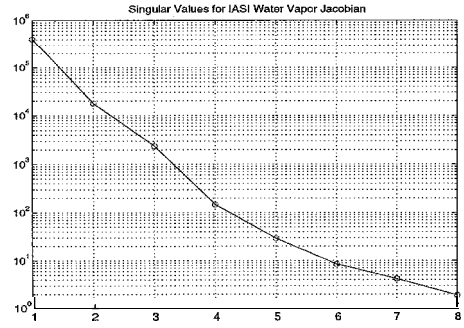
Figure 4: Results of Subset Selection for Full IASI Jacobian.



(a)

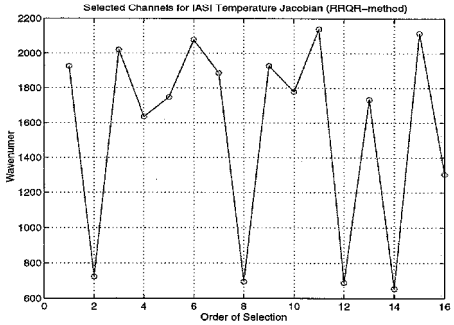


(b)

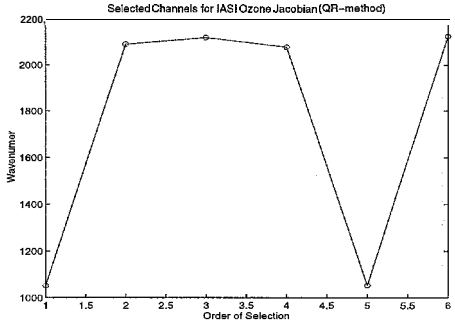


(c)

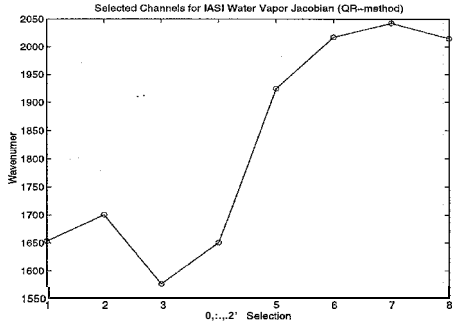
Figure 5: Singular Value Plots for IASI Jacobians: (a) Temperature, (b) Ozone, (c) Water Vapor.



(a)



(b)



(c)

Figure 6: Results of Subset Selection for IASI Channels: (a) Temperature, (b) Ozone, (c) Water Vapor.

Clustering Algorithm for Hyperspectral Data[1]

Luis O. Jimenez
ECE Department
University of Puerto Rico
Mayaguez Campus
Mayaguez, Puerto Rico 00681-5000
Tel. (787)832-4040 Ext. 3031
Email: jimenez@exodo.upr.clu.edu

Yamil Chaar
ECE Department
University of Puerto Rico
Mayaguez Campus
Mayaguez, Puerto Rico 00681-5000
Email: chaar@amadeus.upr.clu.edu

Abstract

A big challenge is to discover the potentially increased number of different data patterns in a very high dimensional data. That will enable us to find labeled samples of the different clusters for supervised classification purposes. This paper will focus in unsupervised classification algorithms were the machine classifies the image with only a small amount of human intervention. Unsupervised classification algorithms are techniques to extract information from Remote Sensing imagery based on machine calculation without prior knowledge of labeled samples and the number of classes. The methodology is implemented by automatically discovering patterns in Remote Sensing data. The main advantage of unsupervised over supervised classification is that it does not require a high level of expertise from the analyst. Clustering is an attempt to find structure in a set of observations that we know a little about, is a form of segmentation. Its objective is to group the data in different clusters that have different statistical patterns. The best known program is the ISODATA algorithm, This program and most of the clustering algorithm only use first order statistics and the spectral response information. Present unsupervised classification algorithms have been developed for images with a small number of bands. They do not take into consideration the inherent hyperdimensional space characteristics. In order to overcome such problems a hyperspectral data clustering has been developed taking into consideration hyperspectral data characteristics and spatial information to discover more patterns with increased accuracy.

Keywords: Pattern Recognition, Remote Sensing, High Dimensionality, Hyperspectral Data, High Dimensional Space, Unsupervised Classification, Statistical Analysis.

[1] This research was partially funded by the Topographic Engineering Center of the U.S. Army Corps of Engineering.

2. High dimensional space characteristics

Most of the problems that need to be solved are related with the inherent properties of hyperdimensional space. The curse of dimensionality has been known for more than three decades. In combinatorial optimization over many dimensions, it is seen as an exponential growth on the computational effort with the number of dimensions. In statistics, it manifests itself as a problem with parameter or density estimation dune to the paucity of data. A study of high dimensional space characteristics and its implication for hyperspectral data analysis has been done previously [1]. On the basis of characteristics of high dimensional data such as that the volumes of hypercubes have a tendency to concentrates in the corners, and in a hyperellipsoid in an outside shell, it is apparent that high dimensional space is mostly empty. Hyperspectral data is usually located in a lower dimensional structure. As a consequence, it is possible to reduce the dimensionality without losing significant information and separability. Another consequence of these characteristics is that local neighborhoods are almost surely empty, requiring the bandwidth of estimation to be large and producing the effect of losing detailed density estimation [2],[3]. Due to the problem of density estimation, nonparametric classification is almost impossible in hyperspectral data. Another characteristic, is the relevance of second order and first order statistics in hyperspectral data. The location of the data clouds is not enough information for image segmentation. The orientation of the clusters, given by the covariance, is important as well. What is called for is the development of algorithms with classification and feature reduction capabilities that utilize the amount of information and separability that hyperspectral data offer while simultaneously avoiding the difficulties inherent in hyperdimensional space. In order to achieve this, it is necessary to continue research in high dimensional space properties and the implications of such properties for remote sensing data analysis.

3. Unsupervised Classification Software System for Hyperspectral Data.

Unsupervised classification algorithms are techniques to extract information from Remote Sensing imagery based on machine calculation without prior knowledge of labeled samples and the number of classes. The methodology is implemented by automatically discovering patterns in Remote Sensing data. The main advantage of unsupervised over supervised classification is that it does not require a high level of expertise from the analyst. Clustering is an attempt to find structure in a set of observations that we know a little about, is a form of segmentation [4]. Its objective is to group the data in different clusters that have different statistical patterns. The best known program is the ISODATA algorithm. This program and most of the clustering algorithm only use first order statistics and the spectral response information. This system has a tendency to produce poor results when applied to hyperspectral data. This is due to the fact that it does not take into consideration the mentioned hyperspectral data characteristics. A hyperspectral data clustering has to take into consideration hyperspectral data characteristics and spatial information to discover more patterns with increased accuracy.

Clustering algorithms depends on two measurements [5]. The first measurement is the distance between a point and the expected values (locations) of the clusters. This distance is used to assign the point to the closer cluster. If the distance only considers frost order statistics, as the two dimensional example in Figure 3a, it will assume that the clusters have a form of a perfect hypersphere with no correlation among the bands. If a second order statistics is used as a parameter in the distance then a situation as the Figure 3b will occur, where you assume a hyperellipsoid form with a possible correlation among the bands. The second case is more flexible, general and more according to real Remote Sensing data. This type of distances that takes into consideration first and second order statistics will be applied. Its results will be compared with the ISODATA algorithm that only uses the location as a parameter.

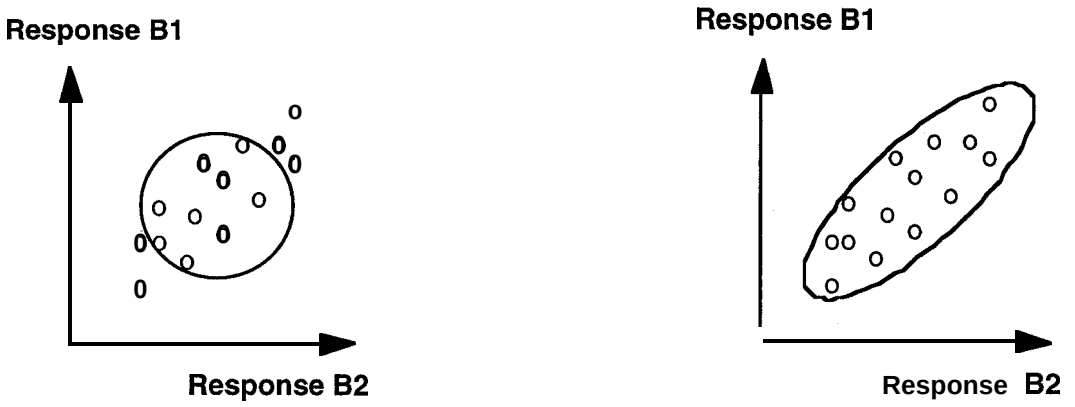


Figure 3. Two dimensional representation of a spectral response at binds 1 and 2. Part a does not takes into consideration second order statistics, it assumes no correlation among the bands. Part b does take into consideration second order statistics, with no assumption about the correlation among the bands.

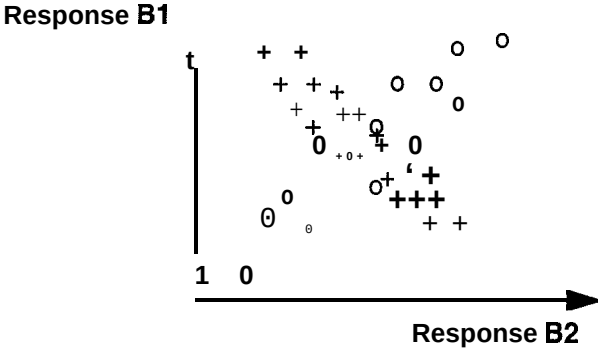


Figure 4. Two dimensional representation of clusters with the same expected value and different covariances.

Another measurement that clustering uses is the distance among the clusters. Present schemes do not use second order statistics. They only measure the cluster's location differences in terms of the Euclidean distance. Figure 4 shows an example where you have two clusters with identical expected value and different covariances. With the present clustering algorithms the clusters in the example will be merge in one because the location difference among them is zero.

4. **Algorithm Description**

One of the methodologies for developing a clustering algorithm is based on Maximum Likelihood estimates. In this particular case we assumed the case of normal mixtures. A clustering algorithm has to solve simultaneously the following equations [Duds and Hart]:

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega_i \mid x_k, \hat{\theta})$$

$$\hat{\mu}_i = \frac{\frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega_i \mid x_k, \hat{\theta}) x_k}{\frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega_i \mid x_k, \hat{\theta})}$$

$$\hat{\Sigma}_i = \frac{\frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega_i \mid x_k, \hat{\theta})(x_k - \hat{\mu}_i)(x_k - \hat{\mu}_i)^T}{\frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega_i \mid x_k, \hat{\theta})}$$

where

$$\hat{P}(\omega_i \mid x_k, \hat{\theta}) = \frac{\hat{P}(x_k \mid \omega_i, \hat{\theta}_i)\hat{P}(\omega_i)}{\sum_{j=1}^{c} \hat{P}(x_k \mid \omega_j, \hat{\theta}_j)\hat{P}(\omega_j)} = \frac{|\hat{\Sigma}_i|^{-1} \exp\left[-\frac{1}{2}(x_k - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1}(x_k - \hat{\mu}_i)\right]}{\sum_{j=1}^{c} |\hat{\Sigma}_j|^{-1} \exp\left[-\frac{1}{2}(x_k - \hat{\mu}_j)^T \hat{\Sigma}_j^{-1}(x_k - \hat{\mu}_j)\right]}$$

where c is the number of classes, n is the number of samples and $\hat{\theta}_i = (\hat{\mu}_i, \hat{\Sigma}_i)$ is the set of estimated parameters.

There are various techniques to simplify the computation and accelerate convergence. One of them is the ISODATA algorithm:

For the x_k find the nearest mean $\hat{\mu}_m$ and approximate $\hat{P}(\omega_i \mid x_k, \hat{\theta})$ as:

$$\hat{P}(\omega_i \mid x_k, \hat{\theta}) \approx \begin{cases} 1 & i = m \\ 0 & \text{otherwise} \end{cases}$$

assuming $(\hat{\Sigma}_i = I)$

This leads to the following procedure:

ISODATA

1. Choose some initial values for the means $\hat{\mu}_1,...,\hat{\mu}_c$ (assuming c classes).
2. Classify then samples by assigning them to the class of the closest mean.
3. Recompute the means as the average of the samples in their class.
4. If any mean changed value, go to 2, otherwise stop.

5. Experiment and Results

The ISODATA algorithm was modified in order to use second order statistics $(\hat{\Sigma}_i \# I,)$ in specific groups covariances information, in the clustering program. The Modified ISODATA is as follows:

Modified ISODATA

1. The initial values for the means $\hat{\mu}_1,...,\hat{\mu}_c$ are taken from the results of the regular ISODATA (assuming c classes). The initial covariances are assumed to be $(\hat{\Sigma}_i = I)$.
2. Classify then samples using the ML classifier.
3. Recompute the means and the covariances for every class.
4. If any mean or covariance changed value, go to 2, otherwise stop.

Both algorithms were applied to a Airborne simulator of Thematic Mapper in Añasco which is in the western region of Puerto Rico. Figure 5 shows the original image that was used. This original image has different crops, roads, an airport, and soils. Figure 6 shows the result of the ISODATA algorithm. Five clusters are map in that image. The spatial structures and variations of the roads and the airport are not obvious in the results. Figure 7 shows the results of the modified clustering algorithms using estimated covariances. The spatial structures of different classes such as the roads and the airport appear more clear in this results than in the regular ISODATA.

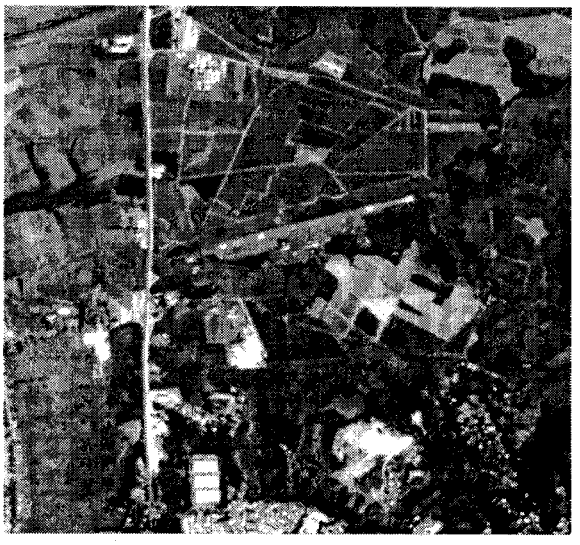


Figure 5. Original Image: Añasco Region

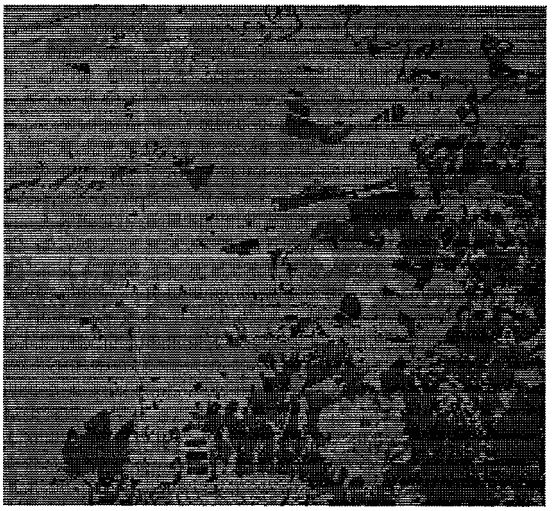


Figure 6. ISODATA Clustering

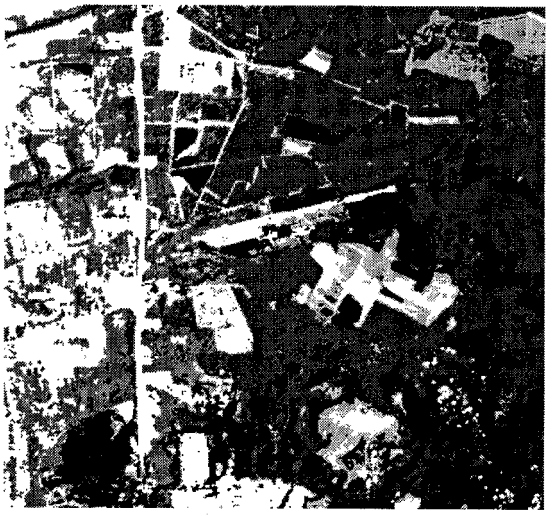


Figure 7. Modified ISODATA Clustering

6. **Conclusion**

It has been shown that for Supervised classification the information provided by the means and the covariances is relevant. In the results shown in this paper we can verify that this is true for unsupervised classification methods as well. This is very important for remote sensing data where the correlation among bands is class dependent. As the number of bands increases, i.e. hyperspectral data this is significant because the correlation is highly dependent on the proximity of the channels. In our particular case the Modified ISODATA was used as a method to enhance the results of Classical ISODATA. We need to explore new methods of finding the initial conditions for both algorithms.

6. **Reference**

[1] L. O. Jimenez and D. A. Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transactions on System, Man, and Cybernetics,* submitted and accepted for publication.

[2] D. W. Scott, *Multivariate Density Estimation,. New* York: John Wiley& Sons, 1992.

[3] J. Hwang, S. Lay and A. Lippman, "Nonparametric Multivariate Density Estimation: A Comparative Study," *IEEE Transactions on Signal Processing,* Vol. *42, No.* 10, pp. 2795-2810, 1994.

[4] M. Nadler, E.P. Smith, *Pattern Recognition Engineering, New* York: John Wiley& Sons, 1993.

[5] J.A. Richards, *Remote Sensing Digital Image Analysis, An Introduction.* 2nd ed. New York: Springler-Verlag, 1993.

[6] R.O. Duda, P.E. Hart, *Pattern Classification and Scene Analysis*, pp. 189-202, 1973.

Eigenvalue Eignevector Constraint based Sufficient Bound for Root Clustering in Circular Regions

CR. Ashok kumar* Frederick Ferguson†
North Carolina A&T State Univeristy
Greensboro NC 27411

December 11, 1997

Abstract

In eigenvalue analysis with real parametric uncertainties in a matrix, robust root clustering focus on finding bounds on real parametric uncertainties which preserve the perturbed eigenvalues in specific subregions of our interest in complex plane. Such bounds analyzed in literature evolves from different strategies. In this paper, we present eigenvalue-eigenvect or constraint based bounds for root clustering in circular regions. The bounds presented here are less conservative than the existing bounds.

1 Introduction:

Root clustering (RC) of a nominal matrix in subregions of the complex plane is well known to satisfy the generalized Lyapunov equations [1]. Robust root clustering (RRC) assumes a matrix with real parametric uncertainties for RC; however, the error matrix family satisfying the generalized Lyapunov equations are not known. Thus, RRC is an approach to determine the uncertain matrix family with the perturbed eigenvalues in subregions of the complex plane [2]. Specifically, RRC in circular regions is useful to analyze both stability and performance of the discrete and continuous time linear uncertain systems. The optimization techniques recently established provide less conservative bounds [3] than the bounds in [2]. Earlier work in this research area is reported in references [4]- [6]. In this paper, eigenvalue-eigenvector constraint based bound for RRC in circular regions is presented.

2 Root Clustering in Circular Regions:

Consider a nominal matrix $A \in R^{n\times n}$ with its eigenvalue spectrum $sp(A) = [\lambda_1 \ldots \lambda_n]$ in a circular region $\mathcal{D}(\beta, r)$, where β is the center and r is the radius. If A is perturbed by an error matrix E, it is of our interest to find the bounds on E, such that $sp(A + E) = [\mu_1 \ldots \mu_n] \in \mathcal{D}(\beta, r)$. Generally, the bounds on E are established as in the cases below:

$$\sigma_{\max}(E) < \mu_{2_{rc}} \qquad \text{(if } E \text{ is unstructured)} \tag{1}$$

*Postdoctoral Research Associate, Center for Aerospace Research
†Director and Adjunct Associate Professor, NASA Center for Aerospace Research

$\sigma_{\max}(E)$ is the maximum singular value of the matrix E. If E is structured, the perturbation locations with uncertain parameters $q_j, j = 1 \cdots r$ are assumed to be known. Thus,

$$E = \sum_{j=1}^{r} q_j E_j \tag{2}$$

where, E_j are the constant matrices. The bounds for structured uncertain matrices are desired on uncertain parameters q_j satisfying:

$$|q_j| < \mu_{2_{rcs}} \quad j = 1 \cdots r \qquad \text{(if } E \text{ is structured)} \tag{3}$$

The notations $\mu_{2_{rc}}$ and $\mu_{2_{rcs}}$ for the bounds are used in reference [2] for all subregions in complex plane representable by second order polynomial. Here, we shall refer them to circular regions. Suppose V is a transformation matrix for the Jordan canonical form of A. Then:

$$J = V^{-1}AV \quad \text{and} \quad sp(A) = sp(J) \tag{4}$$

Since the eigenvalues are invariant under transformation matrices

$$sp(A + E) = sp(V^{-1}(A + E)V) \tag{5}$$

Thus, eigenvalue(μ_i) and eigenvector(v_{e_i}) constraint for the perturbed matrix $(A+ E)$ is:

$$\begin{aligned} \mu_i v_{e_i} &= [V^{-1}(A+E)V]v_{e_i}, \quad i = 1 \cdots n \\ \mu_i v_{e_i} - \beta v_{e_i} &= [V^{-1}(A+E)V]v_{e_i} + \beta V^{-1}Vv_{e_i} \\ (\mu_i - \beta)v_{e_i} &= (J - \beta I_n)v_{e_i} + V^{-1}EVv_{e_i} \end{aligned} \tag{6}$$

For root clustering in circular regions $\mathcal{D}(\beta, r)$, note that $|\mu_i - \beta| < r$, 'd $i = 1 \ldots n$. We shall now establish the overbounds Γ_{us} and Γ_s satisfying $|\mu_i - \beta| < \Gamma_{us}$ (or Γ_s). Using equation 6, note that:

$$\begin{aligned} |\mu_i - \beta| &< \sigma_{\max}(J - \beta I) + \kappa(V)\sigma_{\max}(E), \quad i = 1 \cdots n \\ &= \max_{1 \le i \le n}(|\lambda_i - \beta|) + \kappa(V)\sigma_{\max}(E) = \Gamma_{us} \quad \text{(if } E \text{ is unstructured)} \end{aligned} \tag{7}$$

$$= \max_{1 \le i \le n}(|\lambda_i - \beta|) + \sigma_{\max}\left(\sum_{j=1}^{r} q_j V^{-1}E_j V\right)$$

$$= \max_{1 \le i \le n}(|\lambda_i - \beta|) + \max_j |q_j| \sigma_{\max}\left(\sum_{j=1}^{r} |V^{-1}E_j V|\right) = \Gamma_s \text{ (if } E \text{ is structured)} \tag{8}$$

Here, $|V^{-1}E_jV|$ refers to the magnitude of the individual elements in $V^{-1}E_jV$ and $\kappa(V)$ is the condition number of the transformation matrix V.

2.1 Bounds for Root Clustering in Circular Regions:

From equations 7-8, it is clear that the perturbed eigenvalues μ_i will reside in the circular region $\mathcal{D}(\beta, r)$ if

$$\Gamma_{us} - r < 0 \tag{9}$$

$$\Gamma_s - r < 0 \tag{lo}$$

The conditions in equations 9-10 are the disjointed properties used in reference [6] for the perturbed eigenvalues to reside in a set of dominant disjointed regions. In this paper we extend the technique for robust root clustering in circular regions.

Theorem 1 *Let $\lambda_1 \cdot . . \cdot \lambda_n$ be the eigenvalue spectrum of a nominal matrix $A \in R^{n\times n}$ residing in a circular region $\mathcal{D}(\beta, r)$ with its center at β and radius r satisfying $r > \max_{1\leqslant i\leqslant n}(|\lambda_i - \beta|)$. Let μ_1 ,. . μ_n be the perturbed eigenvalues of the matrix (A + E) and $E \in R^{n\times n}$ the error matrix. Then μ_i will reside in $\mathcal{D}(\beta, r)$ if*

$$\sigma_{\max}(E) \quad < \quad \frac{r \text{ - } \max_{1\leqslant i\leqslant n}(|\lambda_i \text{ - } \beta|)}{K(v)} = \mu_{2_{rc}} \tag{11}$$

$$|q_j| \quad < \quad \frac{r - \max_{1\leqslant i\leqslant n}(|\lambda_i - \beta|)}{\sigma_{\max}(\sum_{j=1}^{r} |V^{-1}E_jV|)} = \mu_{2_{rcs}} \quad j = 1 \cdots r \tag{12}$$

where V satisfies $J = V^{-1}AV$, $|(.)|$ is the absolute value of the scalar (.) or the absolute value of the individual elements of the matrix (.) (if (.) is a matrix), and $\kappa((.))$, $\sigma_{\max}((.))$ are the condition number, maximum singular value of the matrix (.).

Proof: Suppose the perturbed eigenvalues μ_i reside in $\mathcal{D}(\beta, \mathrm{r})$ if

$$\begin{aligned} \sigma_{\max}(E) &< \frac{r - \max_{1\leqslant i\leqslant n}(|\lambda_i - \beta|)}{\kappa(V)} \\ \kappa(V)\sigma_{\max}(E) + \max_{1\leqslant i\leqslant n}(|\lambda_i - \beta|) &< r \\ \Gamma_{us} - r &< 0 \\ |\mu_i - \beta| &< r \qquad \forall \qquad i = 1 \cdots n \end{aligned}$$

Similarly, the proof for the bound $\mu_{2_{rcs}}$ in equation 12. ■

2.2 Example:

The following example for nom *bounded uncertainty* is considered in references [2] and [3].

$$A = \begin{bmatrix} -4.3 & -0.4 \\ 0.2 & -3.4 \end{bmatrix} \quad \text{with } [\lambda_1 = -4.2,\ \lambda_2 = -3.5] \in \mathcal{D}(-4,\ 1)$$

Since the eigenvalues of *A* are distinct, note the matrix *V* is indeed the matrix of eigenvectors diagonalizing *A*. Moreover, it is appropriate to consider the radius $r > \max_{1\leqslant i\leqslant 2}\{|\lambda_i - 4|\} =$ **0.5.** The bound by equation 11 is compared in Table-1.

Table 1: Bounds for norm bounded uncertainty

Bounds	Ref [2]	Ref [3]	Present
$\mu_{2_{rc}}$	*0.0341*	0.4	0.23

For *structured uncertainty* with uncertain parameters q_1 and q_2, the perturbation locations in the above example are assumed to be known as below.

$$E = q_1 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + q_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = q_1E_1 + q_2E_2$$

The bounds for this case are compared in Table-2.

Table 2: Bounds for structured uncertainty

Bounds	Ref [6]	Ref [3]	Present
$\mu_{2_{rc}}$	*0.0317*	*0.4*	*0.2265*

Note that the bounds in reference [3] is a necessary and sufficient condition for q_2, obtained by minimax parameter optimization techniques. The present approach provides a bound which is less conservative than the bounds in references [2] and [6].

3 Conclusions:

This paper provides a new sufficient bound for root clustering in circular regions using eigenvalue and eigenvector constraints. The approach uses Jordan canonical transformation for the nominal matrix. The transformation matrix is indeed a modal matrix if the nominal matrix is diagonalizable. The analytical bounds established in this paper are less conservative than the bounds available in the literature.

References

[1] Gutman, S. and E.I. Jury. A general theory for matrix root clustering in subregions of the complex plane, *IEEE Transactions on Automatic Control,* AC 34, 1981, p157

[2] Yedavalli, R.K. Robust root clustering for linear uncertain systems using generalized Lyapunov theory, *Automatic, v 29,* n 1, 1993, p 237

[3] Luo, J.S. P.P. van de Bosch, and S. Weiland. New sufficient condition for robust root clustering of state space models with structured uncertainty, *IEE Proc- Control Theory Applications, v 143,* n 3, 1996, p244

[4] Yu, W. and K.M. Sobel, Robust eigenstructure assignment with structured state space uncertainty, *Jl. of Guidance,* Control, and *Dynamics, v 14,* n 3, 1991, p 621

[5] Juang, Y.T. Eigenvalue assignment robustness: the analysis for autonomous system matrices, *International Jl. of Control, v 49,* n 5, 1989, p 1787

[6] Yedavalli, R.K. A Kronecker based theory for robust root clustering for linear state space models with real parametric uncertainty, *In the Proc. of American Control Conf., 1993*

[7] Ashokkumar, CR. and R.K. Yedavalli, Eigenstructure perturbation analysis in disjointed domains for linear uncertain systems, *to appear in International Jl. of Control*

98 URC069

Approximation Of Multi-Valued Inverse Functions Using Clustering And Sugeno Fuzzy Inference

Maria A. Walden waldenma@ncat.edu
Marwan Bikdash bikdash@ncat.edu
Abdollah Homaifar homaifar@ncat.edu

NASA Autonomous Control Engineering Center
Department Of Electrical Engineering
NCA&T State University

Abstract: Finding the inverse of a continuous function can be challenging and computationally expensive when the inverse function is multi-valued. Difficulties may be compounded when the function itself is difficult to evaluate. We show that we can use tizzy-logic approximators such as Sugeno inference systems to compute the inverse on-line. To do so, a fuzzy clustering algorithm can be used in conjunction with a discriminating function to split the function data into branches for the different values of the forward function. These data sets are then fed into a recursive least-squares learning algorithm that finds the proper coefficients of the Sugeno approximators; each Sugeno approximator finds one value of the inverse function. Discussions about the accuracy of the approximation will be included.

Keywords: *Fuzzy Clustering, Function Inversion, Recursive Least squares approximation, and Multi-valuedness*

1. Introduction

Sugeno Approximators (SA) have been proven to approximate any continuous function to a high degree of accuracy under certain standard assumptions, as well as other methods that have been used to approximate continuous functions. Some of these other methods involve the use of interpolation (Poggio and Girosi, 1990), neural networks (Li, 1992), (Poggio and Girosi, 1990) and fuzzy inference systems such as Sugeno Approximators (Takagi and Sugeno, 1985) and generalized Sugeno Approximators (Buckley, 1993). These have all been designated as universal approximators, however none of these approximators are geared for approximating multi-valued functions. In this paper, we introduce a new approach for approximating the multi-valued inverse of a continuous function. Three elements are used in our approach: A mechanism to discriminate the region of multi-valuedness, a clustering algorithm to split the training data into separate training sets, and a training algorithm which finds a Sugeno approximation for each training data set. For the values of the inverse function, clearly many Sugeno Approximators are now needed. However, not all of the input/output pairs can be used for training the different SAs. If so, the SAs will be indistinguishable. Therefore a clustering algorithm is required to split the training data into different sets. We demonstrate our results by using examples of one-dimensional inverted functions.

2. Training Sugeno Approximators

A Sugeno Approximator (SA) is a fuzzy engine mapping a vector $x = [x_1, x_2, \ldots . x_n]^T \in \Re^n$ into $u \in \Re$. A typical (k^{th}) rule is of the form:

$$R^K : \text{IF } x_1 \text{ is } A_1^k \text{ and } x_2 \text{ is } A_2^k \text{ and } \ldots \text{ and } x_n \text{ is } A_n^k \qquad 1.1$$

$$\text{Then } u = y^k = c_0^k + c_1^k x_1 \ldots c_n^k x_n \qquad 1.2$$

where xi is the crisp value of the i^{th} component of the input vector x, A_i^k specifies which fuzzy attribute of x_i is being tested by rule k, and the consequence coefficients of the k^{th} fuzzy rule are $c_0^k, c_1^k, \ldots c_n^k$. The output $\hat{u}$ for a SA is obtained by the centroid method of defuzzification.

$$\hat{u} = \sum \beta_k y^k \qquad 1.3$$

Where β_k (X) is the normalized truth value (strength) of the antecedent of the k^{th} rule evaluated with input x, $X = [1, x_1, x_2, \ldots x_n]$ and $c^k = \left[c_0^k, c_1^k x_1, \ldots, c_n^k\right]^T$ which have to be determined.

Let (X^j, Y^j) be the j^{th} training of input/output pair of a total of J pairs. The consequence parameters can then be obtained by minimizing the error index $L = \sum_{j=1}^{J} [Y^j - \hat{u}^j]^2$ where Y^j & $\hat{u}^j$ are the desired and actual output of the given X^j. The simplest approach is to assume that there are as many samples as there are unknown coefficients; i.e, J = K(n+ 1). Then the coefficients are found by solving the set of linear equations:

$$ZC = Y \qquad 1.4$$

where

$$C = \left[c^1; \ldots; c^k\right] Y = \left[Y^1, \ldots, Y^J\right]^T \qquad 1.5$$

and a J*K(n+ 1) matrix, has the form

$$Z = \begin{bmatrix} \beta_1^1 & x_1^1\beta_1^1 & \ldots & x_n^1\beta_1^1 & \ldots & \beta_k^1 & x_1^1\beta_k^1 & \ldots & x_n^1\beta_k^1 \\ \ldots. & \ldots.. & & \ldots... & & \ldots.. & \ldots.. & & \ldots.. \\ \beta_1^J & x_1^J\beta_1^J & \ldots & x_n^J\beta_1^J & \ldots & \beta_k^J & x_1^J\beta_k^1 & \ldots & x_n^J b_k^J \end{bmatrix} \qquad 1.6$$

where $\beta_k^j = \beta_k\left(x^j\right)$ signifies the truth values or strength of the rules at vector X^j. When J = K(n+1) then C = Z-l Y. If J > K(n+1), the least squares solution is found using the pseudo inverse $C = (Z^T Z)^{-1} Z$ 'Y. Alternatively C can be calculated recursively using a standard recursive least-squares algorithm, see (Clifton, 1996) and (Takagi and Sugeno, 1995). Rules can be eliminated if their truth values are below a threshold value that is greater than 0. When a rule doesn't fire, the Z matrix is not full rank and ZC = U has no unique least-squares solution. With the recursive least squares the coefficients are dependent on the initial estimates. The rank deficiency of a Z matrix is compounded by numerical difficulties and if a rule never fires or fires slightly for any given input/output data, the rule is eliminated to make the least squares problem solution unique. For more information refer to (Buckley, 1993), and (Clifton, 1996).

3. Motivating Example

The problem in this paper involves utilizing an SA to approximate the inverse of a one-dimensional continuous function say sin (x) shown in Figure l(a). The inverse, sin-l (x), is shown in Figure 1(b). In Figure 1 (a), we also show the approximation produced by an SA with 6 membership functions. The Root Mean Square Error (RMSE) is used to characterize the approximation error. Figure 1 (b) compares the SA approximate with the inverse sine function. Because of multi-valuedness, the SA interpolates between the

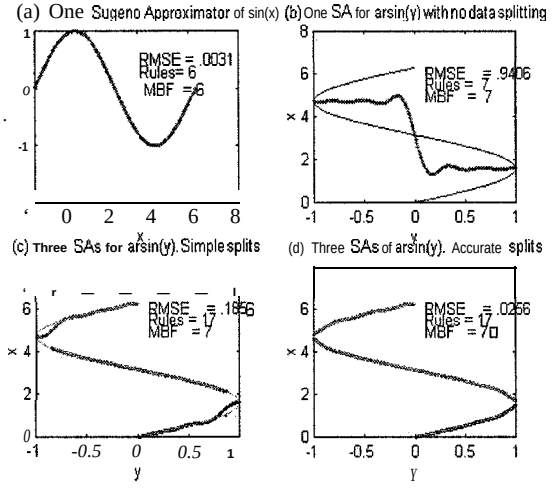


Figure 1: Sugeno Inversion of sin(x) Using Simple Splits

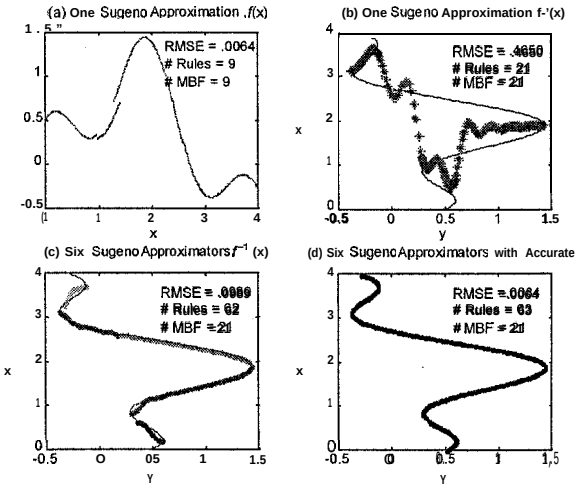


Figure 2: Sugeno Inversion of $f(x) = \sin(x) + .5\cos(3.23x)$ Using Simple Splits

$$\text{RMSE} = \sqrt{\frac{1}{J} * \sum_{j=1}^{J} (Y^j - \hat{u}^j)^2} \qquad 1.7$$

many values, where the dark line represents the interpolated values. Hence, it seems crucial that the branches of the inverse function must be clearly split and data from different branches must be applied to different SAs. From Figure 1 (b), it is obvious that the data can be broken into three parts each representing one branch of the inverse function. First, we arbitrarily split the data into three sets of equal size applying each set to train a Sugeno Approximator. Figure 1 (c) shows the outcome. The SAs were able to approximate the function but still there is a problem with multi-valuedness near the peaks. But the RMSE decreased from 0.9406 to 0.1856. Next we split the data correctly at the peaks; namely, $0 \le x \le \pi/2$; $\pi/2 < x \le 3\pi/2$; and $3\pi/2 < x \le 2\pi$. Applying each data set to an SA, we obtain Figure 1(d) with an RMSE of .0256. To simplify the software implementation, we have chosen 7 gaussian membership functions with equidistant means over the full range of y (-1 to 1). This is more than necessary. Note for example that for the SA representing the top branch of sin-l (x), the membership functions covering $y > 0$ are not really needed, because none of the training or testing data fall in that range. Clearly, for these data sets, at least half of the rules will never fire and hence can be eliminated. Fadïng the simple test $\max_j(\beta_k) < \beta_{thresh}$ will eliminate the k^{th} rule if the truth value never exceeds a given threshold when all training data are used.

This will not only eliminate rules that fall out of range, but also rules that contribute very little to the approximation.

We apply the same procedures to the one-dimensional function shown in Figure 2.

$$y = \sim(x) = \sin(x) + .5*\cos(3.23*x) \qquad 1.8$$

In part (a) we show the original function and its approximation using one SA with 9 membership functions and 9 rules. For one dimensional functions the branches of the inverse function f^{-1} are delineated by the boundaries and peaks of the forward function *f*. These peaks can be found from the zeros of its derivative; i.e. by solving the equation $df/dx = O$. This is how the data were split accurately to obtain part(d) of Figures 1 and 2.

4. Splitting The Data Using Clustering

In anticipation of having to invert higher-dimensional functions, we will consider a more general approach to splitting data into sets fed into different SAs. This approach is a combination of introducing a discriminating function and a fuzzy clustering .

Clustering is a technique that is useful in neural networks, pattern recognition, genetic algorithms and fuzzy logic (G.S.R and et al, Ross, 1995). This process involves grouping data, based on some form of similarity, into joint or disjoint classes. Fuzzy clustering employs fuzzy partitioning where a data point has a degree of membership between O and 1 in each of the respective classes. A normalization stipulation however, must hold, thus requiring the degrees of belongingness of a specific samples into the different classes to add to unity. (Hall and et al, 1992 and Sate, M and Sate, 1995). Fuzzy Clustering finds a (c*n) fuzzy partition matrix U for grouping a collection of J data samples into c classes that minimizes the objective function J_m (Hall et al, 1992) v_k = center of the K^{th} cluster, J = number of data samples, c = number of clusters, μ_{ik} = membership of x_i into Cluster c_k , *m* ' = feature space (weighting parameters), d_{jk} =

$$J_m(U,v) = \sum_{k=1}\sum_{j=1}(u_{jk})^m (d_{jk})^2$$

Euclidean distance from x^j to v^k, $d_{jk} = x^j - v^k$, and $v^k = \sum_{k=1}^{n} u_{jk}^{m'} x^j / \sum_{k=1}^{n} u_{jk}^{m'}$. The weighting parameter *m'*, introduced by Bezdek in 1981, is used to control the amount of fuzziness in the classification process. Figure 3 shows how the training of the different SAs is conducted.

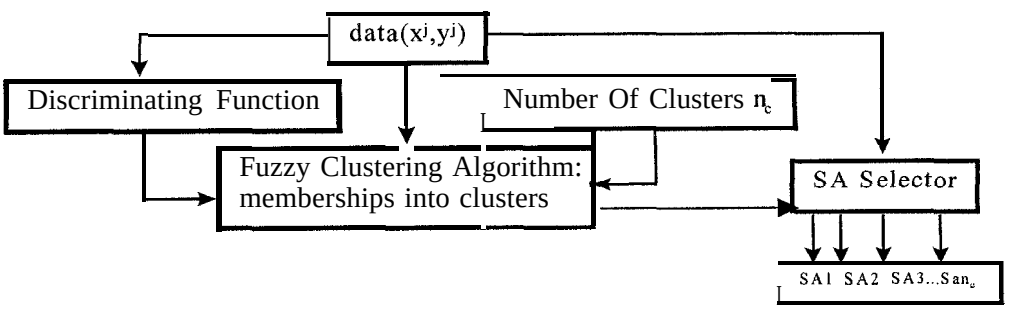


Figure 3: Training A Sugeno Approximator with A Fuzzy Clustering Algorithm

Applying the clustering algorithm to the original data [x, sin (x)] without recourse to a discriminating function leads to the clusters shown in Figure 4(a). We need to introduce a discriminating function that helps in splitting the data at the peaks. The sign of the derivative of the original function A (x,y) = sign[*df/dx*] acts like a good discriminating function because it captures the idea that the slope of f (x) has the same sign over each branch. Using this discriminating function, we apply the Fuzzy Clustering to the same functions used previously. Figure 4 (b) shows the result of applying fuzzy clustering to different samples of [x, sin (x) , +sign(cos (x))]. Hence the data is correctly split at the peaks of the particular function based on the number of branches the function consisted of (see Figure 3).

For ~(x), there exist six branches which call for six clusters. But even with the discriminating function the branches were not identified correctly by the clustering algorithm when the fuzzy clustering algorithm was asked to return 6 clusters. To circumvent this problem without excessive manipulation of the discrimination function, it is only needed to increase the number of clusters input to the clustering algorithm. Increasing the number of clusters to 8 does not solve the problem but increasing it to 10 does. The inverse function was approximated very well with 10 SAs shown in Figure 5 (b).

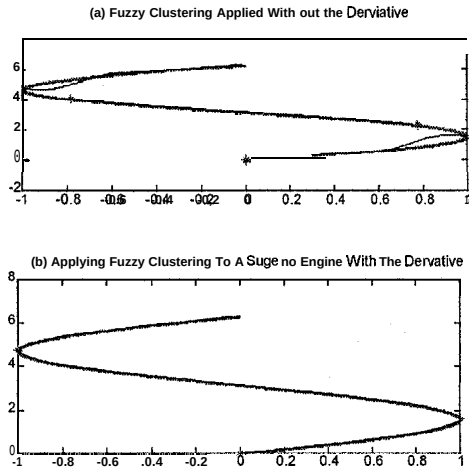


Figure 4: Sugeno inversion using Fuzzy Clustering applied to sin(x)

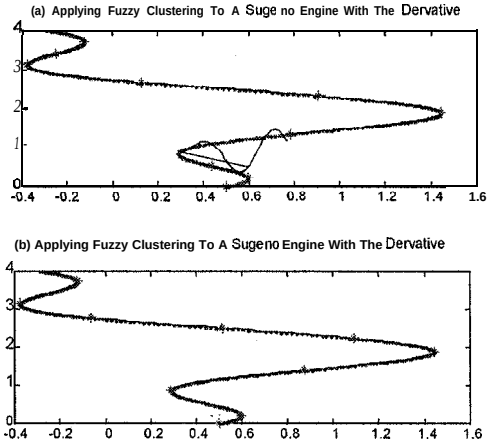


Figure 5: Sugeno Inversion using Fuzzy Clustering applied to f(x) =sin(x)+.5 cos(3.23x)

5. Conclusions

In this paper we have shown that a collection of Sugeno Approximators can successfully approximate inverse-multi-valued functions provided the function is split or clustered correctly, say by using a fuzzy clustering algorithm applied to the data in the form [x, f(x) , A (x, f(x))] where A (x, f(x) is a discriminating function that distinguishes the different branches. Determining the consequence coefficients of the training data is done by utilizing recursive least-squares and by utilizing a beta threshold value for eliminating unnecessary rules from the rule bases. For future work, the techniques discussed in this paper will also be applied to invert higher-dimensional functions.

6. Bibliography

[1] Buckley, J.J. "Sugeno type controllers are universal controllers" , *Fuzzy Sets and Systems 53,* pgs. **299-303, 1993**

[2] Clifton, C. A. Design and Applications of Generalized Sugeno Controllers , North Carolina Agriculture and Technical Thesis, 1992

[3] Hayashi, Y., Buckley, J.J., and Czogala, E. "Fuzzy Expert Systems Versus Neural Networks", Unpublished manuscript,1992

[4] Hall, L. O., Bensaid, A. M., Clarke, L. P., Velthuizen, R. P., Silbiger, M, S. and Bezdek, J. C., "A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Mangetic Resonance Images of the Brain", *IEEE Transactions on Neural Networks,* Vol. 3. No. 5., September 1992, pp. 672-682.

[5] G.-S.R. Jang and C.-T. Sun, E.Mizutani, Neuro-Fuzzy and Soft Computing, Prentice Hall, Upper Saddle River, New Jersey, 1997.

[6] Ross, T, Fuzzy Logic and Controller: Software and Hardware Applications, Prentice Hall, 1995

[7] Sate, M. and Sate, Y., "On A General Fuzzy Additive Clustering Model",*Intelligent Automation and Soft* Computing, Vol. 1, No, 4, pp. 439-448, 1995.

Acknowledgments: We would like to thank NASA Langley/Research Center, Graduate Student Research Program (GSRP) for providing the finding to conduct this research as well as NASA ACE Center at NCA& TSU for partial support,

98 URC070

Efficient Computation of Binomial Coefficients and Applications

Xiao-Ying Zhang[a] and Yong-Ming Lib
[a] Department of Computer Science, Alabama A&M University
[b] Intergraph Corporation, Huntsville, Alabama

Abstract

Binomial coefficients can be computed very efficiently by the construction of Pascal triangle. Analogous to this construction, we shall derive efficient methods to compute $b_k = \sum_{i=0}^{k} C_k^i a_i$ etc. Practical applications of such methods include polynomial basis conversion, reparametrisat ion, and sub division.

1 Introduction

In interpolation and approximation, polynomials are employed for data fitting [1]. In computational geometry, polynomial curves (e.g., Bézier and B-spline polynomial curves) are used to describe the shape of geometric entities [2-3]. For the purpose of evaluating or manipulating a polynomial, it is often required to convert a polynomial of one basis to another or subdivide an interval over which a polynomial is defined. Accordingly, we shall need to compute the following coefficients:

$$b_k = \sum_{i=0}^{k} C_k^i a_i, \quad b_k = \sum_{i=k}^{n} C_i^k a_i, \quad b_k = \sum_{i=0}^{k} C_{n-i}^{k-i} a_i, \cdots \quad (k = 0, 1, \cdots, n)$$

where,

$$C_n^i = \binom{n}{i} = \frac{n!}{(n-i)!\,i!}$$

are called the *binomial coefficients.* Assume a table of binomial coefficients is given. Then, a straight forward method to compute, for example, $b_k = \sum_{i=0}^{k} C_k^i p_i (k = 1,2, \ldots, n)$ is to multiply p_i by C_k^i and then add them together. However, even with the supply of binomial coefficients, this" simple method requires $n(n-1)/2$ multiplications and the same number of additions. In this paper, we shall derive an efficient a'gorithm to compute b_k. Using our method, only $n(n-1)/2$ additions are needed to compute all $b_k = \sum_{i=0}^{k} C_k^i p_i$ $(k = 0, 1, \ldots . n)$ in the *absence* of a table of binomial coefficients.

Our methods are derived analogously to the construction of Pascal triangle, the well-known algorithm to compute binomial coefficients. Some practical applications of our methods are given in this paper.

2 Computing $b_k = \sum_{i=0}^{k} C_k^i a_i$ and $b_k = \sum_{i=0}^{k} C_k^i a_i \beta^i$

We start our discussion by considering computation of binomial coefficients. It is taught in a high school that $C_n^i = C_{n-1}^{i-1} + C_{n-1}^i$. From this property, we can derive an efficient algorithm to compute binomial coefficients

```
c(0,o) = 1
do k=1, n
   c(k,O) = 1
```

```
   c(k,k) = 1
   do i=1,k-1
      c(k,i) = c(k-1,i-1) + c(k-1,i)
   enddo
enddo
```

This method is commonly known as the *construction of Pascal triangle.* It is seen that only $n(n-1)/2$ additions are needed to obtain all $C_k^i (k = 1, 2, \ldots, n; i = 1, 2, \cdots,$ k). When n = 3, the Pascal triangle is given by

$$\begin{array}{ccccc} & & 1 & & \\ & 1 & & 1 & \\ 1 & & 2 & & 1 \\ & 1 & & 3 & \\ & & 3 & & \\ & & & 1 & \end{array}$$

We now consider computation of $b_k = \sum_{i=0}^{k} C_k^i a_i$. For n = 3, we have

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_0 \\ a_0 + a_1 \\ a_0 + 2a_1 + a_2 \\ a_0 + 3a_1 + 3a_2 + a_3 \end{pmatrix}.$$

It is noted that the transformation matrix resembles the property of Pascal triangle. Therefore, analogous to the construction of Pascal triangle, we may derive the following method to compute b_k

```
Copy a(i) into b(i)
do k=1, n
do i=n, k, -1
   b(i) = b(i) + b(i-1)
enddo
```

This method requires only $n(n-1)/2$ additions. For verification, we look at the case where n = 3:

(k = 1)	(k = 2)	(k = 3)
b3=b3+b2=a3+a2	b3=b3+b2=a3+2a2+a1	b3=b3+b2=a3+3a2+3a1+a0
b2=b2+b1=a2+a1	b2=b2+b1=a2+2a1+a0	
b1=b1+b0=a1+a0		

We may extend the above method to compute $b_k = \sum_{i=0}^{k} C_k^i a_i \beta^i$:

```
Copy a(i) into b(i)
do k=1, n
do i=n, k, -1
   b(i) = b(i)* beta + b(i-1)
enddo
```

We now consider the application of the above method. In computational geometry, a Bézier polynomials commonly used to describe the shape of geometric entity. A Bézier polynomial of degree n maybe represented in terms of Bernstein polynomials [2-3]:

$$P(t) = \sum_{i=0}^{n} C_n^i (1 - t)^{n-i} t^i p_i.$$

For efficient evaluation of Bézier polynomial and its derivatives, we often convert it into a canonical polynomial so that Homer method can be employed. By the binomial expansion theorem, we have

$$\sum_{i=0}^{n} C_n^i (1-t)^{n-i} t^i p_i = \sum_{i=0}^{n} \sum_{j=0}^{n-i} (-1)^j C_n^i C_{n-i}^j p_i t^{i+j}.$$

Then, we want to find b_k such that

$$\sum_{k=0}^{n} b_k t^k = \sum_{i=0}^{n} \sum_{j=0}^{n-i} (-1)^j C_n^i C_{n-i}^j p_i t^{i+j}.$$

Since the coefficients of the same degree terms at both sides should be equal, we thus have

$$b_k = \sum_{i=0}^{k} \sum_{j=k-i}^{k-i} (-1)^j C_n^i C_{n-i}^j p_i t^{i+j} = \sum_{i=0}^{k} (-1)^{k-i} C_n^i C_{n-i}^{k-i} p_i.$$

From the relation $C_n^i C_{n-i}^{k-i} = C_n^k C_k^i$, we obtain

$$b_k = \sum_{i=0}^{k} (-1)^{k-i} C_n^k C_k^i p_i = (-1)^k C_n^k \sum_{i=0}^{k} (-1)^i C_k^i p_i.$$

To compute b_k efficiently, we first evaluate $\sum_{i=0}^{k} (-1)^i C_k^i p_i$ using the above method, then multiply the results by $(-1)^k C_n^k$ (noting that $C_n^k = \frac{n-k+1}{k} C_n^{k-1}$). The pseudo code is given as follows:

```
do  i=0, n, 2
   b(i) = a(i)
enddo
do i=1, n, 2
   b(i) = -a(i)
enddo

do k=1, n
do i=n, k, -1
   b(i) = b(i) + b(i-1)
enddo

t emp = 1.0
do i=1, n
   temp = -(n + 1 - i) * temp / i
   b(i) = b(i) * temp
enddo
```

It should be noted that the third block may be replaced by directly multiplying b_k by $(-1)^k C_n^k$ if one's system has a table of binomial coefficients.

3 Compute $b_k = \sum_{i=k}^{n} C_i^k a_i$, $b_k = \sum_{i=k}^{n} C_i^k a_i \beta^{i-k}$

In this section, we are concerned with efficient evaluation of $b_k = \sum_{i=k}^{n} C_i^k a_i$ with $k = 0,1,\text{-}$", n. For n = 3, we have

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_0 + a_1 + a_2 + a_3 \\ a_1 + 2a_2 + 3a_3 \\ a_2 + 3a_3 \\ a_3 \end{pmatrix}.$$

From this special case, we can see that the transpose of a transformation matrix resembles the property of Pascal triangle. Therefore, analogous to the construction of Pascal triangle we can derive the following algorithm to compute $b_k = \sum_{i=k}^{n} C_i^k a_i$:

```
Copy a(i) into b(i)
do k=0, n-1
do i=n-1, k, -1
   b(i) = b(i) + b(i+1)
enddo
```

We may extend the above method to compute $b_k = \sum_{j=k}^{n} C_j^k a_j \beta^{j-k}$:

```
Copy a(i) into b(i)
do k=0, n-1
do i=n-1, k, -1
   b(i) = b(i) + b(i+1) * beta
enddo
```

To demonstrate an application of the above method, we consider reparametrising a canonical polynomial

$$P(t) = \sum_{i=0}^{n} a_i t^i, \qquad t \in [a, b]$$

by $t = [(b+a) + (b-a)u]/2$ such that $P(u)$ is defined over [–1, 1]. Let $\alpha = (b-a)/2$ and $\beta = (b+a)/(b-a)$. Then, $t = \alpha(\beta + u)$. Therefore,

$$P(u) = \sum_{i=0}^{n} a_i \alpha^i (\beta + u)^i.$$

By the theorem of binomial expansion, we have

$$P(u) = \sum_{i=0}^{n} \sum_{j=0}^{i} a_i \alpha^i C_i^j \beta^j u^{i-j}.$$

We want to find b_k such that

$$\sum_{k=0}^{n} b_k u^k = \sum_{i=0}^{n} \sum_{j=0}^{i} a_i \alpha^i C_i^j \beta^j u^{i-j}.$$

Since the coefficients of the same degree terms on both side should be equal, we have

$$b_k = \sum_{i=k}^{n} \sum_{j=i-k}^{i-k} a_i \alpha^i C_i^j \beta^j = \sum_{i=k}^{n} a_i \alpha^i C_i^{i-k} \beta^{i-k} = \sum_{i=k}^{n} C_i^k (a_i \alpha^i) \beta^{i-k}.$$

Therefore, an efficient method to compute b_k would be

```
b(0) = a(0)
temp = alpha
do i=1, n
   b(i) = a(i) * temp
   temp = temp * alpha
enddo

do k=0, n-1
do i=n-1, k, step=-1
   b(i) = b(i) + b(i+1) * beta
enddo
```

4 Compute $b_k = \sum_{i=0}^{k} C_{n-i}^{k-i} a_i$, $b_k = \sum_{i=0}^{k} C_{n-i}^{k-i} a_i \beta^{k-i}$

In this section, we are concerned with efficient evaluation of $b_k = \sum_{i=0}^{k} C_{n-i}^{k-i} a_i$ with $k = 0, 1,$, n. For $n = 3$, we have

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_0 \\ 3a_0 + a_1 \\ 3a_0 + 2a_1 + a_2 \\ a_0 + a_1 + a_2 + a_3 \end{pmatrix}.$$

From this special case, we can see that the transpose of a transformation matrix resembles the property of Pascal triangle. Therefore, analogous to the construction of Pascal triangle, we give the following method to compute b_k:

```
Copy a(k) into b(k)
do k=0 , n
do i=1, n-k
   b(i) = b(i) + b(i-1)
enddo
```

This method can be extended to compute $b_k = \sum_{i=0}^{k} C_{n-i}^{k-i} a_i \beta^{k-i}$:

```
Copy a(k) into b(k)
do k=0, n
do i=1, n-k
   b(i) = b(i) + b(i-1) * beta
enddo
```

To illustrate an application of the above method, we consider reparametrising a vector-valued rational curve r(t) = $(x(t), y(t), z(t))$ by $t = t(u)$ such that

$$r(u) = r(t),\ r'(u)|_{u=0} = \alpha \mathbf{r}'(t)|_{t=0}.$$

Such reparametrisation is required whenever we want change the derivatives at the end points without affecting the geometric shape of the curve. In addition to the requirement of modifying the derivative at $t = O$, *we* also want to keep the degree of curve and the parameter interval unchanged. To meet all these requirements, the following bilinear transformation is a choice ,.

$$t = \frac{\alpha u}{(\alpha - 1)u + 1} = \frac{\alpha u}{\beta u + 1} \qquad (u \in [0,1],\ t \in [0,1]).$$

By the chain rule of differentiation we have $\mathbf{r}'(u) = \mathbf{r}'(t)t'$ **(u).** Then, it is readily checked that $r'(u)|_{u=0} = \alpha \mathbf{r}'(t)|_{t=0}$.

We denote r(t) = $(x(t), y(t), z(t))$ in homogeneous coordinate system by R(t) = (X(t), Y(t), *Z(t), W(t))* and write R(t) in the canonical form R(t) = $\sum_{k=0}^{n} \mathbf{a}_k t^k$, where $\mathbf{a}_k = (x_k, y_k, z_k, w_k)$ are vectors in homogeneous coordinates. Replacing t in R(t) by t = t(u), we have

$$R(u) = \sum_{i=0}^{n} \left(\frac{\alpha u}{\beta u + 1}\right)^i = \frac{1}{(\beta u + 1)^n} \sum_{i=0}^{n} \mathbf{a}_i (\alpha u)^i (\beta u + 1)^{n-i}.$$

Since R(u) is represented in homogeneous coordinates, the common factor $1/(\beta u + 1)^n$ can be discarded. Therefore, we have

$$R(u) = \sum_{i=0}^{n} \mathbf{a}_i (\alpha u)^i (\beta u + 1)^{n-i}.$$

By the binomial expansion theorem we obtain

$$\mathbf{R}(u) = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \mathbf{a}_i \alpha^i C_{n-i}^j \beta^j u^{i+j}.$$

We now want to find $\mathbf{b}_k$ such that

$$\sum_{k=0}^{n} \mathbf{b}_k u^k = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \mathbf{a}_i \alpha^i C_{n-i}^j \beta^j u^{i+j}.$$

Since the same degree terms at both side should be equal, we can then derive

$$\mathbf{b}_k = \sum_{i=0}^{k} \sum_{j=k-i}^{k-i} \mathbf{a}_i \alpha^i C_{n-i}^j \beta^j = \sum_{i=0}^{k} \mathbf{a}_i \alpha^i C_{n-i}^{k-i} \beta^{k-i}.$$

Let $\bar{\mathbf{a}}_i = \mathbf{a}_i \alpha^i$. Then, $\mathbf{b}_k = \sum_{i=0}^{k} C_{n-i}^{k-i} \bar{\mathbf{a}}_i \beta^{k-i}$ (k = **0,1, ""-,** n), which can be computed very efficiently using the method we discussed earlier.

5 Conclusions

Binomial coefficients can be computed very efficiently by the construction of Pascal triangle. Based on an observation of this construction, we derived efficient methods to compute $b_k = \sum_{i=0}^{k} C_k^i a_i$, $b_k = \sum_{i=k}^{n} C_i^k a_i$, etc. Applications of these methods in computational geometry and approximation were also discussed. Although not all possible combinations of the binomial coefficients and a_i were considered in this paper, one can analogously derive a respective method to meet his own requirement.

References

[1] Davis, P.J. (1975), Interpolation *& Approximation,* Dover Pub., Inc.

[2] Hoschek, J. and Lasser, D. (1993), *Computer Aided Geometric Design,* A.K. Peters, Ltd.

[3] Faux, I.D. and Pratt, M.J. (1979), *Computational Geometry for Design and Manufacture,* Ellis Horwood Ltd., London.

Effect of Fresh Poultry Litter and Compost on Soil Physical and Chemical Properties

Stacy Carr*, Teferi Tsegaye#, Ahmed Fahsi#, and Tommy Coleman*

* Michigan Technological University
# Center for Hydrology, Soil Climatology, and Remote Sensing
Alabama Agricultural and Mechanical University

Abstract

Application of poultry litter and compost as a substitute for fertilizer not only uses unwanted waste and decreases expenditures for commercial fertilizer, it adds nutrients to soil for plant uptake. The properties of soil affected by poultry litter were analyzed to determine the positive and negative aspects of using this substitute fertilizer. This study focused on changes associated with saturated hydraulic conductivity, bulk density, nitrate concentrations, and pH after application of varying concentrations of poultry litter and compost. Soil samples from Tennessee Valley Substation in Alabama were analyzed in a laboratory at Alabama A&M University. As a result of the application of fresh poultry litter and compost, we found that the saturated hydraulic conductivity increased and the bulk density decreased, while the pH was generally not affected. Using poultry litter and compost as an alternative commercial fertilizers could be adapted by the farming community to protect the sustainability of our environment. Unwanted waste is used productively and soil is enriched for farming.

Introduction

Hundreds of millions of chickens are produced each year for consumption. These chickens produce several thousand metric tons of litter per year. Litter is a mixture of feces, bedding, and water containing several nutrients of great importance for plant growth. The practice of applying litter, composted or fresh, to fertilize farming soil is gaining popularity. The nitrogen availability from litter is a main focus due to the high requirement of nitrogen by crops. It is important to note that only certain amounts of nitrogen are available in the first year. Its availability to plants must be studied to minimize environmental concerns as a result of runoff and leaching.

Nitrate leaching into groundwater systems is of great concern due to health and environmental implications. Many human health concerns arise from high concentrations of nitrate in drinking water. One such concern is the “blue baby syndrome”. Not only are health concerns involved, but environmental pollution of ammonia-nitrogen is toxic to fish and other aquatic organisms as well. Nitrogen input originating from ‘agriculture is most frequently cited as the cause of groundwater and surface water contamination by nitrates, although Adams et al. (1994) reported that a manure application rate providing approximately twice a crop’s nitrogen needs can be used with minimal transport of nitrate-nitrogen to groundwater when manure is applied to deep, well-drained soils and the crop is harvested.

Use of fresh litter and compost also increases organic matter content in the soil. This increase leads to increased water holding capacity, less susceptibility to erosion, improved soil structure, increased infiltration rates reducing runoff, and reduced crusting problems. Saturated hydraulic conductivity is important in understanding the movement of nitrate and other contaminants. Analysis of soil water retention is important due to its relationship with plant available water, infiltration, drainage, hydraulic conductivity, irrigation, and solute movement. Another effect is the increased soil buffering capacity against fluctuations in soil pH.

The use of poultry litter, as a substitute to traditional fertilizers, is an excellent way to minimize costs and to use unwanted waste. However, surface and groundwater standards may be violated due to runoff of this litter. Therefore, the litter concentrations, used as fertilizer, must be monitored to avoid such violations. The objectives of this research project are 1) to compare the effect of poultry litter on soil physical properties and 2) to compare the distribution of nitrate-nitrogen resulted from organic and inorganic sources of nitrogen in the upper top 30 cm of soil profile.

Methods and Materials

Varying degrees of poultry litter were applied to a field plot of the Tennessee Valley Substation in Belle Mina, northern Alabama. The field plot was divided into four replications and 20 treatments per replication. The plot size of the experiment field was 6 meters by 9 meters. Samples were collected at four locations within each plot using Global Positioning System (GPS) to georeference the sampling coordinate points,

Initial Preparation

The fertilizers used were urea, composted (CL) and fresh poultry litter (FL), two vitrification inhibitors (carboxy methyl pyrazole-CMP) treatments (with and without), and three nitrogen (N) rates (40, 80, and 120 kg N/ha). (Figure 1). Two nitrogen tests, with and without inhibitor, were included. Phosphorus and potassium fertilizers were applied to all plots to mitigate differences in phosphorus and potassium among urea and poultry litter treatments. The vitrification inhibitor was sprayed directly on the fertilizer source and on the soil for each control treatment plot. After the fertilizers were applied, the soil was tilled and mixed with a field cultivator immediately to minimize volatile loss of the vitrification inhibitor from the treated plots.

Figure 1: A schematic of the replications and treatments for the varying sections.

Replication 1									
80-Urea +CMP	40-CL+ CMP	80-FL+ CMP	120-FL+ CMP	CMP	40-FL+ CMP	80-CL+ CMP	Control	40-Urea	80-Urea
40-Urea +CMP	80-CL	40-FL	120-Urea + CMP	40-CL	120-Urea	120-CL	80-FL	120-FL	120-CL+ CMP
Replication 2									
40-FL+ CMP	40-FL	CMP	Control	80-Urea +CMP	80-CL+ CMP	40-CL	120-FL	120-CL	80-FL
80-Urea	120-Urea +CMP	120-FL +CMP	120-CL+ CMP	120-Urea	40-CL+ CMP	80-FL+ CMP	40-Urea +CMP	40-Urea	80-CL
Dedication 3									
40-FL+ CMP	40-CL	40-Urea +CMP	80-Urea +CMP	80-CL	120-Urea	40-FL	120-CL	120-FL+ CMP	80-FL+ CMP
CMP	120-Urea+ CMP	Control	120-CL+ CMP	120-FL	40-Urea	80-FL	80-Urea	40-CL+ CMP	80-CL+ CMP
Replication 4									
80-CL	40-CL+ CMP	40-Urea +CMP	80-FL+ CMP	80-CL+ CMP	1 20-FL	40-FL	120-CL+ CMP	CMP	40-CL
80-Urea +CMP	120-CL	40-Urea	40-FL+ CMP	120-FL+ CMP	80-FL	120-Urea	Control	80-Urea	120-Urea + CMP

** CMP= Inhibitor; FL= Fresh Litter; CL= Compost Litter; 40, 80, 120= kg N/ha

Sampling

Undisturbed and disturbed soil samples were taken in March 1997 from the research plots. Undisturbed soil cores, 7.6-cm length by 7.6-cm diameter, were collected using a Uhland core sampler (Blake and Hartge, 1986) for determination of soil physical properties for 0-8 centimeter soil depth. Samples were carefully trimmed, wrapped in plastic, transported to the laboratory, and stored in a refrigerator at 4^0C prior to analysis. These samples were used for hydraulic conductivity and bulk density analyses. Disturbed soil samples were obtained for nitrate-nitrogen, organic matter, and pH determination using two inch diameter Giddings probes (Murphy and Riley, 1962; Carter, 1993) sectioned at depth intervals of 0-15 and 15-30 centimeters.

Hydraulic Conductivity and Bulk Density

Upon thawing, the samples were weighed, then placed in a pan of water to be saturated for 24 hours before being weighed again. Saturated hydraulic conductivity (k_{sat}) was determined by pending 3.6-cm of water on top of each core under constant head conditions. A 24 hour equilibration period was allowed, after which outflow measurements were taken at 15, 30, 45 and 60 minute periods to calculate k_{sat} using Darcy's equation,

$$Q = \frac{-k A \partial h}{\sim} \quad (1)$$

Following k_{sat} determination, soil water retention was determined for each core at matric potentials of O, -2.0, -4.0,-6.0, -100, -500, -1000, and -1500 kilopascals (kPa). Matric potential values of -2.0 to -1500 kPa were achieved using a pressure plate apparatus (Klute et. al., 1986). The bulk density values were determined for each soil core after drying the soil core in an oven at 105°C over 24 hours. Soil water retention was expressed in terms of volumetric water content using the bulk density of individual cores for the conversion. Porosity for each soil core was determined by assuming a soil particle density of 2.65 gm/cm^3.

Nitrate Concentrations

A five gram soil sample was weighed on an electromagnetic balance. Twenty milliliters (mL) of de-ionized water was then added to the soil, which produced a 1:4 ratio. The sample was placed on a shaker for one hour. The solution was next filtered through Whatman 42 (125 millimeters) filter paper to extract soil nitrate-nitrogen. The liquid was collected in 20 mL bottles and 5 mL was transferred to autosampling vials. The vials were then placed in a dionex vial rack to be analyzed by the ion chromatography machine.

pH

The pH was determined by first mixing a 1:1 ratio of distilled water and soil. A pH meter was then placed in the mixture to determine the soil pH.

Results and Discussion

Hydraulic Conductivity and Bulk Density

The outflow volume collected during the laboratory measurement of k_{sat} is very crucial to minimize the measurement error. Measurement of k_{sat} by collecting outflow at different time periods (15 minutes, 30 minutes, 45 minutes and 60 minutes) indicated that k_{sat} values for each time period were highly correlated (Table 1). The correlation among the four time periods were very close, but the 30 and 45 time periods exhibited a correlation value of 1.00. The correlation between 15 and 60 minutes is 0.99, a value very close to one. It is advantageous, from a time restricted viewpoint, to use the 15 minute interval to determine k_{sat} when making measurements for a large number of samples. It is also important to realize that the k_{sat} value could be affected by timing of outflow. If the pores are not allowed to become saturated, not all pores will conduct uniformly. Likewise if the time period is too long, the smaller particles may disperse and plug pores, decreasing outflow.

Table 1: Correlation among the varying time periods used to calculate k_{sat}.

	15 minutes	30 minutes	45 minutes	60 minutes
15 minutes	1.00	0.95	0.94	0.99
30 minutes	0.95	1.00	1.00	0.93
45 minutes	0.94	1.00	1.00	0.94
60 minutes	0.99	0.93	0.94	1.00

Based on the above information, the k_{sat} values were obtained using the mean values for the 15, 30, 45 and 60 minute time periods. Generally, k_{sat} increases with litter applications due to high organic matter present in the soil. Saturated hydraulic conductivity had the greatest value (9.83 cm/hr) for the 80 kg N/ha urea and the 120 kg N/ha fresh litter plus inhibitor applications. The lowest k_{sat}, 1.84 cm/hr, resulted from the 120 kg N/ha application. The control treatment exhibited a k_{sat} value of 3.09 cd-n-, which is relatively low compared to most of the treated plots.

Soil water retention curves were shown in Figures 2a, 2b, and 2c. Application of fresh litter contributed to the increase of water retention in the micropore region at higher suction. The left side of the graph (lower pressure values) indicates where macropore flow dominates and water mainly drains faster. At the higher end of the suction, water will be tightly bound to the soil, making it difficult to become available to plants. Poultry litter usually increases pore size distribution and water holding capacity of a soil. In general, there was no clear evidence of increase in water holding capacity due to the addition of compost litter or fresh litter. It may take longer than one or two year application periods to see noticeable differences among treatments.

The general trend found in this study was that the bulk density decreased with application of fresh and compost litter. The lowest bulk density recorded was 1.29 g/cm^3 for 80 kg N/ha fresh litter plus inhibitor and the largest value was 1.46 g/cm^3 for 120 kg N/ha urea plus inhibitor. The control value was 1.43 g/cm^3. Porosity is a function of bulk density and, therefore, varied similar to bulk density. The range for porosity was between 0.45 and 0.52, with the porosity of the control at 0.46.

Nitrate Concentrations

Using t-tests for statistical analysis, it was determined that the mean nitrate concentrations varied with depth. As can be seen in Table 2, there is a significant difference between 0-15 cm and 15-30 cm, showing that nitrate-nitrogen leaches into the ground surface.

Table 2: Mean NO3-N concentrations for the 0-15 cm and 15-30 cm depths.

Depth	Mean NO3-N
0-15 cm	2.44[a] ppm
15-30 cm	3.29[b] ppm

Means followed by a different letter are significantly different from each other at probability level of 5%.

The results of the general linear models procedure showed that the only significant nitrate-nitrogen concentration difference was between the 80 kg N/ha fresh litter plus inhibitor treatment and the rest of the treatments. The remaining 19 treatments did not lead to nitrate-nitrogen concentrations that were significantly different from each other (Figure 3).

pH

Again, using the general linear models procedure, there was little significant difference found among the treated soils in regards to pH. The average pH values for all 20 sections for each replication are shown in Table 3. These values were not found to be statistically different from each other.

Table 3: This table details the mean pH for each replication including each of the 20 treatment sections.

Replication	pH
1	7.06[a]
2	6.99 [a]
3	6.96 ‘
4	6.94 [a]

Means followed by a different letter are significantly different from each other at probability level of 5%.

The average pH resulting from the individual treatments were also found to be quite similar, though some were found to be statistically different from each other. Table 4 lists the average soil pH for each treatment.

Figure 2a: The Volumetric Water Content as: Function of Pressure

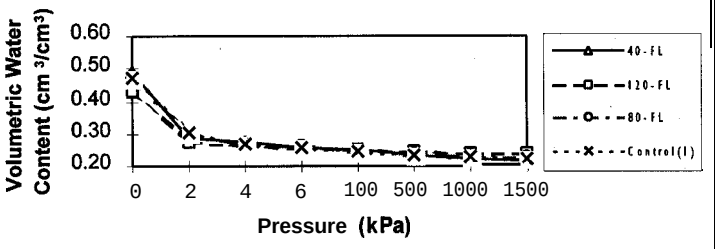


Figure 2b: The Volumetric Water Content as a Function of Pressure

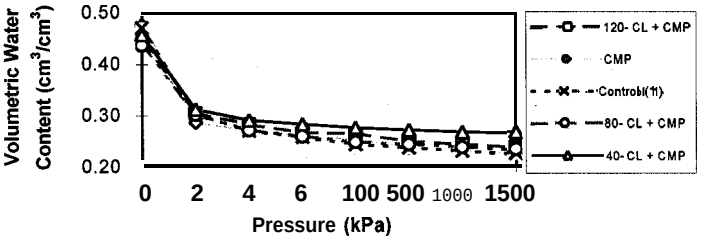


Figure 2c: The Volumetric Water Content as a Function of Pressure

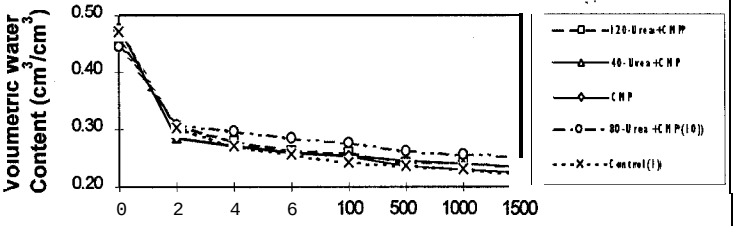


Table 4: Average soil pH for each treatment.

Treatment	pH	Treatment	pH	Treatment	pH
Control	7.02[ab]	40-CL+CMP	7.09[ab]	120-Urea	6.78[b]
CMP	6.95[ab]	80-Urea	6.90[ab]	120-Urea+CMP	7.07[ab]
40-Urea	6.95[b]	80-Urea+CMP	6.83"'	120-FL	7.18"
40-Urea+CMP	6.81[ab]	80-FL	6.85[ab]	120-FL+CMP	7.01[ab]
40-FL	6.99[ab]	80-FL+CMP	6.95[ab]	120-CL	7.17"
40-FL+CMP	6.99[ab]	80-CL	7.07[ab]	120-CL+CMP	7.11[ab]
40-CL	7.04[ab]	80-CL+CMP	7.01[ab]		

Means followed by a different letter are significantly different from each other at probability level of 5%.

Table 4 shows that the 120 kg N/ha urea treatment is statistically different from the 120 kg N/ha compost litter treatment and the 120 kg N/ha compost litter plus inhibitor treatment. The other values are similar.

Summary and Conclusion

Poultry litter is high in nutrients needed for plant growth and improves soil structure. Use of this litter increases soil organic matter content leading to increased water holding capacity, improved soil structure, and increased infiltration rates, which will reduce runoff. The application of poultry litter and compost increased the soil saturated hydraulic conductivity and water holding capacity and decreased the bulk density. Poultry litter and compost can be used as an alternative to commercial fertilizers by farmers to recycle unwanted waste and protect the sustainability of the environment.

Acknowledgment

The authors express their special thanks to Mr. Johnny Boggs and Mr. Harold Anthony for their tremendous laboratory and computer assistance. This work was supported by Grant No. NCCW-0084, from the National Aeronautics and Space Administration (NASA), Washington, DC. Any use of the trade, product or firm names is for the descriptive purpose only and does not imply endorsement by the U.S. Government.

References

Adams, P.L., et al. 1994. Poultry Litter and Manure Contributions to Nitrate Leaching through the Vadose Zone. ***Soil Science Society of America Journal,*** Vol. 58 No 4, pp 1206-1211.

Blake, G. R., and K. H. Hartge. 1986. Bulk Density. In. A. Klute et al. (Ed.), Methods of Soil Analysis. Part I. 2nd ed. ***Agronomy*** 9:363-375.

Carter, M.R. 1993. Soil Sampling and Methods of Analysis, ***Canadian Society of Soil Science,*** Lewis Publisher, CRC Press, Inc., Florida, , p 843.

Klute, A. 1986. Water retention: Laboratory methods. In. A. Klute et al. (Ed.), Methods of Soil Analysis. Part I. 2nd ed. ***Agronomy*** 9:635-662.

Murphy, J. and J.P. Riley. 1962. A Modified Single Solution Method for the Determination of Total Organic Phosphorus in Soil, ***Journal of Soil Science,*** 6:254-267.

Development of a Coupled Hydrological/Sediment Yield Model for a Watershed at Regional Level

Narayan Rajbhandari[1], William Crosson[2], Teferi Tsegaye[1], Tommy Coleman[1]
Yaping Liu[1], and Vishwas Soman[2]

[1]Center for Hydrology, Soil Climatology, and Remote Sensing
Plant and Soil Science Department
Alabama A&M University
Normal, AL 35762.
[2]Global Hydrology and Climate Center
977 Explorer Blvd.
Huntsville, AL 35806.

Introduction

Development of a hydrologic model for the study of environmental conservation requires a comprehensive understanding of individual-storm affecting hydrologic and sedimentologic processes. The hydrologic models that we are currently coupling are the Simulator for Hydrology and Energy Exchange at the Land Surface (SHEELS) and the Distributed Runoff Model (DRUM). SHEELS runs continuously to estimate surface energy fluxes and sub-surface soil water fluxes, while DRUM operates during and following precipitation events to predict surface runoff and peak flow through channel routing. The lateral re-distribution of surface water determined by DRUM is passed to SHEELS, which then adjusts soil water contents throughout the profile. The model SHEELS is well documented in Smith et al. (1993) and Laymen and Crosson (1995). The model DRUM is well documented in Vieux et al. (1990) and Vieux and Gauer (1994). The coupled hydrologic model, SHEELS/DRUM, does not simulate sedimentologic processes. The simulation of the sedimentologic process is important for environmental conservation planning and management. Therefore, we attempted to develop a conceptual frame work for coupling a sediment yield model with SHEELS/DRUM to estimate individual-storm sediment yield from a watershed at a regional level.

The sediment yield model that will be used for this study is the Universal Soil Loss Equation (USLE) with some modifications to enable the model to predict individual-storm sediment yield. The predicted sediment yield does not include wind erosion and erosion caused by irrigation and snow melt. Units used for this study are those given by Foster et al. (1981) for SI units.

Sediment Yield Model Development

Explicit patterns of land use, soils, and topography are required to run the hydrologic model, SHEELS/DRUM. Slope and aspect at each grid cell are derived from a Digital Elevation Model (DEM) and control the lateral movement of surface water. During a storm event, the hydrologic model is driven by gridded rainfall and predicts the hydrologic parameters such as surface runoff, peak flow, and channel flow. These hydrologic parameters will be utilized by the sediment yield model, USLE, to predict individual-storm sediment yield from each grid cell. The total sediment reaching the watershed outlet will be estimated using the channel sediment routing process.

The USLE model is designed to predict long term annual soil loss using the equation:

$$Y = R*K*L*S*C*P \quad (1)$$

where, Y is the sediment yield in tons per hectare, *R is* the rainfall energy factor (SI unit), *K* is the soil erodibity factor (SI unit), *L is the* slope length factor (unitless), S is the slope steepness factor (unitless), C is the cover and management factor (unitless), and *P is the* conservation practice factor (unitless).

The R factor

A major weakness of equation (1) for an individual-storm event is the failure of the R factor to adequately account for peak rate and total volume of surface runoff. In the USLE model, the R factor is the product of rainfall energy (E) and maximum 30-min intensity (I_{30}) divided by a constant. In an individual-storm event, soil loss is largely influenced by total volume of surface runoff and peak rate. Williams (1975) and Onstad and Foster (1975) evaluated the R factor and proposed following equations for individual-storm events.

$$R_w = 27.06\ A^{0.12}\ (Q\ q_p)^{0.56} \qquad (2)$$

$$R_o = 0.5\ R + 3.42\ Q\ q_p^{\frac{1}{3}} \qquad (3)$$

where R_w and R_o are the rainfall energy factors proposed by Williams and Onstad and Foster respectively, A is the drainage area, Q is the volume of surface runoff, and q_p is the peak flow rate.

The K factor

The K factor is a measure of a soil's resistance to the erosive powers of rainfall energy and runoff. The USDA-SCS (1978) has developed estimates of K based on textural classification for topsoil, subsoil, and residuals. The textural classification documents will be used to estimate the K factor in this study.

The L and S factors

The L and S factors are accounted for both rill and interill erosion impacts. McCool et al. (1993) recommended the following equations to estimate the L and S factors:

$$L = \left[\frac{\lambda}{22.1}\right]^m \qquad (4)$$

where λ is the slope length in meter, and m is a variable slope length exponent which is related to the ratio of rill to interill erosion, β, by

$$m = \frac{\beta}{1+\beta} \qquad (5)$$

For a slope which is moderately susceptible to erosion, β is estimated as:

$$\beta = \frac{11.16\ \sin\theta}{3.0\ (\sin\theta)^{0.8} + 0.56} \qquad (6)$$

where θ is the slope angle.

β is multiplied by 2 for a soil which is highly susceptible to erosion. The multiplication is 0.5 for a soil which is less susceptible.

For a slope length less than 5 meter, the S factor is estimated as:

$$S = 3.0\ (\sin\theta)^{0.8} + 0.56 \qquad (7)$$

For a slope length greater than 5 meter where rill erosion will be initiated, the S factor is estimated as

$$S = 10.8\ \sin\theta + 0.03 \qquad \sin\theta < 0.09 \qquad (8)$$

$$S = 16.8\ \sin\theta - 0.50 \qquad \sin\theta \geq 0.09 \qquad (9)$$

Equation (9) will be replaced by equation (10), below, under condition where thawing of recently tilled soils is occurring and surface runoff is the primary factor causing erosion.

$$S = 4.25 \ (\sin \theta)^{0.6} \tag{lo}$$

The C Factor

The C factor accounts for the effects of cover above the ground, ground cover, incorporated residue, and surface roughness on soil erosion. The C factor is estimated as:

$$C = \exp[(-0.2231 - C_m) \exp(-0.00115 S_c) + Cm] \tag{11}$$

where SC is the soil cover in $kg.ha^{-1}$ and C_m is the minimum value of C. The value of C_m is estimated from the average annual C factor using the equation as:

$$C_m = 1.463 \text{ in } (C_a) + 0.1034 \tag{12}$$

The value of C_a is estimated from table prepared by Wischmeier and Smith (1978).

The P Factor

The P factor is the ratio of soil loss from any conservation support practice to that with up- and downslope tillage. It is an important parameters for evaluating the effects of counter tillage, strip cropping, terracing, surface drainage, and dry land farm surface roughening. The P factor is estimated using information provided in Wischmeier and Smith (1978).

Channel Sediment Routing

The above equations (1) to (12) will be used to estimate sediment yield from each grid cell. The total sediment yield from the watershed will be estimated using channel sediment routing process. The routing process consists of sediment deposition and sediment degradation. The sediment deposition in the channel is based on the Stokes's low, which is defined as:

$$S_{fx} = 411 \ d^2 \ T \tag{13}$$

where S' is the sediment of particle size d which will fall to a depth of x during the travel time T.

The final deposition, S_{dep}, is estimated as:

$$s_{dep} = S_{rec} \ (1 - DR) \tag{14}$$

where S_{rec} is the sediment received by the channnel. *DR is the* delivery ratio which is estimated as:

$$DR = \frac{(1 \ 0.5) \ s'}{Y} \qquad S_{fx} \leq Y \tag{15}$$

$$DR \cdot \frac{0.5 \ Y}{s_{fx}} \qquad S_{fx} > Y \tag{16}$$

where Y is the depth of flow.

Sediment degradation in the channel routing will be estimated using Williams's (1980) and Bagnold's (1977) definition of stream power which is defined as:

$$s_{degd} = [\alpha_{sp}\ \gamma^{1.5}\ (dur)\ (w)\ (Y\ (WS)\ V_c)^{1.5}]\ [1\ +\ K\ C\ ]\ [1\ -\ DR] \quad (17)$$

where α_{sp} is a parameter dependent on maximum stream power for the reach, γ is the density of water, dur is the flow duration in hour, w is the channel width at flow depth, WS is the water surface slope, and V_c is the velocity in the channel, and K and C are the USLE factors for the channel.

The total sediment leaving the watershed is then estimated as:

$$\sum S_{out} = \sum S_{rec} - \sum S_{dep} + \sum S_{degd} \quad (18)$$

Coupling Strategy

In the coupled model, the sediment yield model, USLE, will receive the hydrologic parameters such as rainfall energy (E) rainfall intensity (I_{30}), runoff volume (Q), peak runoff rate (q_p), soil texture classes, grid cell area, slope length, and slope angle in each grid cell from SHEELS/DRUM as an input. The USLE will then estimate sediment yield for each grid cell. The channel routing components of SHEELS/DRUM will read the USLE outputs and will estimate the sediment deposition and sediment degradation using the above equations (13 -17). Finally, the coupled model will estimate the total sediment leaving the watershed using equation 18.

Testing Storm Soil Loss Simulation

We tested the USLE model using the R_w and R_o values for individual-storm events to evaluate the model's effectiveness for predicting individual-storm sediment yield. We obtained field measurement sediment yield data from a small agricultural watershed at the Oxford Tobacco Research Station in Granville County, North Carolina. The size of the watershed was 1.50 ha. The watershed contained an upland agriculture with a grass vegetated zone at the down slope edge and adjacent forested zone (O. 16 ha.) that received runoff from the field. We evaluated the model for the individual-storm sediment yield from the field during 1992. The agricultural crop, tobacco, was planted in late April and was harvested in mid-August through mid-October. The details about the watershed and the sediment analysis are discussed in Rajbhandari et al. (1996).

The USLE parameters -K, L, S, and P remained constant for the agricultural field and were estimated at 0.044, 1.877,0.338, and 0.38 respectively. The C factors for the planting period, harvest period, and seed bed preparation were estimated at 0.05, 0.2 and 0.2 respectively. Calculated values of R_w, R_o, and R_u are compared in Figure 1. For the majority of the stormflow events, R_w exceeded R_o. The differences were considerably large during the large stormflow events. The case for R_o and R_u during the large stormflow events was also similar.

Owing to the different values of R_w and R_o, the predicted sediment yields at the outlet of the field also differed accordingly (Figure 2). The USLE model using R_w predicted accordingly the measured sediment yield during the tobacco plantating period. On the another hand, the model using R_o predicted closely during the harvest and seed bed preparation periods. The former model predicted the sediment yield remarkably higher than the latter model during the high stormflow events (Julian days 248 and 249). Laflen et al. (1990) also found similar results.

Conclusion

Since the hydrologic model, SHEELS/DRUM, simulates surface runoff, peak flow rate, and channel routing, an opportunity exists to utilize the sediment yield model, USLE, with some modifications for predicting sediment yields for individual-storm events at a regional level watershed. The model using R_w did not respond well during the tobacco harvesting period, soil work period, and high storm event period. In contrast, the model using R_o did not respond well during the tobacco plantating period. These results raised a considerable debate as to which model to use. At this stage, we plan to include both options for comparative study for users. In thenear future, these findings will be further validated with more data for different crop systems in a regional watershed level.

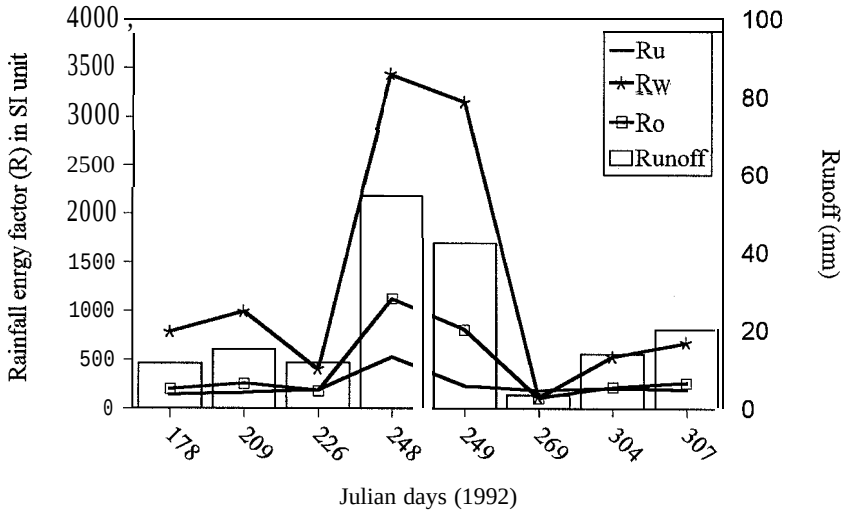


Figure 1. Comparison of the rainfall energy factors - Rw, Ro, and Ru as identified by Willams (1975), Onstad and Foster (1975), and USLE respectively. The data was obtained from the Oxford Tobacco Research Station in Granville County, North Carolina.

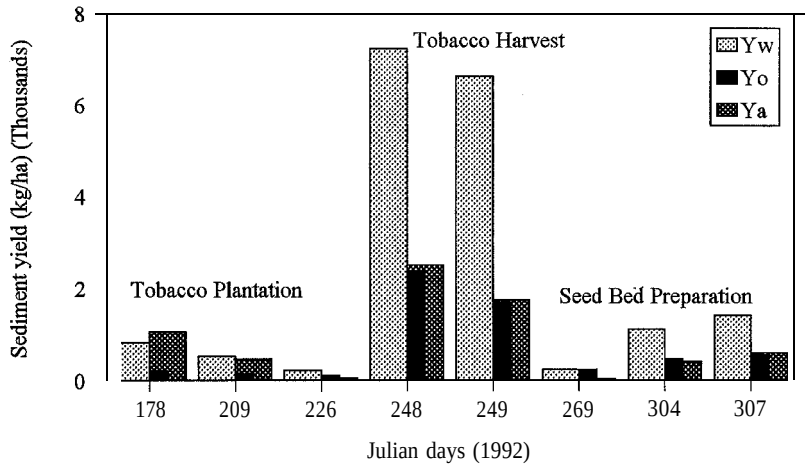


Figure **2.** Predicted and observed values of sediment yield from a small agricultural field at the Oxford Tobacco Research Station in Granville County, North Carolina, during and after tobacco plantating in 1992. The predicted sediment yields using the USLE model with Williams's modification and Onstad-Foster's modification are respectively represented by Y_w and Y_o. The field measurement sediment yield is represented by Y_a.

References

Bagnold, R. A. 1977. Bedload transport in natural rivers. Water Resources Research, 13(2):303-312

Foster, G. R., D. K. McCool, K. G. Renard, and W. C. Moldenhauer. 1981. Conversion of the Universal Soil Loss Equation to SI metric units. J. Soil Water Conservation. 36:355-359.

Laflen, J. M., G. R. Foster, and C. A. Onstad. 1990. Computation of Universal Soil Loss Equation R and C Factors for simulating individual-storm soil loss, pp 125-127. In EPIC-Erosion/Prodctivity Impact Calculator, Model Documentation. USDA, ARS, Tech. Bulletin No, 1768.

Laymen, C. A., and W.L. Crosson, 1995: Regional-scale hydrology with a new land surface model. Preprint Vol. of Conf. on Hydrology (January 15-20; Dallas, TX), Amer. Meteor. Sot., Boston, MA.

McCool, D. K., Foster, G. R., and Weesies, G. A. 1993. Slope length and steepness factor. In Predicting Soil Erosion - A Guide to Conservation Planning with the RUSLE. USDA-ARS Special Publication.

Onstad, C. A., and G. R. Foster. 1975. Erosion modeling on a watershed. Trans. Am. Sot. Agric. Eng. 18:288-292.

Rajbhandari, N.B., J.D. Gregory, C. Franklin, and J.E. Parsons. 1996. Dispersal of agricultural runoff through a level spreader in a forested filter zone enhances sediment and nutrient removal, Proceedings, Water Quality Conference. March 19-20, 1996, North Carolina State University, Raleigh, USA.

Smith, E. A., W.L. Crosson, H.J. Cooper and H.-Y. Weng, 1993. Estimation of surface heat and moisture fluxes over a prairie grassland. Part III: Design of a hybrid physical/remote sensing biosphere model. J. Geophys. Res., 98,4951-4978.

Soil Conservation Service. ! 978. Water management and sediment control for Urbanizing areas. USDA - SCS, Columbus, OH.

Vieux, B. E., V.F. Bralts, L.J. Segerlind and R.B. Wallace, 1990. Finite element watershed modeling: One-dimensional elements. J. Water Res. Planning and Management., 116,803-819.

Vieux, B.E. and N. Gauer, 1994. Finite element modeling of storm water runoff using GRASS GIS. Microcomputers in Civ. Eng., 9,263-270.

Williams, J. R. 1980. SPNM, a model for predicting sediment, phosphorus, and nitrogen from agricultural basin. Water Resources Bulletin, 16(5):843-848.

Williams, J. R. 1975. Sediment yield predictions with Universal Soil Loss Equation using runoff energy factor, pp 244-252. In Present and Prospective Technology for Predicting Sediment Yields and Sources. U. S. Dept. Agric., Res. Serv. (Publ) ARS-S-40.

Wischmeier, W. H., and D. D. Smith. 1978. Predicting rainfall erosion losses. U. S. Dep. Agric., Agric. Handbook. 537.

Acknowledgment

We acknowledge the Oxford Research Project, Department of forestry, North Carolina State University, Raleigh, NC., for providing data for this study.

Change Detection Analysis in Urban and Suburban Areas Using Landsat Thematic Mapper data: Case of Huntsville, Alabama

Dana Kuan*, A. Fahsi#, S. Steinfeld†, and T. Coleman#
* University of Maryland at College Park
# Center of Hydrology, Soil Climatology and Remote Sensing
Alabama Agricultural and Mechanical University
† Pennsylvania State University

ABSTRACT

Two Landsat Thematic Mapper (TM) images, from July 1984 and July 1992, were used to identify land use/cover changes in the urban and suburban fringe of the city of Huntsville, Alabama. Image difference was the technique used to quantify the change between the two dates. The eight-year period showed a 16 % change, mainly from agricultural lands to urban areas generated by the settlement of industrial, commercial, and residential areas. Visual analysis of the change map (i.e., difference image) supported this phenomenon by showing that most changes were occurring in the vicinity of the major roads and highways across the city.

INTRODUCTION

Huntsville, an industrial city of north Alabama, is currently one of the most changing urban areas in the southeastern region of the United States. These changes have been driven by an increasing growth of industrial and commercial companies leading to a rapidly growing population. Consequently, the dynamic interaction between these changes and the demographic expansion of the city causes increasing changes in land use/cover of the urban and suburban areas. This has resulted in the loss of agricultural fields to make room for the expanding urban and suburban establishment.

The purpose of this study was to map and quantify the land use/cover changes in north Alabama over an eight-year time span using Landsat TM images. By showing the change between an eight-year period, this study may help managers and decision makers better understand the present regional growth and prepare the city of Huntsville and its surrounding areas for future expansion.

The monitoring of land use/cover changes has long been a concern for developers, managers, and decision makers. Accurate quantification of this phenomenon requires knowledge of the spatial and temporal variability of the occurring changes. Various techniques have been suggested to monitor land use/cover changes in urban and suburban areas using satellite remotely sensed data (Azerzaq et al., 1995; Green et al., 1994; Lambin and Strahler, 1994a; Price et al., 1992; Westmoreland and Stow, 1992; Byrne et al., 1990; Ehlers et al., 1990; Martin and Howarth, 1989; Martin, 1986; Jensen, 1982). The traditional method of visual comparison using aerial photographs is usually tedious, time consuming, and often subject to omission errors. Therefore, most of the recent techniques are based on satellite remotely sensed data. With the help of satellite imagery, change detection is especially advantageous for large areas and the analysis is much more easily accomplished.

Lambin and Strahler (1994b) showed that the detection of land cover changes processes by remote sensing is improved when using both spectral (e.g., vegetation index and surface temperature) and spatial indicators of surface condition. With its capability of high spectral and spatial resolution, Landsat TM data has been a major source for change detection analyses.

A fundamental assumption when using satellite remotely sensed data for change detection is that any change in land use/cover causes a change in spectral response of that feature. Several methods have been developed to quantify change detection between multi-date images. These include: Change detection using Write Function Memory Insertion, Multi-date Composite Image change detection, Image Algebra change detection, Post-Classification Comparison change detection, Multi-date change detection using a binary mask applied to date 2, Multi-date change detection using ancillary data source as date 1, Manual, on-screen digitization of change, Spectral Change Vector Analysis, and Knowledge-based Vision Systems for detecting change (Jensen, 1996). Image difference (Jensen, 1996; Griffits, 1988), Transformed Tassled Cap (Fung and LeDrew, 1987b), and Principal Component Analysis

(Eastman and Fulk, 1993; Singh, 1989; Fung and LeDrew, 1987a) are generally the most common techniques used in change detection. Other most recent techniques may involve ancillary data to improve the analysis.

Among the various existing change detection quantification methods, we adopted the image difference technique for its simplicity but also because of the pixel by pixel subtraction between the two images.

METHODOLOGY

Characteristics of the study site

Huntsville, the economical and industrial city of Alabama, shelters approximately 180,000 people on 168 square mile area. Although the city is still under moderate population density, it is one of the most rapidly growing urban areas in the region. The settlement of an increasing number of industrial and commercial companies, such as NASA's Marshal Space Flight Center, Intergraph, and others, constitute one of the major incentives to attract workers and organizations to the area. Because of its inherent power and opportunities to attract people, the city and its suburbs are under a continuous modification from agricultural to residential areas. These characteristics make this site adequate for change detection analysis. The study area encompasses a portion of Huntsville urban area and extends over the Madison County, northeast of the city of Huntsville. The area covered by the site is about 25km by 25km.

Data and equipment used

One of the most important considerations in multi-date composite image change detection analysis is the time of data acquisition. Jensen (1996) reported that images acquired in summer (i.e., sunny periods) show a good contrast among surface features, particularly in urban and suburban areas. The author also suggested the use of images acquired during the same period of the year to reduce the effects due to different sun angles, changes in vegetation physiology, and difference in surface soil moisture. In this study, we used two Landsat TM images from July 1984 and July 1992. Data processing and analysis were performed utilizing Sun Spare 10 Unix Workstation and Erdas/Imagine 8.2 image processing software.

Data processing

Radiometric rectification

Traditionally, change detection has been determined by classifying the individual multi-date images and subtracted from one another. This technique may introduce enormous errors due to the fact that, in supervised classification, training sites used to classify an image are delineated separately for each image (Chuvieco and Congalton, 1988), which may result in different errors for each classified image. Similarly, in unsupervised classification, pixels of similar ground cover may be classified under different classes because of the spectral differences between the two images due to sensor, atmospheric condition, and ground conditions.

In this study, we used a technique developed by Hall et al. (1991), which consists essentially of registering one image (subject image) to another (reference image) to appear as if they were both imaged with the same sensor and under similar atmospheric and illumination conditions. This technique removes radiometric noises so that only changed features are accounted for (Gong, 1992). The 1992 image was used as a reference because of its lower minimal values in all bands, suggesting that the data is less affected by the atmosphere. The technique uses a linear transformation between the two images to convert the subject image to the reference image so that the two images are restored to a common radiometric response. The linear transformation is calculated by relating a set (i.e., the means of a sample of points) of the darkest and the brightest pixel values from the two images such that:

$$Y_i = a, + b_i X_i \quad (1)$$

Therefore, the transformed means of the radiometric control sets in the subject image to the reference image can be expressed as (Hall et al., 1991):

$$D_{ri} = a_i + b_{si} D_{si} \quad \text{for the dark control sets}$$
$$B_{ri} = a_i + b_{si} B_{si} \quad \text{for the bright control sets} \qquad (2)$$

By solving the system of equations, we obtain:

$$b_i = \frac{B_{ri} - D_{ri}}{B_{si} - D_{si}} \quad \text{and,} \qquad (3)$$

$$a_i = \frac{D_{ri} B_{si} - D_{si} B_{ri}}{B_{si} - D_{si}} \qquad (4)$$

where D_{ri}, D_{si}, B_{ri}, and B_{si} are the dark (D) and the bright (B) control sets for the reference image (r) and the untransformed subject image (s) in the i^{th} band (i).

Image differencing

To produce a difference or *residual* image, the two images must be geometrically registered (Novak, 1992). The resulting image represents the change between the two multi-date images. Theoretically, pixels with zero values indicate no change between the two images and pixels with positive or negative values are areas that have changed in land use/cover. The computation equation is given by Singh (1989) as:

$$DX_{k,ij} = \frac{X_{k,ij}(t_2) - X_{k,ij}(t_1)}{2} + 127 \qquad (5)$$

and by Jensen (1996) as:

$$DX_{k,ij} = [X_{k,ij}(t_2) - X_{k,ij}(t_1)] + 255 \qquad (6)$$

where :

$DX_{k,ij}$ is the difference in brightness for pixel $X_{k,ij}$ between time 1 (t_1) and time 2 (t_2),
k is the spectral channel, and
i,j is the pixel location in band k (i = row and j = column).

In an 8-bit analysis, the values in the difference image may range from -255 to 255. In the above equations (5 and 6), a constant of 127 (in Singh's case) or 255 (in Jensen's case) was added to convert the results into a positive number, The outcome is a new image, where each band is approximately normally distributed. Pixels indicating no change are distributed around the mean (approximately the value of 127) and pixels indicating change occur at the tails of the histogram. Decision on the limit between change and no change is very important in change detection analysis. The analyst may interactively use various thresholds before selecting the optimum one (Fung and LeDrew, 1988). Often, one standard deviation from the mean is used, which was the case in this study.

After selecting the threshold to determine the change areas between 1984 and 1992, the resulting image was converted into a binary format, where zeros indicate no change and ones indicate change.

RESULTS

The change detection image of the urban and suburban fringe of Huntsville, obtained by comparing two multi-date images (1984 and 1992), showed a great amount of change in the land use/cover over an eight-year period (Fig. 1). The results presented in this paper do not quantify change detection by land use/cover category, they only present changed versus non-changed areas. The statistical results indicated that about 16 % of the land use/cover inside the study site has changed between 1984 and 1992. The visual inspection of the difference image (Fig. 1) showed that most of the changes occurred in the vicinity of major roads and highways across the city and its suburbs, where

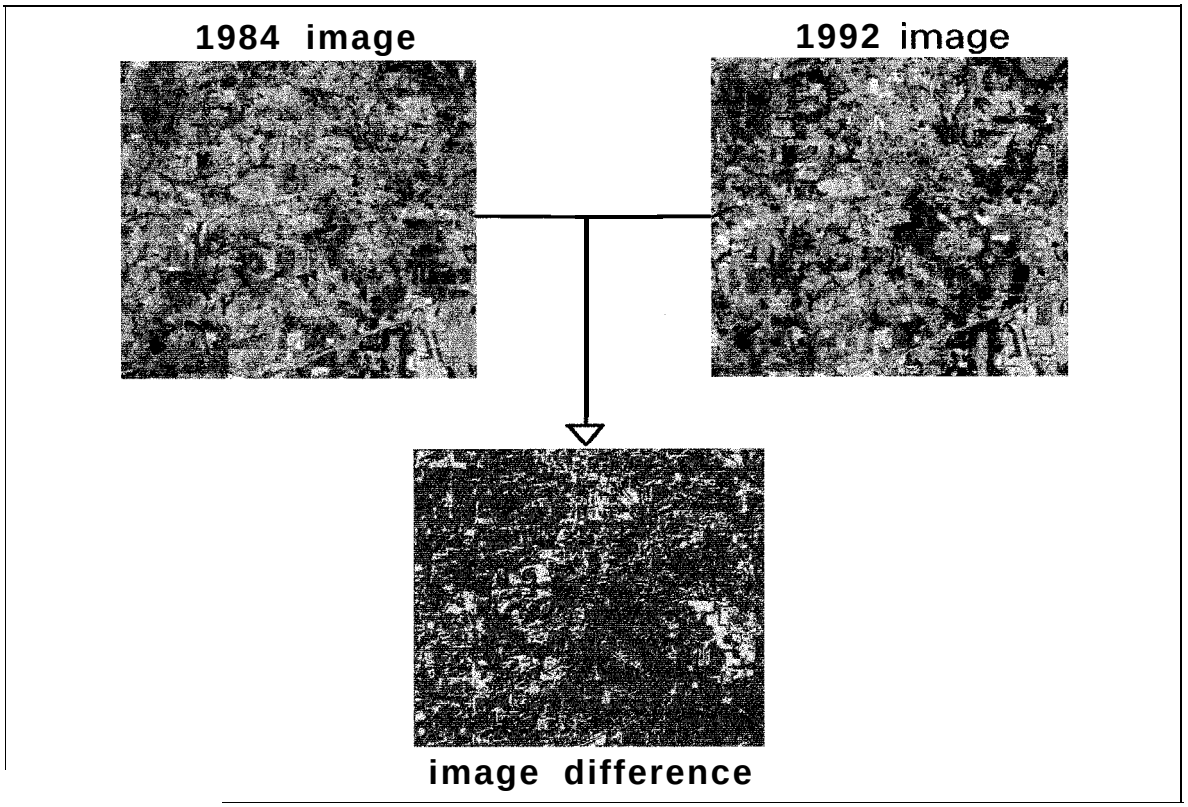


Figure 1. Image differencing illustrating change/no change of the urban and suburban areas of Huntsville, AL between 1984 and 1992.

industrial and commercial activities are located. Areas of no-change are mostly parks, farms, and recreation sites. Considerable changes are shown in the Madison area and along the major highways running through the city.

Although the map shows considerable changes between 1984 and 1992, not all the changes are necessarily due to land use/cover modification. Because of the approximate threshold used to quantify change versus no-change (i.e., one standard deviation from the mean value in the difference image), the difference image may exhibit some errors. Another source of errors may come from spectral response difference of the same cover type but at different development stages between 1984 and 1992.

CONCLUSIONS

Through this study, we have demonstrated that satellite remotely sensed data, such as Landsat-TM, are adequate for detecting changes in urban and suburban areas. Landsat-TM data analysis showed that the urban and suburban areas of Huntsville, had a considerable change in land use/cover over an eight-year span. This can be directly attributed to the industries and commercial centers that are settling in the city, attracting an increasing number of workers. This phenomenon has led to a continuous population growth resulting in loss of agricultural land in favor of industrial, commercial and residential settlements. 16 % of agricultural lands were converted to urban areas over an eight year period.

The pace of change in the urban and suburban fringe of the city of Huntsville, Alabama, are far from decelerating, which may require further studies to better plan for the future development of the area.

ACKNOWLEDGMENTS

The authors express their special thanks to Mr. Webb Tadesse and Mr. Donvila Williams for their tremendous computer and image processing assistance. This work was supported by Grant No. NCCW-0084, from the National Aeronautics and Space Administration (NASA), Washington, DC. Any use of the trade, product or firm names is for the descriptive purpose only and does not imply endorsement by the U.S. Government.

REFERENCES

Azerzaq M., M. Assafi, and A. Fahsi. 1995. The Use of Spot-HRV images for Suburban Change Detection Assessment in Casablanca (Morocco). Télédétection des Milieux Urbains et Périurbains. *AUPELF-UREF - Actes des Journées Scientifiques de Liège*, Oct. 2-5, 1995, Liege, Belgium, pp. 275-282 (in French).

Byrne, G. F., P. F. Grapper, and K. K. May. 1990. Monitoring Land Cover Changes by Principal Component Analysis of Multitemporal Landsat Data. *Remote Sensing of Environment,* 10:175-184.

Chuvieco, E. and R. G. Congalton, 1988, "Using Cluster Analysis to Improve the Selection of Training of Statistics in Classifying Remotely Sensed Data," *Photogrammetric Engineering & Remote Sensing 54 (9): 1275-1281.*

Eastman, J. R. and C. J. Fulk, 1993, "Long Sequence Time Series Evaluation Using Standardized Principal Components," *Photogrammetric Engineering & Remote Sensing, 59* (6):99 1-996.

Ehlers, M., M. A. Jadkowski, R. R. Howard, and D. E. Brostuen. 1990. Application of SPOT Data for Regional Growth Analysis and Local Planning. *Photogrammetric Engineering and Remote Sensing, 56(2): 175-180.*

Fung, T. and E. LeDrew, 1988. "The Determination of Optimal Threshold Levels for Change Detection Using Various Accuracy Indices," *Photogrammetric Engineering & Remote Sensing, 54* (10): *1449-1454.*

Fung, T. and E. LeDrew, 1987a, "Application of Principal Components Analysis for Change Detection," *Photogrammetric Engineering & Remote Sensing, 53(12): 1649-1658.*

Fung, T. and E. LeDrew, 1987b. Land Cover Change Detection with Landsat-MSS and TM Data in the Kitchener - Waterloo Area, Canada. Technical Papers of the ASPRS-ACSM Annual Convention. *American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping*, Falls Church, Virginia, Vol. 6, pp. 81-89.

Gong, P., E. F. LeDrew, and J. R. Miller, 1992, "Registration-noise Reduction in Difference Images for Change Detection," *Photogrammetric Engineering & Remote Sensing, 13* (4):773-779.

Green, K., D. Kempka, and L. Lackey, 1994, "Using Remote Sensing to Detect and Monitor Land-cover and Land-use Change," *Photogrammetric Engineering & Remote Sensing, 60* (3):33 *1-337.*

Griffits, G. H. 1988. Monitoring Urban Change from Landsat-TM and SPOT Satellite Imagery by Image Differencing. *Proceedings of the International Geoscience and Remote Sensing Symposium.* European Space Agency, Paris, France. Vol. 1, pp. 493-497.

Hall, F. -G., D. -E. Strebel, J. -E. Nickson, and S. -J. Goetz. 1991. Radiometric Rectification: Toward a Common Radiometric Response Among Multidate, Multisensor Images. *Remote Sensing of Environ., 35* (l): 11-27.

Jensen, J. R. and D. L. Toll 1982. "Detecting Residential Land Use Development at the Urban Fringe," *Photogrammetric Engineering & Remote Sensing, 48(4):629-643.*

Jensen, J. R. 1996. Introductory Digital Image Processing: A Remote Sensing Perspective, Chapter 9, Second Edition, Prentice Hall, New Jersey, pp. 337.

Lambin, E. F. and A. H. Strahler. 1994a. Change-vector Analysis: A Tool to Detect and Categorize Land-cover Change Processes Using High Temporal-resolution Satellite Data. *Remote Sensing of Environment*, *48:231-244.*

Lambin, E. F. and A. H. Strahler. 1994b. Indicators of Land-cover Change for Change-vector Analysis in Multitemporal Space at Coarse Spatial Scales. *International Journal of Remote Sensing, 15:2099-2119.*

Martin, L. R. G. and P. J. Howarth. 1989. Change-Detection Accuracy Assessment Using SPOT Multispectral Imagery of the Rural-Urban Fringe. *Remote Sensing of Environment,* 30:55-66.

Martin, L. R. G. 1986. Change Detection in the Urban Fringe Employing Landsat Satellite Imagery. Plan Can., 26(7): 182-190.

Novak, K., 1992, "Rectification of Digital Imagery," *Photogrammetric Engineering & Remote Sensing, 58 (3):339-344.*

Price, K. P., D. A. Pyke, and L. Mendes, 1992, "Shrub Dieback in a Semiarid Ecosystem: The Integration of Remote Sensing and GIS for Detecting Vegetation Change," *Photogrammetric Engineering & Remote Sensing, 58 (4) 455-463.*

Singh, A. 1989. Review article: Digital Change Detection Techniques using Remotely Sensed Data. *International Journal of Remote Sensing,* 10(6):989-1003.

Westmoreland, S. and D. A. Stow, 1992, "Category Identification of Changed Land-use Polygons in an Integrated Image Processing/Geographic Information System," *Photogrammetric Engineering & Remote Sensing, 58* (11):1593-1599.

Correlation Between Precipitation and Crop Yield for Corn and Cotton Produced in Alabama

Carol E. Hayes, Auburn University, Auburn, AL
Donald J. Perkey, Institute for Global Change Research and Education[1] Huntsville, AL

1.0 Introduction

The total production of corn in Alabama has increased during the past 10 years (Alabama Agricultural Statistical Service 1995). As a result, the economic influence of the corn industry in Alabama has increased. For this reason, it is important to be able to identify and possibly predict the effects of factors influencing corn yield. One such factor is the amount of rainfall. Water stress during silking and tasseling effects the yield more directly than stress during any other time in the life of the plant. The three weeks following this period are the grain-filling period and are also crucial. During this time, the corn kernels increase in weight and size. The silking and tasseling time depends on the planting date, temperature, soil type and other factors. Generally, silking and grain filling occur during the ninth through twelfth weeks after emergence (Wright et al.). In Alabama, corn is usually planted in March, resulting in an early June silking date. The exact planting date varies with regional weather and is typically earlier in southern Alabama and later in northern Alabama (Henderson et al.). This results in a silking date of early June in southern Alabama, and as late as early July in northern Alabama.

Cotton is another important Alabama crop. Alabama ranked eleventh in the nation in cotton production in 1995. Again, water is an important factor influencing yield. Cotton plants are most susceptible to water stress during the period of bloom. Blooms first appear three to four weeks after emergence and peak bloom occurs three to four weeks later. Cotton bloom typically begins July 5-15 (Guthrie et al. 1994). Following peak bloom, the number of flowers on the plants decreases gradually (Monks et al. 1996). Temperature is another important factor in determining yield. If the temperature is excessively hot, the plant cools itself through transpiration to maintain an optimum temperature. This process requires moisture. If high temperatures coincide with low moisture, the plant wilts, reducing yield. Also, in times of drought, stomata on the leaves close. This hampers the process of photosynthesis through which the plant produces the energy necessary for growth. In order to survive, the plant is forced to consume stored carbohydrates that would normally be used for boll filling and new growth, thus reducing crop yield.

In this study, variations in precipitation during the time of corn silking are compared to Alabama corn yields. Also, this study compares precipitation variations during bloom to Alabama cotton yield. The goal is to obtain mathematical correlations between rainfall during the crop's critical period and the crop amount harvested per acre.

2.0 The Data

Two data set types, one containing precipitation data and one, crop yield data, are needed for this study. Each data set was processed independently. Corn yield data were obtained from the Alabama Extension Office at Auburn University and the Alabama Agricultural Statistics Office. The information is printed in the Alabama Agricultural Statistics Annual Bulletin for each year. Data were obtained by county from the earliest data available (1939) through 1996. Data for cotton yield by county were available as early as 1948, but were missing for the years of 1962 and 1963. Data were also missing for various counties in different years.

The counties selected for the study represent the main producers of each crop in the state. High production counties were selected because data were consistently available, and it was hypothesized that a county average would be less affected by one farm with significantly different yields in a county with many farms, than in a county with only a few farms. The state was divided into three regions – south, central and north – in order to determine if the effects of precipitation varied from south to north. Five to six counties were selected in each region (Table 1).

Precipitation data were obtained from National Climatic Data Center (NCDC). These data were originally collected by the National Weather Service and given to NCDC to perform quality control. These data were selected because they were available for various cities across the state of Alabama and because they were quality controlled.

[1] The Institute for Global Change Research and Education is jointly operated by the University of Alabama in Huntsville and Universities Space Research Association.

Crop	Northern Counties	Central Counties	Southern Counties
Corn	DeKalb (3)	Calhoun	Baldwin (4)
	Jackson (1)	Dallas	Crenshaw (10)
	Lauderdale (6)	Elmore	Escambia (9)
	Limestone (5)	Marengo	Geneva
	Madison (2)	Talladega	Houston
	Marshall (7)		Pike (8)
Cotton	Colbert	Autauga	Baldwin (8)
	Lauderdale (7)	Dallas	Escambia (6)
	Lawrence (9)	Elmore	Geneva (4)
	Limestone (1)	Shelby	Henry (10)
	Madison (3)	Tuscaloosa	Houston (2)
			Monroe (5)

Table 1: Counties included in study for corn and cotton. Numbers represent rank for top 10 corn-or cotton-producing counties in Alabama.

Rainfall data were in the form of total precipitation per month recorded at a site. The precipitation values for each site in a given county were averaged to obtain a number representing the whole county average rainfall for the month. This county-average value was used to represent the precipitation amount for any location within the county. Because of the local nature of summer precipitation, this assumption may lead to some scatter in the statistics.

3.0 Methodology

In order to compare the precipitation data to the crop yield data, each data set was expressed as a deviation from the normal or expected value. Deviations were calculated independently for precipitation and crop yield.

3.1 Precipitation Data

The mean monthly precipitation and its standard deviation was calculated for each county. After removing outlier data points that were defined as more than three standard deviations away from the mean, the mean for each month was recalculated. Counties within each region were also compared to find outliers. The regional mean precipitation for each month was calculated by averaging the data from the counties within each region. The standard deviation was found. Again, any points in excess of three standard deviations from the mean were omitted. Only one data point was removed by this process. For this reason, the average for each county was not recalculated. For each year, the deviation was calculated by subtracting the mean value from the actual rainfall recorded.

3.2 Crop Yield Data

Crop yield data were also processed individually by county. Figure 1 shows the plot of crop yield versus year for corn in DeKalb county. This example shows considerable year-to-year variation and indicates that overall yields have increased with time. This increase is attributed to various technological advances, such as improved fertilizers, insecticides, crop rotation practices, etc. For the purposes of this study the variations from this increasing trend are hypothesized to be the result of fluctuations in precipitation. A least-squares mathematical equation was calculated to represent the trend. The "goodness of fit" was determined by the r-squared value. R-squared represents the percent of variance explained by the regression line. An r-squared value of 1 indicates that the line fits the data perfectly. A perfect fit was not expected because of the variation in precipitation as mentioned above. For corn, the best fit was found using an exponential curve. For cotton, a linear trend was selected. The trend was calculated for each county.

The deviation from the trend or "expected" value was then calculated for each year by subtracting the expected yield, based on the trend, from the actual yield. Mean deviation from the trend was calculated by county, as was the standard deviation of the deviation data. Any data point more than two standard deviations above or below the mean was deleted from the set. Comparisons were also made among the counties in each region. Using the five or six counties studied per region, the mean and standard deviation in yield for each year were calculated. Again, data points differing from the mean by more than two standard deviations were removed. This process revealed that much of the data from Baldwin county for corn yield was significantly different from that of other counties in the southern

region. Baldwin was therefore omitted from the corn data set. No counties were found to be significant outliers in cotton data. After removing outliers, anew trend was found and the trend deviation recalculated.

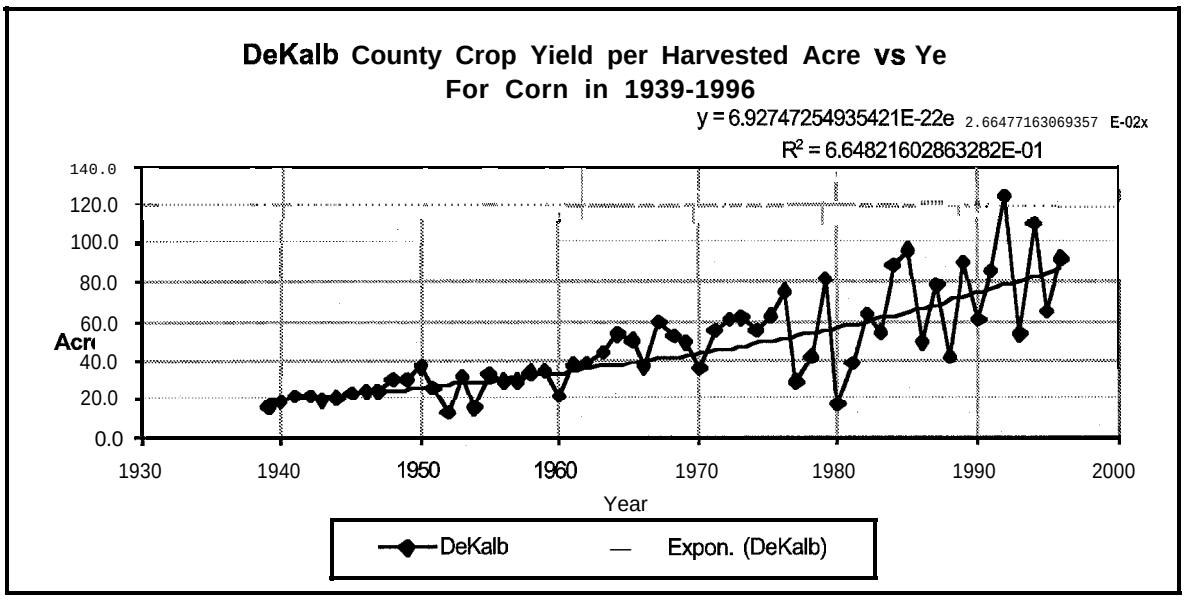


Figure 1: DeKalb county crop yield per harvest acre versus year for corn from 1939 to 1996.

3.3 Precipitation and Crop Yield

With the data in the form of deviations from the average or expected value, rainfall and crop yield data could be compared. It was now necessary to statistically determine the month during which precipitation should be compared to yield. Based on the physiology of the plants, corn was expected to demonstrate an extreme sensitivity to rainfall anomalies during silking and grain filling which varies throughout the state. Statistically the critical month for each county was determined using the correlation coefficient as guidance. The spring or summer month that demonstrated the best correlation between rainfall deviation and yield deviation was selected for comparison. In general, as expected July was the month selected, with June the month selected for some southern counties. Using the same procedure, it was expected that the correlations would be greatest during the month in which cotton plants were in bloom. According to the correlation coefficients, the critical month for cotton was found to occur in July in most counties. Thus, for both corn and cotton the statistical tests agreed with those expected from plant physiology.

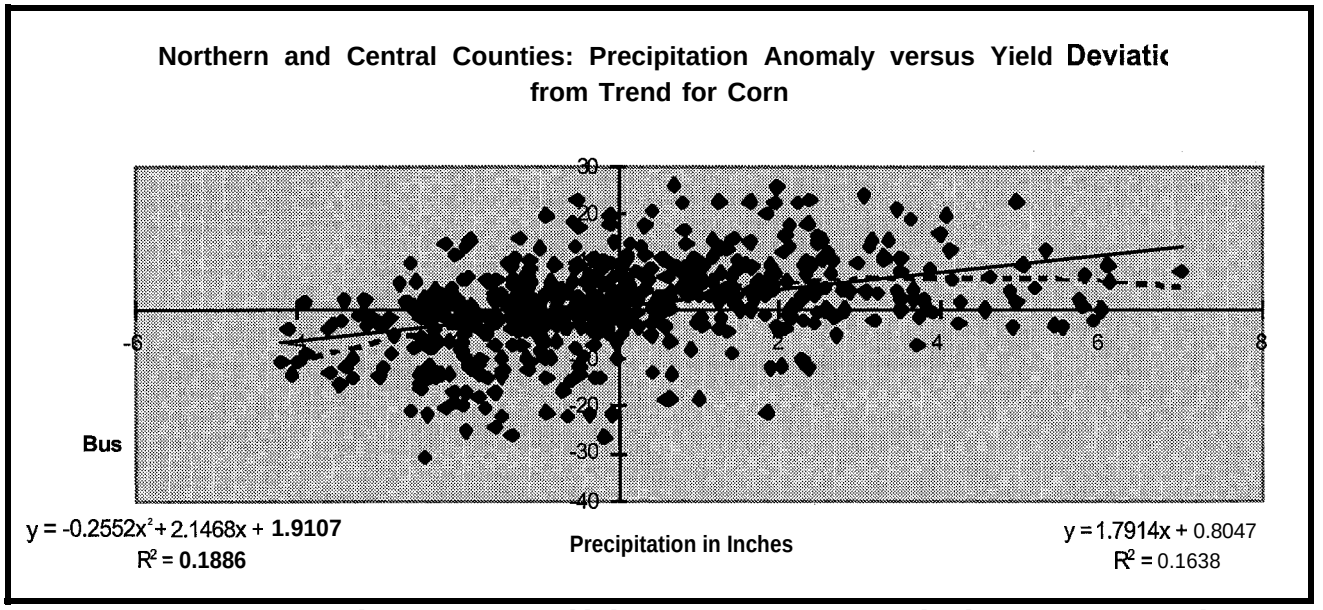
Once the data sets had been statistically tested individually and the critical month determined, it was necessary to determine whether data should be grouped by single counties, region or for the entire state. To provide guidance, t- and f-tests and correlation coefficient calculations were performed. The t- and f-tests yield the probability that the data sets represent data from different populations. The t-test evaluates the mean, while the f-test compares the variance. In each case, a value of 0.05 or lower was sought. This can be interpreted as showing a 95% probability that the data sets were from different populations and should not be grouped. The correlation coefficient yields the percent to which the two sets vary together. A value of 0.75 or higher was interpreted as showing that the two sets were correlated and should be grouped. The data sets were compared for each pair of counties. Precipitation and yield data were evaluated independently. Precipitation data were compared for the growing season months of May, June, July, and August individually. Also yield deviation data were compared for each crop.

Results indicated that there were no significant differences in corn yield between counties, thus the entire state should be grouped together. Although in some cases correlations existed between neighboring counties, this pattern was not consistent and could not be used to divide the data into regions. No significant differences were found between counties for cotton as well. Again, some strong correlations were found, within the northern region, but the correlations were not supported by the f- and t-tests. Therefore, cotton yield data did not support that the data should be divided into regions for comparison.

Precipitation data yielded different results for different months. For example, the data for May showed that the counties within the north, central and south regions demonstrated correlations with each other, and significantly differed from counties from other regions. In contrast, rainfall data for June indicated differences between southern and

central counties, and southern and northern counties. There were no significant difference between northern and central counties. Therefore, calculations involving rainfall data from June were divided into two regions, one including southern counties, and one including northern and central counties. July and August rainfall data exhibited the same pattern as June, with the exception that Baldwin and Escambia counties were shown by the t- and f-tests to be significantly different from other counties, in all regions of the state.

Figure 2 shows the plot of precipitation anomaly versus yield deviation for corn in the northern/central region. For each region, two trend equations were calculated, one linear, and one second-order polynomial. The significance of the fit was calculated by performing the f-test on the r-squared value. Both plots for corn and cotton expressed significant fits for the linear and second-order trend lines. The polynomial equations were selected because they demonstrated a higher r-squared values and thus, higher significance levels.



igure 2: Precipitation anomaly versus crop yield deviation from the expected value for corn m northern and central Alabama. Outliers beyond three standard deviations for precipitation and two standard deviations for yield have been omitted.

4.0 Results and Discussion

4.1 Corn

The results for corn are summarized in Table 2. Based on r-squared values, the second-order polynomial curve provided a better fit than the linear fit. The slope of the curve was evaluated at various levels of rainfall. This slope represents the approximate increase in yield that would result from a 1 inch increase in rainfall during the critical month for the crop. Table 3 shows these results. In both regions of the state, rainfall has a much greater effect on yield during a time with below average rainfall than in a time with average or more. In northern and central Alabama, during a period with rainfall 2 inches below normal, one additional inch of rain would cause an increase in yield of 3.16 bushels per acre. Based on the average yield per acre for corn in northern and central counties in the 1990's, this represents a 4% increase in crop yield. For the same counties, in a period with 2 inches above average rainfall, an additional inch of rain only increases yield by 1.12 bushels per acre, a 1.4% increase in yield.

A similar pattern is observed in the southern counties. In southern Alabama, during a period with rainfall 2 inches below normal, a 1 inch increase in rainfall results in a 2.57 bushel per acre, or 3 .29%, increase in yield. In a period with precipitation 2 inches above normal, a 1 inch increase in rainfall results in only a 0.19 bushel per acre, or 0.2 5%, increase in yield. Table 3 also gives farther information regarding the change in yield as rainfall changes. For both regions, as rainfall increases, the slope decreases. A positive slope depicts that more rain is beneficial to yield. A negative slope means the opposite, increased precipitation decreases yield. Optimum yield occurs when the slope equals zero. According to the above equations and slopes, optimum critical month rainfall for northern counties is 4 to 5 inches above the average rainfall, or around 9 inches of rain during the month of silking and tasseling. For

southern Alabama, optimum rainfall is around 2 to 3 inches above the average, or 8 inches of rain. The variation in optimum rainfall may be due to differences in soils, or other factors not considered in this study.

	Second-order Polynomial Fit	r-squared	Linear Fit	r-squared
Northern and Central Counties	$y = -0.2552x^2 + 2.1468x + 1.9107$	0.1886	$y = 1.7914x + 0.8047$	0.1638
Southern Counties	$y = -0.2974x^2 + 1.3819x + 1.6955$	0.1549	$y = 0.8287x + 0.3074$	0.0787

Table 2: Trend equations for corn yield per acre in regions of Alabama.

	Precipitation Anomaly (Inches)										
	-10	-5	-2	-1	0	+1	+2	+3	+4	+5	+10
Northern and Central Counties (Yield: 77.9)											
Slope	7.251	4.699	3.168	2.657	2.147	1.636	1.126	0.617.	0.105	-0.405	-2.957
% Yield Change	9.31	6.03	4.07	3.41	2.76	2.10	1.45	0.79	0.14	-0.52	-3.80
Southern Counties (Yield: 78.2)											
Slope	7.330	4.356	2.572	1.977	1.382	0.787	0.192	-0.403	-0.997	-1.592	-4.566
% Yield Change	9.38	5.57	3.29	2.53	1.77	0.01	0.25	-0.51	-1.28	-2.04	-5.84

Table 3: Slope and percent deviation in corn crop yield due to precipitation variations by region. Yield refers to the average yield per acre harvested in 1990 to 1996.

4.2 Cotton

Analysis of the crop yield versus precipitation for cotton showed more scatter than for corn. Correlation coefficients for cotton in July were typically smaller than those for corn. However, the same procedure was followed to interpret the data. Trend equations were plotted for precipitation versus crop yield deviations, and r-squared values were calculated. Table 4 displays the trend equations and corresponding r-squared values. The slope for each region was calculated, using the second-order polynomial trend lines as shown in Table 5. These were interpreted for cotton as they were for corn, and were found to follow the same pattern. In northern and central counties, a 1 inch increase in critical month precipitation during a period with 2 inches below average rainfall, would result in a yield increase of 23.1 pounds per acre. Based on the average yield per acre for the 1990's, this represents a 3.6% increase. In southern counties, a 1 inch increase in critical month precipitation during a period with 2 inches below average rainfall would result in a 14.8 pound per acre, or 2.4%, increase in yield. During a period with 2 inches above average rainfall, an increase of an inch in rainfall would result in a 10.7 or 4.5 pound per acre increase in yield in northern or southern Alabama, respectively. As with corn, the effect of increased precipitation during the critical month is greater in a time with less rain than in a time with more rain. Also, the optimum amount of rainfall is shown to be above the average rainfall for the region. In north and central Alabama, optimum rainfall appears to be 5 to 6 inches above average, or about 10 inches of rain. In the southern counties, optimum rainfall is 3 to 4 inches above average, also about 9 to 10 inches total.

	Second-order Polynomial Fit	r-squared	Linear Fit	r-squared
Northern and Central Counties	$y = -1.5484x^2 + 16.955x + 19.893$	0.0843	$y = 14.631x + 13.098$	0.0751
Southern Counties	$y = -1.2899x^2 + 9.6688x + 19.844$	0.0621	$y = 9.2362x + 12.703$	0.0545

Table 4: Trend equations for cotton yield per acre in regions of Alabama.

5.0 Conclusions

The results of this study show that there is a statistically significant correlation between precipitation and crop yield for corn and cotton grown in Alabama. Increased precipitation has the greatest effect on corn crop yield during the month when the plant is silking and tasseling. This was found to occur in June or July, depending on the region of the state. Cotton crop yield was found to be most influenced by precipitation during the bloom stage, which was found to occur most often in July, but varied with region. Increased precipitation was also found to have a greater effect on crops during dry intervals than intervals with above average rainfall. While these trends were observed, there was some uncertainty, caused by low correlation coefficients and r-squared values. The f-test performed on the

r-squared values indicate that the relationships were significant, that is, the correlations between precipitation and yield were not the result of chance. However, low r-squared values suggest that the relationships explain less than 20% of the variance in the data. For such equations to be very beneficial in predicting actual outcomes of increased precipitation or irrigation, explanation of a larger percentage of the variance would be desired.

	Precipitation Anomaly (Inches)										
	-10	-5	-2	-1	0	+1	+2	+3	+4	+5	+10
Northern and Central Counties (Yield: 642)											
Slope	47.923	32.439	23.149	20.052	16.955	13.858	10.761	7.665	4.568	1.471	-14.013
% Yield Change.	7.46	5.05	3.60	3.12	2.64	2.16	1.68	1.19	0.71	0.23	-2.18
Southern Counties (Yield: 623)											
Slope	35.467	22.568	14.828	12.249	9.669	7.089	4.509	1.929	-0.650	-3.230	-16.129
% Yield Change	5.69	3.62	2.38	1.97	1.55	1.14	0.72	0.31	-0.10	-0.52	-2.59

Table 5: Slope and percent deviation in cotton crop yield due to precipitation variations by region. Yield refers to the average yield per acre harvested in 1990 to 1996.

The weak correlation may be caused in part by the nature of the water needs of the plants. Precipitation data used in this study did not indicate the intensity of the rainfall. Cotton thrives in months with frequent showers followed by hot sun. The monthly averages used express a month with sporadic heavy rain storms as being equivalent to a month with frequent showers. Such differences would be expected to have a large impact on yield, but could not be detected from the data used in this study. An analysis of the intensity and distribution of rain throughout the month might yield better correlations than those obtained considering only monthly totals of rainfall.

This study also failed to consider the effects of factors other than precipitation on crop yield. Factors such as soil type, temperature, humidity, disease and insect infestation among others, must also be considered. In order to make a conclusive statement about the effect of precipitation alone, further research should be conducted accounting for the effects of these additional factors. Further research should also examine the effects of weather patterns such as El Niño-Southern Oscillation, which effect precipitation and temperatures in Alabama.

6.0 Acknowledgments

Part of this study was conducted during an internship by Carol Hayes at the Global Hydrology and Climate Center through the Center for Hydrology, Soil, Climatology and Remote Sensing (HSCaRS) at Alabama A&M University, funded by NASA (Cooperative Agreement No. NCC8- 140). Other funding for this study was provided by the Alabama State Climatology Office located at the University of Alabama in Huntsville (UAH). Thanks are extended to Mr. Max Runge and others with the Alabama Cooperative Extension Service for their assistance in providing and interpreting crop yield data, to Mr. Bob Clymer of UAH for his assistance with rainfall data, to Dr. Charles Laymen of IGCRE and HSCaRS for his assistance in determining the statistical significance of the findings.

References

Alabama Agricultural Statistical Service, 1995: Corn Acreage, Yield and Production. *Alabama Agricultural Statistics - Annual Bulletin.* 92p.

Guthrie, D., Albers, D., Edmisten, K., Hake, K., and Kerby, T., 1994:1993 Crop Yields — Explaining the Highs and Lows, *Cotton Physiology Today. 5,* January, 4p.

Henderson, J., Ball, D.M., and Mask, P.L.: Soybean Rotation Options for Alabama Farmers, *Agriculture and Natural Resources: Agronomy, Alabama Cooperative Extension Service, Auburn University Alabama.* Circular ANR-460, 4p.

Monks, C.D., Burmester, C.H., Patterson, M. G., and Delaney, D., 1996: Monitoring Cotton Growth and Development in Alabama, *Alabama Cooperative Extension System.* Circular ANR-995, 6p.

Wright, D. L., Rhoads, F.M., and Stanley R. L.: Irrigated Corn Production, *Florida Agricultural Experiment Stations, Florida Cooperative Extension Service, University of Florida.* Circular 486, 16p.

AN EXPERIMENT WITH ROBUST FUZZY CONTROLLER USING DELTA-SIGMA MODULATION

Jeremy Payton, Tim Peterson, Djuro Zrilic'
NASA Center for Autonomous Control Engineering
New Mexico Highlands University
Department of Engineering
Fax: (505) 454-3306
E-mail: Payton@edison.nmhu.edu
Las Vegas, NM 87701

ABSTRACT:

An experiment, which combines a fuzzy logic controller and proportional controller with delta-sigma modulation, was performed. To demonstrate the robust performance of the proposed system, we have built an experimental model to control the temperature of a plant chamber. Set point responses were plotted for system response with and without disturbances. It was proven that the combined system of fuzzy and proportional control with delta-sigma modulation has a significant advantage over conventional control in the presence of external disturbances.

Key words: Delta-Sigma Modulation, Proportional Control, Fuzzy Logic.

I. INTRODUCTION

The combined role of fuzzy logic control and conventional control was first examined by He et al [1]. They proposed to connect a conventional PI controller and a fuzzy controller in series. They showed that a fuzzy controller alone is not capable of tracking changes in transient response to sudden load disturbances or variations in the system's parameters. Zrilic' et al [2], proved that the robustness of the control system can be significantly improved by using delta-sigma modulation (DSM), compared to the approach of He's et al, when the disturbance level in the supply line or load line is significant. To verify the simulation results of reference [2], we have built an experimental model according to Figure 1. The significance of this approach is that delta-sigma modulation (DSM) improves the disturbance performance of the control system as is shown in the following section.

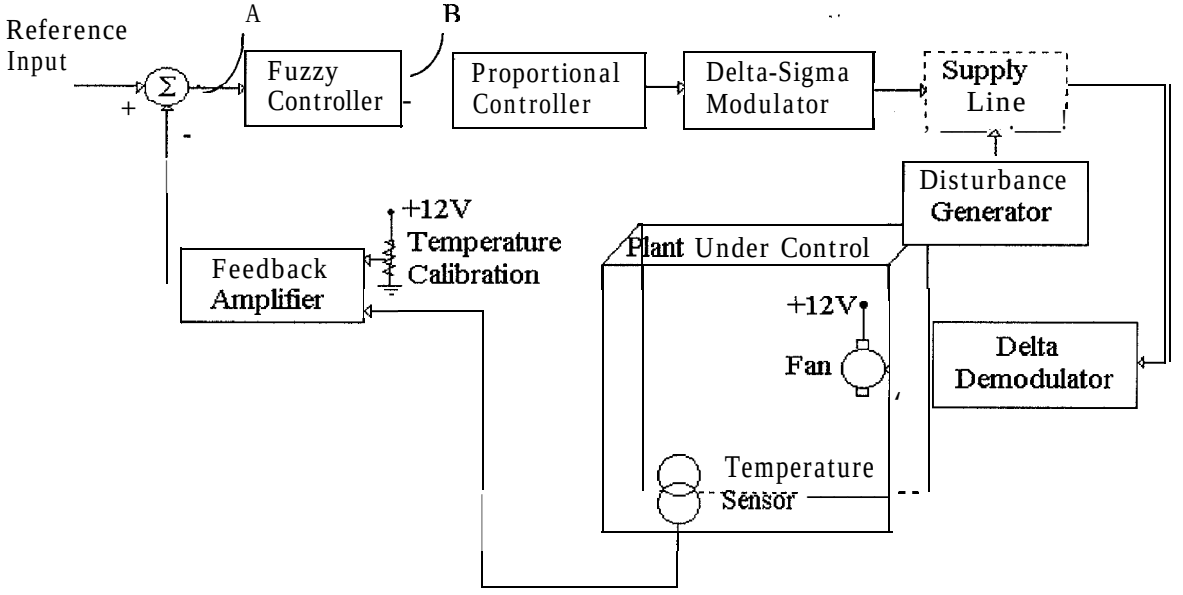


Figure 1: Block Diagram of proposed System.

II. EXPERIMENTAL RESULTS

According to the block diagram in Figure 1, the feed forward part consists of a fuzzy controller, a proportional controller, an analog-to-digital converter (which is in our case a delta-sigma modulator), a supply line with a disturbance generator, and a plant under control (chamber with fan and temperature sensor). The temperature sensor and the feedback amplifier are part of feedback of the system. The delta-sigma demodulator of the proposed scheme resides as an integral part of the plant, which may be at a remote or isolated location. To demonstrate robust performance of the proposed system, we included a disturbance generator to generate errors at rates of 10-1, 10-2, and 10-3. Usually disturbances are random in nature, here, however; the errors are periodic bursts, since this is the worst case in practice. Four different cases are examined as follows:

Case #1: Proportional Controller only

To verify the performance of the proposed system, we have built an experimental test bed. The hardware block diagram of this system is shown in Figure 2. As a bench- mark, we are using proportional control only. This means that all blocks Ii-em Figure 1 are by-passed except for the proportional controller, the supply line with disturbances, and the plant under control.

Figure 3a is a plot of the output voltage from the temperature transducer. Figure 3b is a plot of the response of the proportional controller without disturbances.

Figure 4a is a plot of the output of the temperature transducer where digital error bursts are initiated in the supply line. Figure 4b is a plot of the response of the proportional controller when the burst error rate is 10-1. Clearly, the proportional control system is very sensitive to impulsive noise.

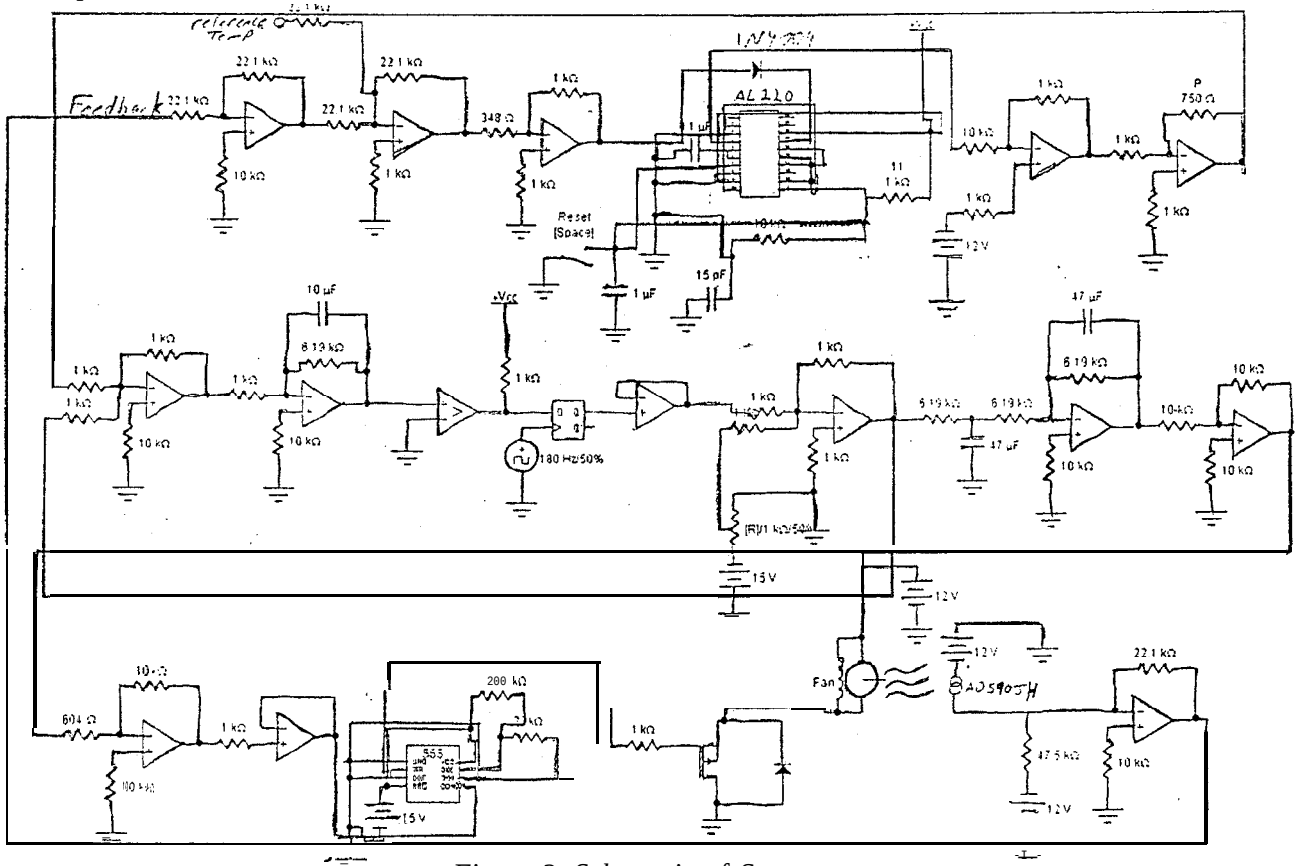


Figure 2: Schematic of System.

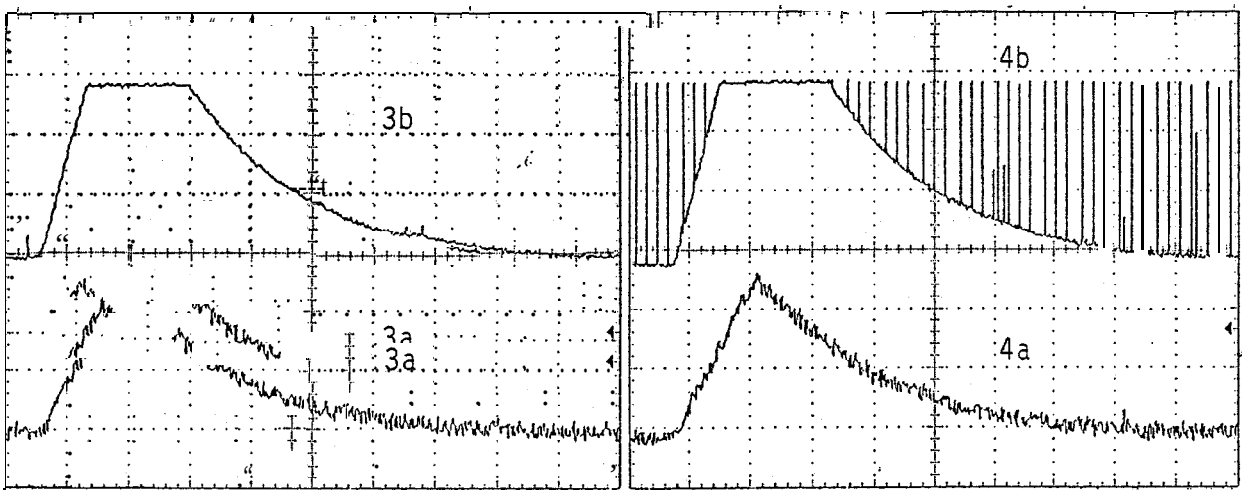


Figure 3a,b: On Left plots Proportional Controller without disturbances and system response.
Figure 4a,b: On right plots Proportional Controller with disturbances and system response.

Case #2: Proportional and Fuzzy Controllers **only**

In this experiment, both fuzzy and proportional controllers were used. Figure 5a is a plot of the output of the temperature transducer without supply line disturbances. It *is* worth mentioning that high frequency noise is coming into the system from fuzzy logic controller oscillator. Figure 5b is a plot of the joint controllers' response just before it is input into the plant under control. Figure 6a is a plot of the transducer output when supply line burst errors are initiated with a rate of 10^{-1}. Figure 6b is a plot of the relevant response. In comparison to Figure 5b, a significant increase of impulsive noise is seen.

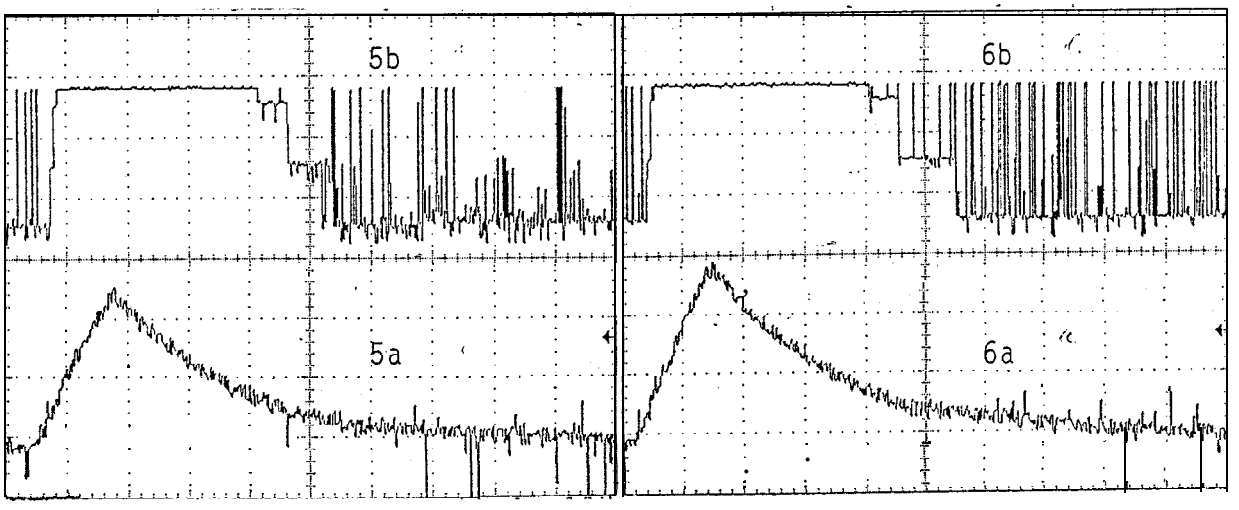


Figure 5a,b: On left plots Proportional and Fuzzy Control without disturbances.
Figure 6a, b: On right plots Proportional and Fuzzy Control with disturbances.

Case #3: Proportional Controller and **DSM** System only

In this experiment, both proportional controller and DSM were used. Figure 7a,b are plots without disturbances when the combination of the proportional controller and DSM system were connected to control a system according to Figure 1.

Figure 8a,b are plots of the Proportional Controller when impulsive disturbances are initiated at a rate of 10-1. We see that impulsive noises are eliminated, but the system response has a certain oscillation.

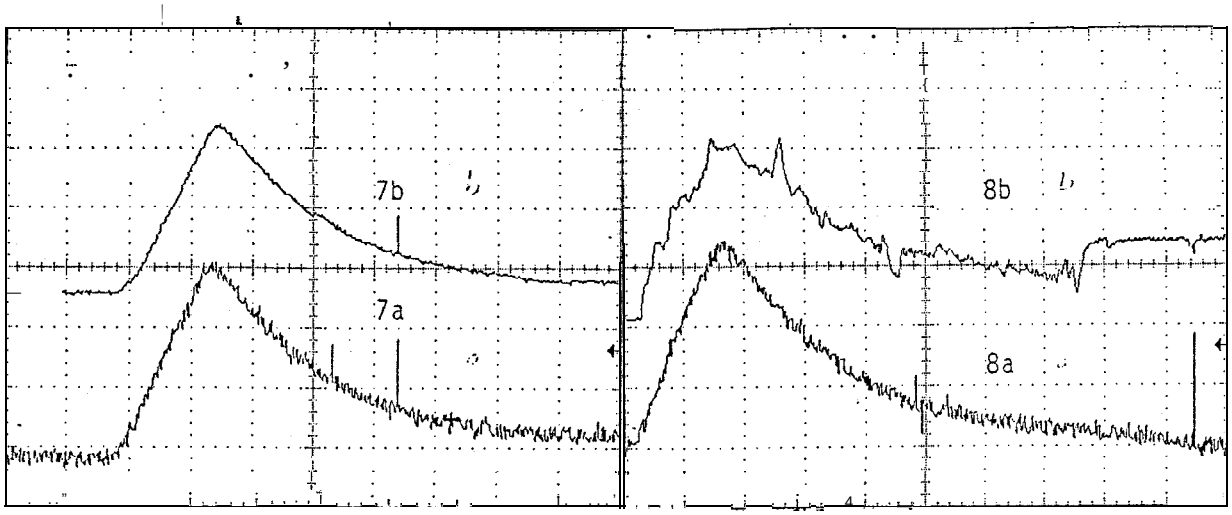


Figure 7a,b: On left plots Proportional Controller with Delta-Sigma Modulation without disturbances.
Figure 8a,b: On right plots Proportional Controller with Delta-Sigma Modulation with disturbances.

Case #4: Fuzzy and Proportional Controllers with **DSM** system.

In this experiment, both fuzzy and proportional controllers with DSM were used. Figure 9b is a plot of the response of the proposed system from Figure 1, without disturbances.

Figure 10b is a plot of the response when disturbance rate of 10-1 was initiated. Both oscillations and impulse noise are greatly reduced using the current proposed approach. Again, it is worth mentioning that high frequency noise came from the oscillator of the fuzzy controller, which operates at nearly 2MHz.

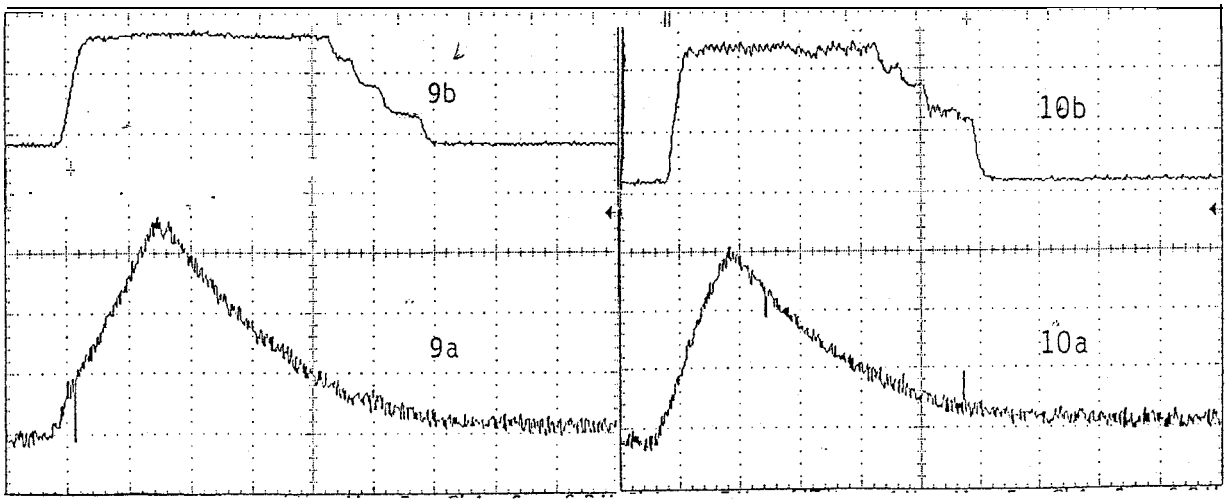


Figure 9a,b: On left plots Fuzzy Control Proportional Control and Delta-Sigma Modulation without disturbances.
Figure 10a,b: On right plots Fuzzy Control Proportional Control and Delta-Sigma modulation with disturbances.

I. CONCLUSION

We have shown that the combination of fuzzy and conventional controllers with delta-sigma modulation can have a significant benefit when supply disturbances have to be suppressed. Although disturbances can enter into any point of the system, our experiment was performed with disturbances entering the supply line only. References [1] and [2] consider the influence of load and input disturbances as well.

APPENDIX A

According to Figure 1 we have listed the Fuzzy Controller rules below with reference to A(Error) and B(response).

If Error is sml_+ then Response = O.
If Error is smd_+ then Response = 50.
If Error is med_+ then Response= 100.
If Error is lmd_+ then Response = 150.
If Error is lrg_+ then Response = 250.

The output of the chip is between 0-5V and it is divided into 256 different levels. The fuzzy rules have a symmetric triangular shape, and error is classified as follows: small(sml), small-medium(smd), medium(med), large-medium(lmd), and large(lrg).

REFERENCES

1. S. Z. He et al, "Design of an On-line Rule-adaptive Fuzzy Control System." Proceeding for IEEE Conference on Fuzzy Systems, pp. 83-91, 1992.
2. D. Zrilic' et al, "A New Robust Fuzzy Controller Using Sigma-Delta Techniques." Proceeding of IFSA '97. Volume 3, pp. 266-272, 1997.

General Solutions for a Carrier Based Phase-Locked Loop Servo Normalized for Type II and Type III Systems

Asad M. Madni[1] and John M. Jumper[2]

[1]BEI Sensors& Systems Company
Corporate Offices
13100 Telfair Avenue
Sylmar, CA 91342
Tel: (818) 364-7215
Fax: (818) 362-1836
Email: bei1madni@aol.com

[2]BEI Sensors& Systems Company
Precision Systems and Space Division
1100 Murphy Drive
Maumelle, AR 72113
Tel: (501) 851-4000
Fax: (501) 851-5476
Email: mjumper@beissd.com

ABSTRACT

The carrier based phase-locked loop system is unique in it's ability to provide accurate rate control over a wide range of velocities including zero. The design of this type of servo system if viewed as a unique application is a complex and formidable design problem. However, there are an infinite number of general solutions to the phase-locked loop design problem. Presented in this paper are normalized general solutions that apply to all cases of "idealized" carrier based phase-locked loop type II and type III systems, referred to as the Ultra-Loc® [1]. A detail hardware design and principle of operation of this system was presented by the authors at the 1997 NASA URC Technical Conference held in February in Albuquerque, New Mexico [2]. This system was the basis of the extremely slow motion servo control technique used in the Hubble Space Telescope's (HST) Star Selector Servo Subsystem (SSSS).

DEVELOPMENT OF A NORMALIZED TYPE II TRANSFER FUNCTION

The process of normalizing transfer functions in filters and closed-loop systems is a powerful technique used to simplify the general solution types. Normal forms of transfer functions or general solutions that could be applied to servo systems provide the means to adapt solutions to systems with large or small motors, high or low resolution feedback sensors, and high or low bandwidth applications. There are an infinite number of general solutions to the type II control design problem. However, by selecting the appropriate frequency as the normalization point, a single manageable transfer function can be developed. Figure 1 shows the "idealized" Ultra-Loc® servo system. If the Integral Gain Factor (I) is set to zero, a type II system results.

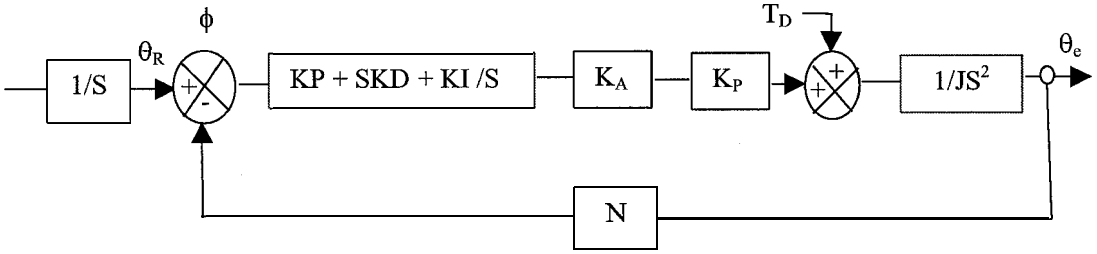


Figure 1. Idealized Ultra-Loc® Servo System

The open loop gain (G) of a type II system is given by:

$$G = \frac{KDS + KP}{S^2} \tag{1}$$

and the closed loop gain of a type II system is given by:

$$\frac{G}{1+G} = \frac{KDS + KP}{S^2 + KDS + KP} \tag{2}$$

The basis for the normalized solution presented here lies in the common ω_N (resonance radian frequency) term found in both open and closed loop gain equations. Also, if the denominator of equation (2) is compared to the quadratic factor $(S^2 + 2\xi\omega_N S + \omega_N^2)$, then the closed loop gain is shown to be:

$$\frac{G}{1+G} = \frac{2\xi\omega_N S + \omega_N^2}{S^2 + 2\xi\omega_N S + \omega_N^2} \tag{3}$$

Similarly the open loop gain is:

$$G = \frac{2\xi\omega_N S + \omega_N^2}{S^2} \tag{4}$$

If a value ξ is chosen that is less than unity, the system will have complex poles and the transient response will have oscillatory components. If the value ξ is greater than unity, the system will be slower to respond than for $\xi = 1$. Thus, a critically damped solution with<= 1 is selected as a basis for a normalized solution.

The system designer now has three variables, K, D and P to select. To determine these values the maximum acceleration requirement must be known or conservatively estimated. This leads directly to KP. The value of KP is found knowing the maximum error signal from the phase comparator (producing maximum system torque) occurs at $\varepsilon = \pi/N$. The end of its linear region where N the number of Sine/Cosine cycles in the feedback sensor. The maximum acceleration is given by two expressions. $\alpha_{max} = (KP)\varepsilon_{max} = KP\,\pi/N$ and $\alpha_{max} = T_{max}/J$.

By combining these two expressions it is seen that: $KP = NT_{max}/\pi J$. (5)

With the product KP defined in terms of system variables, one could select any P and then calculate K. From equation (3) values of KP and KD are:

$$KP = \omega_N^2 , \quad KD = 2\omega_N \quad \text{and} \quad KD = 2\sqrt{KP} \tag{6}$$

Equations (3) and (4) can be reduced to normalized form by using the substitution $S = \omega_N P$. Knowing that:

$$\frac{G}{1+G} = \frac{2\omega_N S + \omega_N^2}{S^2 + 2\omega_N S + \omega_N^2} = \frac{2\omega_N S + \omega_N^2}{(S+\omega_N)^2} \tag{7}$$

this substitution yields a closed loop gain of:

$$\frac{G}{1+G} = \frac{2P+1}{(P+1)^2} \tag{8}$$

and an open loop gain of:

$$G = \frac{2P+1}{P^2} \tag{9}$$

Equations (8) and (9) are the normalized transfer functions for a type II system. Figure 2 shows the normalized gain and phase characteristics of the open-loop function while Figure 3 is the normalized closed-loop.

The closed loop gain as represented in equation (8) can be restated in complex form as:

$$\frac{G}{1+G} = \frac{1 + j2\omega}{(1+j\omega)^2} \tag{lo}$$

The magnitude (A) of this closed loop gain response is:

$$A = \sqrt{\left(\frac{1+4\omega^2}{1+2\omega^2+\omega^4}\right)} \tag{11}$$

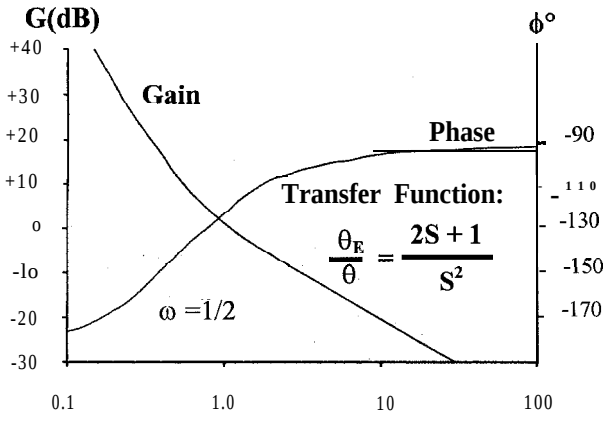


Figure 2. Open-Loop Gain and Phase vs. Radian Frequency

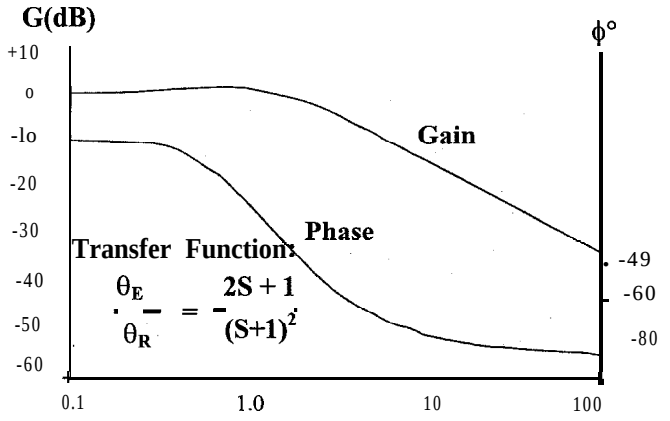


Figure 3. Closed-Loop Gain and Phase vs. Radian Frequency

and the phase of the closed loop response is:

$$\phi = \tan^{-1}(2\omega) - 2\tan^{-1}(\omega) \tag{12}$$

The frequency at which the 3db gain loss occurs is found when the magnitude of (A) becomes $\sqrt{2}/2$. This produces the relationship in equation (11) of:

$$\frac{1}{2} \cdot \frac{1+4\omega^2}{1+2\omega^2+\omega^4} \quad \text{or} \quad \omega^4 - 6\omega^2 - 1 = 0 \tag{13}$$

The only real root of equation (13) is $\omega^2 = 3 + \sqrt{10}$ or $\omega = \sqrt{3+\sqrt{10}} = 2.48$ radians/see. Using equation(n), the polynomial that represents a maximum gain condition is :

$$2\omega^4 + \omega^2 - 1 = 0 \tag{14}$$

From this equation, the positive root $\omega^2 = \frac{1}{2}$ or $\omega = \sqrt{2}/2$ is found. At this frequency, the magnitude of the closed loop gain (A) = $\sqrt{4/3}$ = 1.1547. The open loop gain is given by equation (9). The magnitude (A_{oL}) of the open loop gain is described by equation (15) and the phase (ϕ_{OL}) is described by equation (16):

$$A_{OL} = \frac{\sqrt{1+4\omega^2}}{(0^2} \tag{15}$$

$$\phi_{OL} = \tan^{-1}(2\omega) \cdot \pi \tag{16}$$

DEVELOPMENT OF A NORMALIZED TYPE III TRANSFER FUNCTION

As previously stated, there are an infinite number of general solutions to the Type III control design problem. However, by selecting the appropriate frequency as the normalization point, a single manageable transfer function can be developed. The open loop gain for such a Type III system is described by:

$$G = \frac{KDS^2 + KPS + KI}{S^3} \tag{17}$$

The closed-loop gain is given by:

$$\frac{G}{1+G} = \frac{KDS^2 + KPS + KI}{S^3 + KDS^2 + KPS + KI} \tag{18}$$

The basis for the normalized form presented herein lies in the denominator of equation (18). This cubic function can have either three real roots or one real root and a complex conjugate pair of roots. If system parameters are chosen that produce complex roots, then the transient response of the system will have oscillatory (sinusoidal) terms. Certainly this condition needs to be avoided and the three real roots selected. In most cases it is desirable that all three roots be large so that the system time constant is small. Closer examination of the denominator of equation (18) yields the product.

$$(S+A)(S+B)(S+C) = S^3 + (A+B+C)S^2 + (AC+AB+BC)S + ABC \quad \text{where} \quad ABC = KI \tag{19}$$

The maximum value of the product of the three roots occurs when all roots are equal. That is to say if any root is larger than the remaining two, they must be smaller. Therefore, the largest minimum root occurs when all roots are equal. This condition results in the smallest time constant. From these arguments, the normalized radian frequency ω_x is selected such that:

$$(S+\omega_x)^3 = S^3 + 3\omega_x S^2 + 3\omega_x^2 S + \omega_x^3 \tag{20}$$

By comparing equation (20) with the denominator of (18) it is obvious that:

$$KI = \omega_x^3 \quad , \quad KP = 3\omega_x^2 \quad \text{and} \quad KD = 3\omega_x \tag{21}$$

As in the case of the type II system, in order to determine the appropriate choices for K, P, D and I, the acceleration requirements must be known or conservatively estimated. This knowledge leads directly to the product KP. Knowing the maximum feedback phase comparator error is $(\epsilon = \pi/N)$, where N is the encoder cycles of sine/cosine and is proportional to the maximum error signal, under static conditions the maximum acceleration is:

$$a_{max} = KP\epsilon_{max} = KP\,\frac{\pi}{N} \tag{22}$$

However, acceleration is also given by:

$$a_{max} = \frac{T_{max}}{J} \quad \text{where} \quad T_{max} = \text{peak torque and } J = \text{total inertia} \tag{23}$$

Therefore, the product KP can be seen as: $KP = \frac{N\,T_{max}}{\pi J}$ (24)

and KI and KD can now be expressed in terms of KP:

$$KI = \left(\frac{KP}{3}\right)^{312} \quad \text{and} \quad KD = \sqrt{3KP} \tag{25}$$

The open-loop gain can now be restated using (21) as: $G = \frac{3\omega_x S^2 + 3\omega_x^2 S + \omega_x^3}{S^3}$ (26)

and the closed-loop gain as: $\frac{G}{1+G} = \frac{3\omega_x S^2 + 3\omega_x^2 S + \omega_x^3}{(s+\omega_x)^3}$ (27)

Equations (26) and (27) are converted to a normalized form by the substitution: $S = \omega_x P$

The normalized open-loop gain is: $G = \frac{3P^2 + 3P + 1}{P^3}$ (28)

and the normalized closed-loop gain is: $\frac{G}{1+G} = \frac{3P^2 + 3P + 1}{(P+1)^3}$ (29)

Equations (28) and (29) are the normalized transfer functions for a Type III system. Figure 4 illustrates the normalized gain and phase characteristics of the open-loop function, while Figure 5 shows the normalized closed-loop functions. These figures are derived directly from equations (28) and (29). The quadratic solution applied to equation (29) shows three poles at $\omega = 1$, and complex zeros at:

$$\omega = \frac{-3 \pm j\sqrt{3}}{6} = \frac{1}{\sqrt{3}} \angle \pm 30°$$

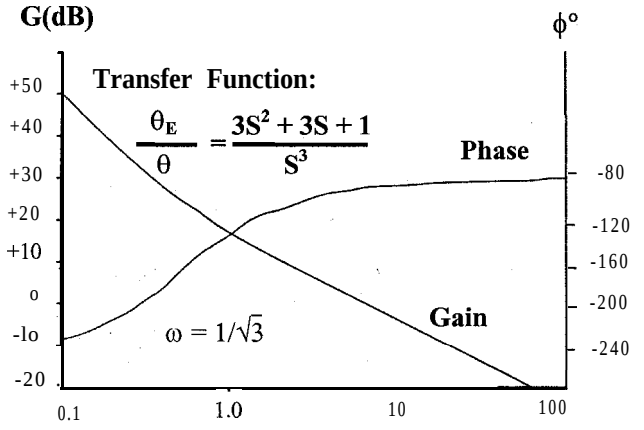


Figure 4. Open-Loop Gain and Phase vs. Radian Frequency

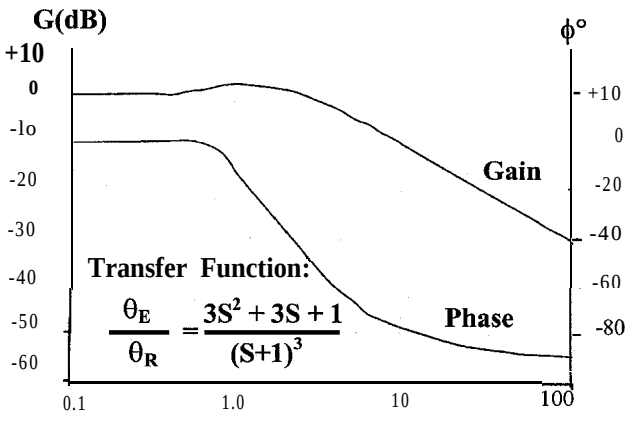


Figure 5. Closed-Loop Gain and Phase vs. Radian Frequency

The complex frequency form of the closed-loop gain equation (29) is given by:

$$\frac{G}{1+G} = \frac{(1-303^2) + j3\omega}{(1-303^2) + j(3\omega - 63^3)} \qquad (30)$$

where the magnitude (A) is:

$$A = \sqrt{\frac{(1-3\omega^2)^2 + 9\omega^2}{(1-3\omega^2)^2 + (3\omega - \omega^3)^2}} \qquad (31)$$

and the phase (ϕ) is:

$$\phi = \tan^{-1}\left(\frac{3\omega}{1-3(0^2}\right) - 3\tan^{-1}(0)) \qquad (32)$$

The 3 db frequency is found when the magnitude (A) of the gain function becomes $\sqrt{2}/2$ which yields:

$$\omega^6 - 15\omega^4 - \omega^2 - 1 = 0 \qquad (33)$$

Since equation (33) is similar to $x^3 - 15x^2 - x - 1 = O$ which has only one real root at $x = 15.07$ and $\omega = \sqrt{x} = 3.882$ radians/see. Similarly, the maximum closed-loop gain is found from equation (33).

$$3\omega^6 + 2\omega^4 - 5\omega^2 - 4 = 0 \qquad (34)$$

The positive root here is $\omega^2 = 4/3$ or: $\omega = \sqrt{4/3} = 1.1547$. At this frequency the closed-loop gain (A) is 1.28714 or 2.1829 db. From the open-loop gain function of the normalized Type III system equation (28) the expression for amplitude (A_{OL}) of the open-loop gain function is:

$$A_{OL} = \frac{\sqrt{1 + 3\omega^2 + 9(0^4}}{\omega^3} \quad (35)$$

and the phase (ϕ_{OL}) is:

$$\phi_{OL} = \tan^{-1}\left(\frac{3\omega}{1-3(\omega)^2}\right) - \frac{3\pi}{2} \quad (36)$$

Comparison of Type II and Type III Systems

Using the normalized solutions presented above, we are able to determine the system response for most all applications. However, the two types of systems were not normalized to the same frequency. The Type II system was normalized to ω_N where $\omega_N = \sqrt{KP}$. The Type III system was normalized to $\omega_N = (KI)^{1/3}$. In each case, the normalization frequency produced the most responsive system for its type and at this frequency multiple poles occur in the closed loop response. This arbitrary choice resulted in simple forms of transfer functions.

The basic linking factor between the two solutions is the factor KP determined from the maximum acceleration specification, producing $\omega_N = \sqrt{KP}$ as a common linking factor. A comparison of equations (21) and (6) for KP indicates the basic connecting relationship:

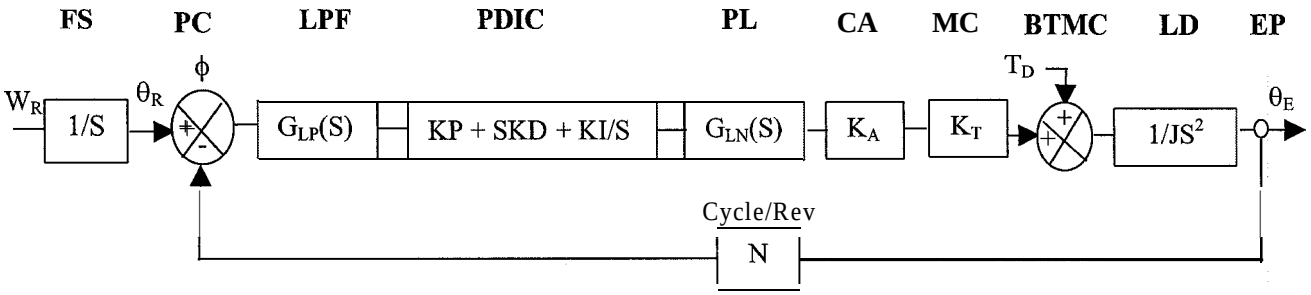
$$\omega_X = \omega_N/\sqrt{3} \quad (37)$$

The substitution of equation (37) into any of the frequency results of the Type III analysis will produce relationships that can be directly compared to the Type II results.

Design Example: Hubble Space Telescope's Star Selector Servo Subsystem

The Star Selector Servo Subsystem is presented as a Type III design example. This two axis servo system positions interferometric optics within the Fine Guidance Sensors in the Hubble Space Telescope to locate and track guide stars. The system provides the fundamental pointing reference for the Telescope. The system has 21-bits (0.61 8 arcsecond) position resolution and a velocity range from 0.5 arcsecond per second to 16,384 arcseconds per second.

The previously developed normalized equations represent a system optimized for the best time constant over both small and large angles. The SSSS is optimized for tracking at very slow step rates. Figure 6 represents the idealized SSSS block diagram. Figure 7 shows the actual servo design implemented to accomplish the ultra slow tracking rates stated above. Table 1 shows the selected parameters for the system



FS = Frequency Synthesizer
PC = Phase Comparator
LPF = Low Pass Filter
PDIC = PDI Compensation
PL = Phase Lead
CA= Current Amplifier
MC = Motor Constant
BTMC = Bearing Torque
LD = Load
EP = Encoder Position
Motor Cogging

Figure 6. Idealized Star Selector Servo System

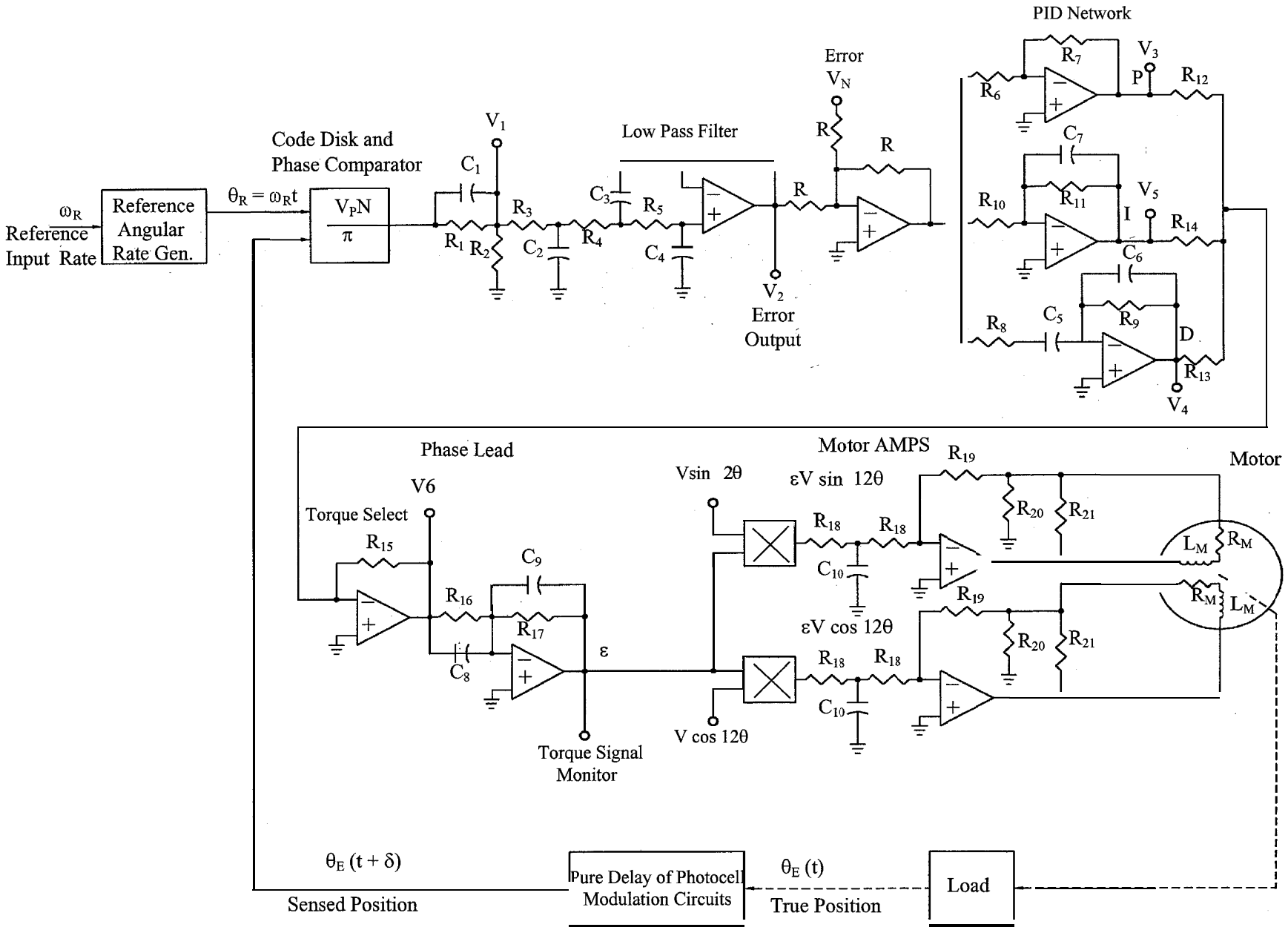
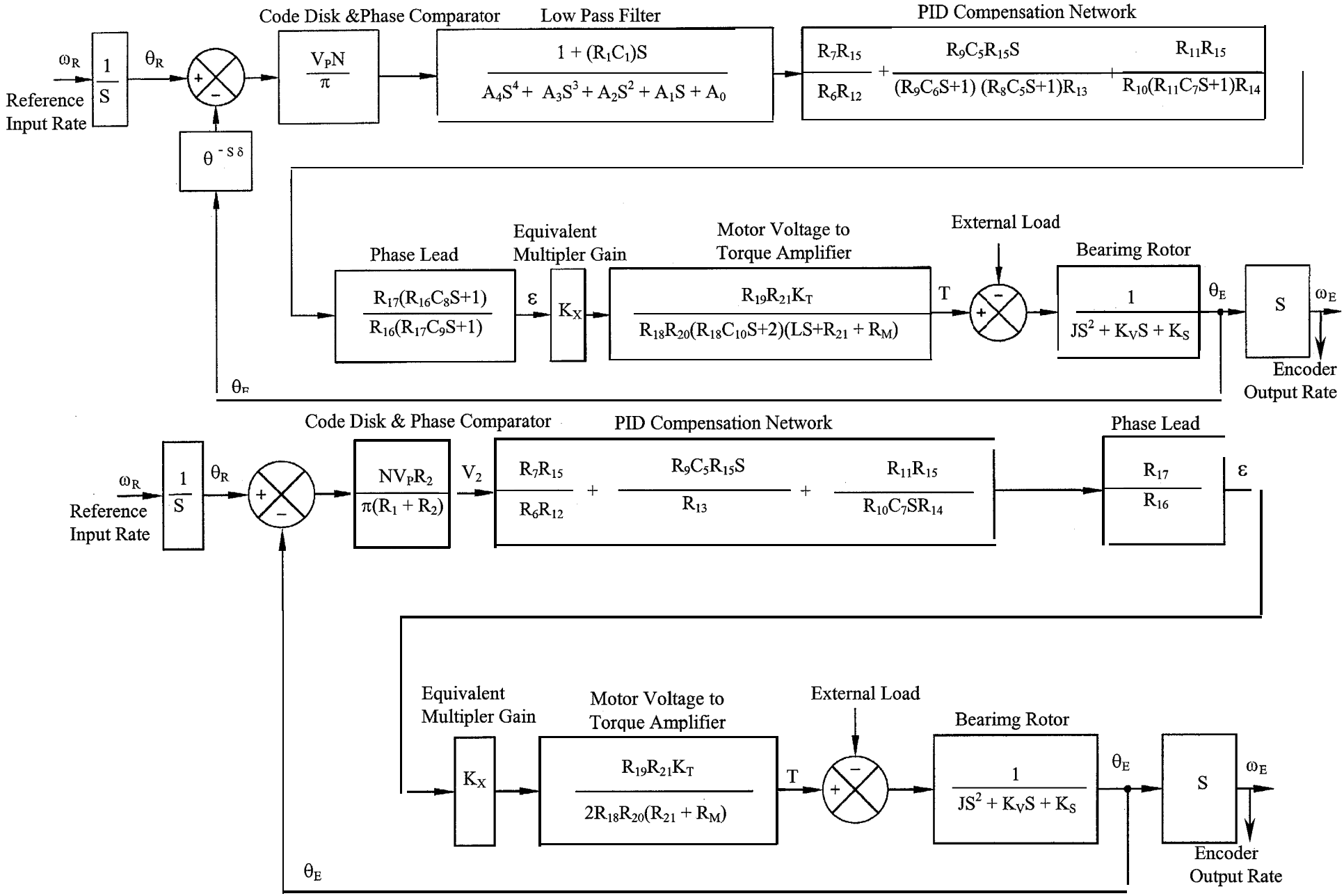
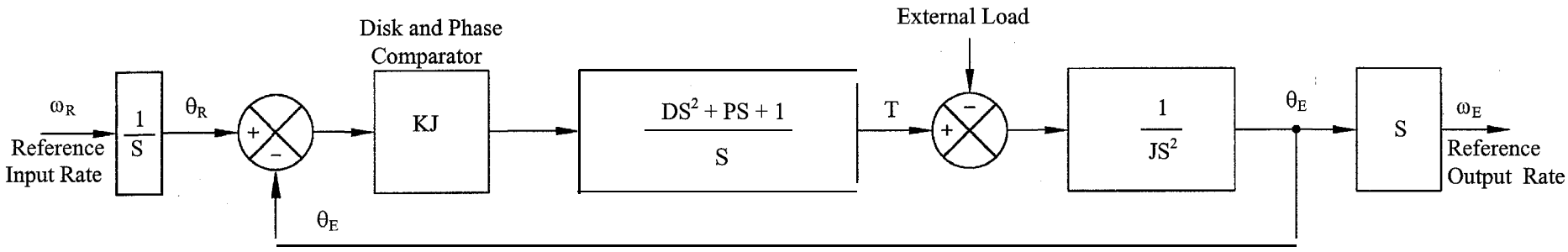


Figure 7. Actual Servo Design


Code Disk &Phase Comparator
Low Pass Filter
PID Compensation Network
Reference Input Rate
ω_R
θ_R
$\frac{1}{S}$
$\frac{V_P N}{\pi}$
$\frac{1 + (R_1C_1)S}{A_4S^4 + A_3S^3 + A_2S^2 + A_1S + A_0}$
$\frac{R_7R_{15}}{R_6R_{12}} + \frac{R_9C_5R_{15}S}{(R_9C_6S+1)(R_8C_5S+1)R_{13}} + \frac{R_{11}R_{15}}{R_{10}(R_{11}C_7S+1)R_{14}}$
$\theta^{-S\delta}$
Phase Lead
Equivalent Multipler Gain
Motor Voltage to Torque Amplifier
External Load
Bearimg Rotor
$\frac{R_{17}(R_{16}C_8S+1)}{R_{16}(R_{17}C_9S+1)}$
ε
K_X
$\frac{R_{19}R_{21}K_T}{R_{18}R_{20}(R_{18}C_{10}S+2)(LS+R_{21}+R_M)}$
T
$\frac{1}{JS^2 + K_VS + K_S}$
θ_E
S
ω_E
Encoder Output Rate
θ_E
Code Disk & Phase Comparator
PID Compensation Network
Phase Lead
Reference Input Rate
ω_R
θ_R
$\frac{1}{S}$
$\frac{NV_PR_2}{\pi(R_1 + R_2)}$
V_2
$\frac{R_7R_{15}}{R_6R_{12}} + \frac{R_9C_5R_{15}S}{R_{13}} + \frac{R_{11}R_{15}}{R_{10}C_7SR_{14}}$
$\frac{R_{17}}{R_{16}}$
ε
Equivalent Multipler Gain
Motor Voltage to Torque Amplifier
External Load
Bearimg Rotor
K_X
$\frac{R_{19}R_{21}K_T}{2R_{18}R_{20}(R_{21} + R_M)}$
T
$\frac{1}{JS^2 + K_VS + K_S}$
θ_E
S
ω_E
Encoder Output Rate
θ_E



Idealized Model

$$K = \left(\frac{NV_P R_2}{\pi(R_1 + R_2)}\right)\left(\frac{R_7 R_{15}}{R_6 R_{12}}\right)\left(K_X\right)\left(\frac{R_{17} R_{19}}{2R_{16} R_{18} R_{20}}\right)\left(\frac{R_{21}}{R_{21} + R_M}\right)\left(\frac{K_T}{J}\right)$$

$$A_0 = (R_1 + R_2)/R_2$$

$$A_1 = R_1C_1 + R_3C_2 + R_1(R_3/R_2)C_2 + R_1C_2 + R_5C_4 + R_1(R_5/R_2) \lesseqgtr + R_4C_4 + R_1(R_4/R_2)C_4 + R_1(R_3/R_2)C_4 + R_1C_4 + R_3C_4$$

$$A_2 = R_1R_3C_1C_2 + R_1R_5C_1C_4 + R_3R_5C_2C_4 + R_1R_3(R_5/R_2)C_2C_4 + R_1R_5C_2C_4 + R_4R_5C_3C_4 + R_3R_5C_3C_4 + R_1R_4(R_5/R_2)C_3C_4 + R_1R_3(R_5/R_2)C_3C_4 + R_1R_5C_3C_4 + R_1R_4C_1C_4 + R_3R_4C_2C_4 + R_1R_3C_1C_4 + R_1R_3(R_4/R_2)C_2C_4 + R_1R_4C_2C_4$$

$$A_3 = R_1R_3R_5C_1C_2C_4 + R_1R_4R_5C_1C_3C_4 + R_3R_4R_5C_2C_3C_4 + R_1R_3R_5C_1C_3C_4 + R_1R_3R_4(R_5/R_2)C_2C_3C_4 + R_1R_4R_5C_2C_3C_4 + R_1R_3R_4C_1C_2C_4$$

$$A_4 = R_1R_3R_4R_5C_1C_2C_3C_4$$

Figure 7. Actual Servo Design (Continued)

Table 1.
Selected Parameters for the System

Carrier Frequency	$F_c = 6.628$KHz
Encoder Fine Cycle Resolution	N = 16,384 cycles/rev.
Moment of Inertia	J = 0.05208 lb-ft-S^2
Maximum Motor Torque	$T_M = 65.15$ oz-in

$$KP = \frac{N\ T_M}{\pi\ J} = 33{,}979\ , \qquad K1 = \frac{KP}{(3)^{3/2}} = 1{,}306{,}407 \quad \text{and} \quad KD = \sqrt{3KP} = 319 \tag{38}$$

If P = 1, then K = 33,979, I = 35.48 and D = 9.39 x 10^{-3}

$$\omega_N = \sqrt{KP} = 184 \text{ rad/see}, \qquad f_N = \frac{\omega_N}{2\pi} = 29.28 \text{ Hz} \quad \text{and} \quad f_{3db} = 2.241, \quad f_N = 65.74\text{Hz} \tag{39}$$

Phase margin = 71.57° and Gain margin = 19.06 dB.

In order to optimize the system for the Hubble's very slow rates, increases in the loop gain and integral gain were necessary. The state variable model required a loop gain increase of 20% and an integral gain of 87.2, this produced a bandwidth of 120Hz. Calculations were made for step rates of 0.5, 8, 64, 512 and 4096 arcseconds per second, respectively.

For each rate, the time response for nominal and twice (2X) and half (½ X) values of proportional, derivative and integral gain values were taken. The twice derivative gain could not be shown because of induced oscillation due to this high gain. Derivative gain (1.25) was selected as adequately indicative of the trend induced by increasing D.

The trends evident from these plots are for low arcsecond/second rates, increasing integral gain results in lower positional error and a tighter lock to the required rate step. However, at high rates the opposite result occurs. At 4096 arcseconds/second, an integral gain below nominal values stated above tracks more closely with the step rate with positional error little improved. Although derivative gain could be "tweeked" a little for improvement at high or low step rates, the nominal D-gain is close to ideal for all rates. Variations in proportional gain had the least effect on step rate and positional error. Only small changes in gain and phase margins were evident.

System optimization was found to be a function of the step rate. This behavior could be caused in part by the bearing torque variance between starting and running torque. However, as the slow sidereal rates dominate performance requirements, the system was optimized for an 8 arcsecond per second rate and the specific K, P, D and I values determined. Figure 8 shows the final closed-loop gain, open-loop gain and phase plots for the system. From these plots the phase and gain margins were 51 degrees and 10 db respectively.

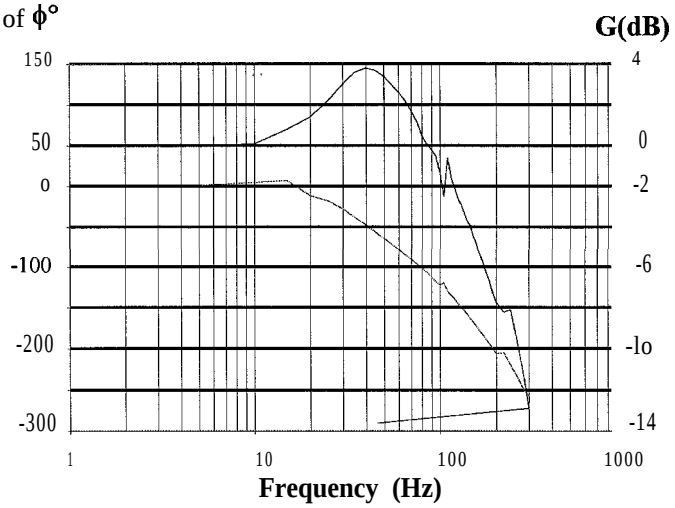


Figure 8. Gain-Phase Plot for Hubble SSSS

CONCLUSION

General solutions normalized for type II and type III carrier based phased-lock loop servo system are presented. While there are an infinite number of general solutions to such systems, by selecting the appropriate frequency as the normalization point, a single manageable transfer function can be developed. The Hubble Space Telescope's Star Selector Servo Subsystem is presented as a design example and the gain constants, closed-loop bandwidth, open-loop phase and gain margins, position errors, and velocity errors are evaluated.

ACKNOWLEDGMENT

The authors wish to express their gratitude to Linet Aghassi for her patient and meticulous typing of this manuscript.

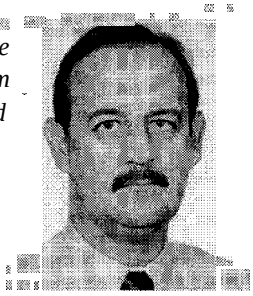
REFERENCES

[1] Hafle, Ralph S., "Wide Range, High Accuracy Electronically Programmed Speed Control System", U.S. Patent No. 3,974,428, issued August 10, 1976. Assignee: Baldwin Electronics, Inc., Little Rock, AR.

[2] Madni, Asad M. and Jumper, John M., "An Extremely Slow Motion Servo Control Technique used in the Hubble Space Telescope's Star Selector Servo Subsystem", Proceedings, 1997 NASA URC Technical Conference, Albuquerque, New Mexico, pp. 483-490.

BIOGRAPHIES

Dr. Asad M. Madni *is President and CEO of BEI Sensors and Systems Company, Inc., located in Sylmar, California. Prior to joining BEI in 1992, he was with Systron Dormer Corporation (A Thorn/EMI Company) for 18 years where he served in various senior level technical and executive positions, eventually as Chairman, President and CEO. He received the A.A.S. degree from RCA Institutes, Inc., B.S. and MS. degrees from University of California, Los Angeles (UCLA) and the Ph.D. degree from California Coast University (CCU), all in engineering. He is also a graduate of the Certificate Program in Engineering Management at the California Institute of Technology, the Executive Institute at Stanford University and the Program for Senior Executives at MIT Sloan School of Management. Dr. Madni is an internationally recognized authority with over 25 years of experience in intelligent system design and signal processing. He is credited with over 60 refereed publications and numerous patents resulting in "industry firsts". He serves in an advisory capacity for several professional and academic organizations. Dr. Madni is also the recipient of numerous awards and honors and is listed in 15 Who's Who publications including Who's Who in America. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), Life Member of the Association of Old Crows (AOC) and Member of the American Association for the Advancement of Science (AAAS), New York Academy of Sciences (NYAS) and the Society of Automotive Engineers (SAE).*



Mike Jumper *is presently serving in the capacity of Business Development Manager for BEI Sensors and Systems Company's Precision Systems and Space Division,. He was the System Mechanical Design Engineering for the Hubble Star Selector Servo Subsystem during the engineering and production programs. His responsibilities at BEI have included work on the GOES Scanning Mirror Motor/Encoders, UARS High Gain Antenna Position Sensors and the development of the 24 and 25-bit high accuracy commercial optical encoders. He received a BS degree in Mechanical Engineering from the University of Arkansas in 1980 prior to joining BEI that same year.*

HARDWARE IMPLEMENTATION OF A DELTA-SIGMA DECODER-DIGITAL APPROACH

Joseph Winkles, Loren Johns-Kaysing, Dj. Zrilic, G. Petrovic
NASA Center for Autonomous Control Engineering
New Mexico Highlands University
Department of Engineering
e-mail: j .wink@edison.nmhu. edu
Fax: (505) 454-3306
Las Vegas, NM 87701

Abstract: A simplified realization of a digital delta-sigma decoder, proposed in reference [2], is presented. Digital realization consists of an N bit-shift register, up-down counter, and some additional logic. The proposed realization can serve as a DM-PCM converter as well.

Keywords: digital delta-sigma decoder, pulse code modulation, digital signal processing

1. INTRODUCTION

The basic idea for this work is proposed by Zrilic *et al.* [2], Figure 3b, and the general principles of delta-sigma modulation are given in reference [1]. This work is a continuation of the work given in reference [3]. To process delta-sigma modulated (DSM) oversampled pulse density stream, dedicated hardware is needed. Here we propose to process a DSM stream with ordinary DSP hardware. To do so, a DM to PCM converter is needed.

In this paper, we propose to realize the idea given in reference [2] by Zrilic *et al*.

2. SIMULATION AND REALIZATION RESULTS

Proposed digital realization of a delta-sigma decoder is given in Figure 1.

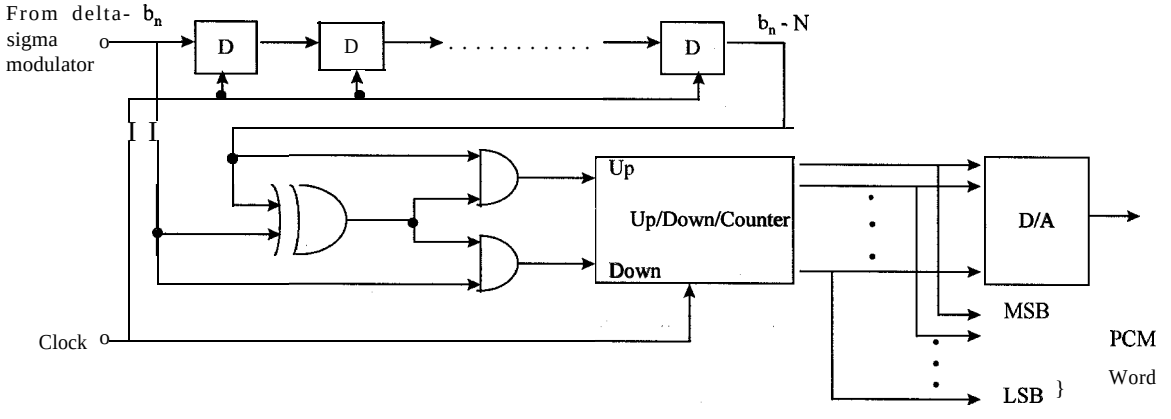


Figure 1. Proposed digital realization of delta-sigma decoder [2].

Consider binary values of-l and +1 or a logic “0” and “1”, respectively. Two AND gates and one XOR gate can be used to realize up count when input UP=1, stop when UP=DOWN=0. Binary states of the counter present pulse code modulation (PCM) words. These PCM words can be further used for additional digital signal processing with ordinary DSP hardware if needed, or A/D converted into analog signal.

To verify the proposed algorithm, we simulated it using *Electronic Workbench software* first. Figure *2* presents the electric diagram of the Electronics Workbench simulation.

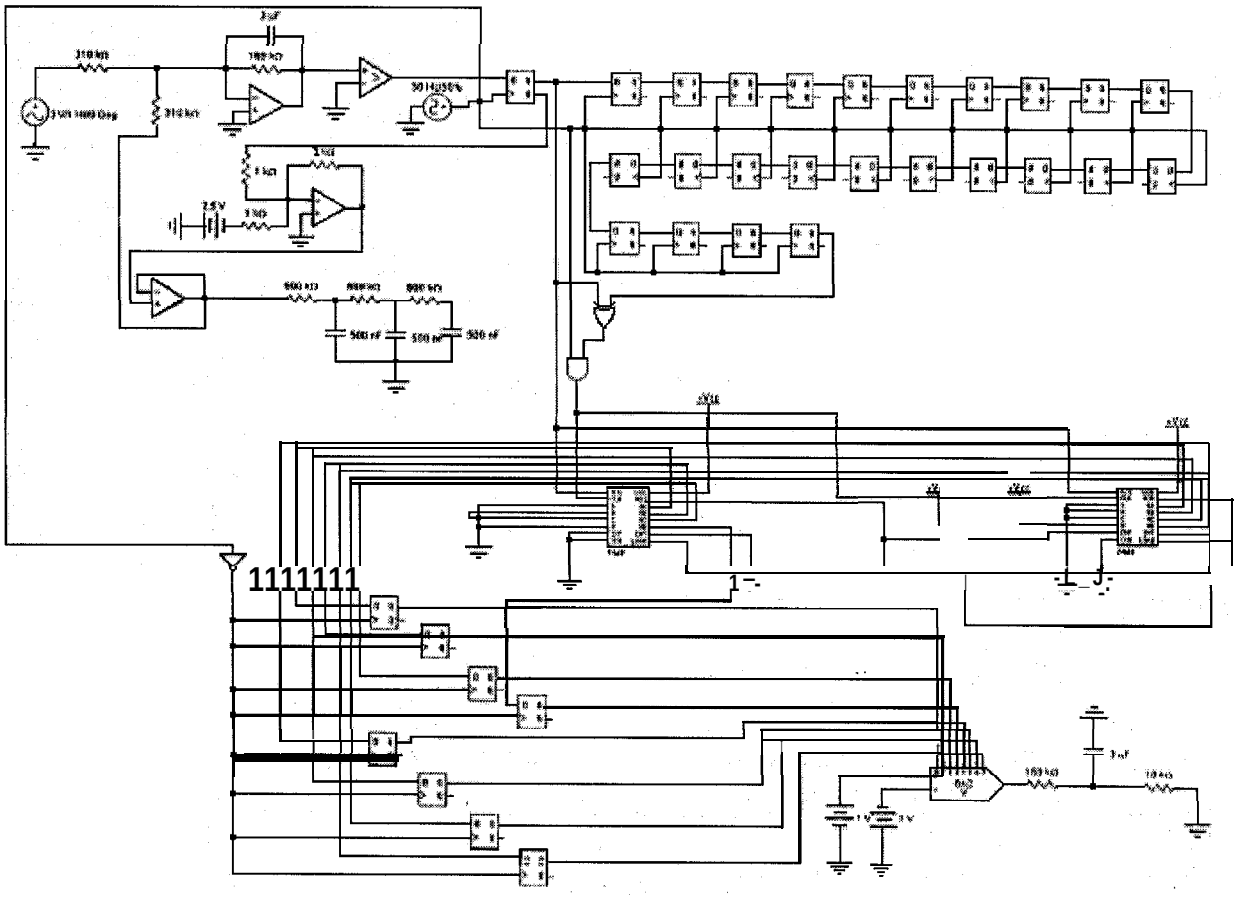


Figure 2. Simulation and Realization Electric Diagram

The output of DSM is fed into an N=24 shift register. At the same time, for control purposes, output of the DSM is fed into a 3rd order low pass filter. The first and last bits of the shift register are fed into combinational logic, The output of this logic is fed as the UP-DOWN command of the 8-bit counter. To synchronize the outputs of the counter, D flip-flops are used. The outputs

of the D flip-flops are fed into a DAC to acquire an analog output. Figure 3 presents the results of the simulation when a sinusoidal input is DSM converted.

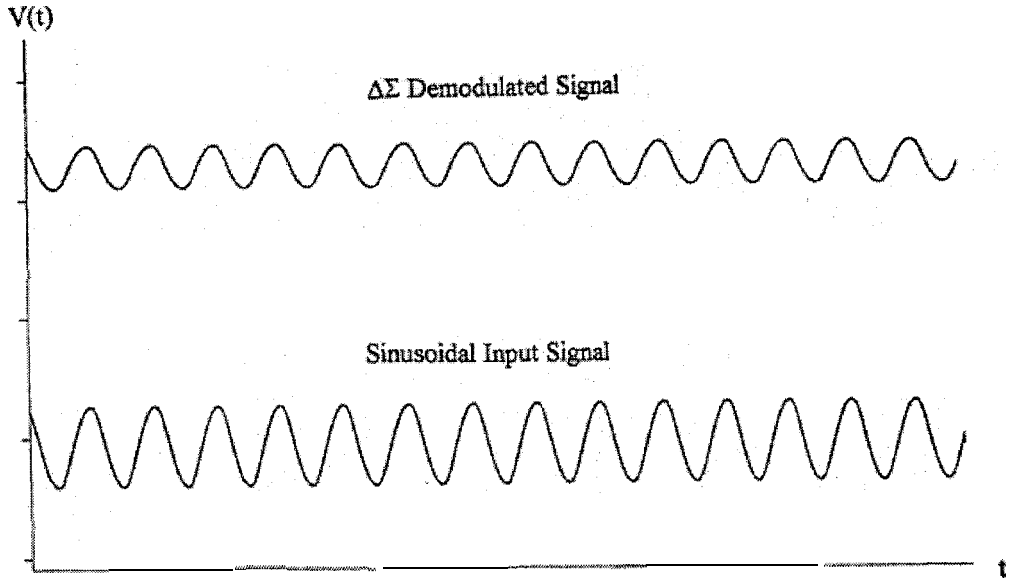


Figure 3. Results of Simulation for sinusoidal input

We see a pretty good agreement except that the output is delayed. Figure 4 presents the results of the realization according to the simulation diagram.

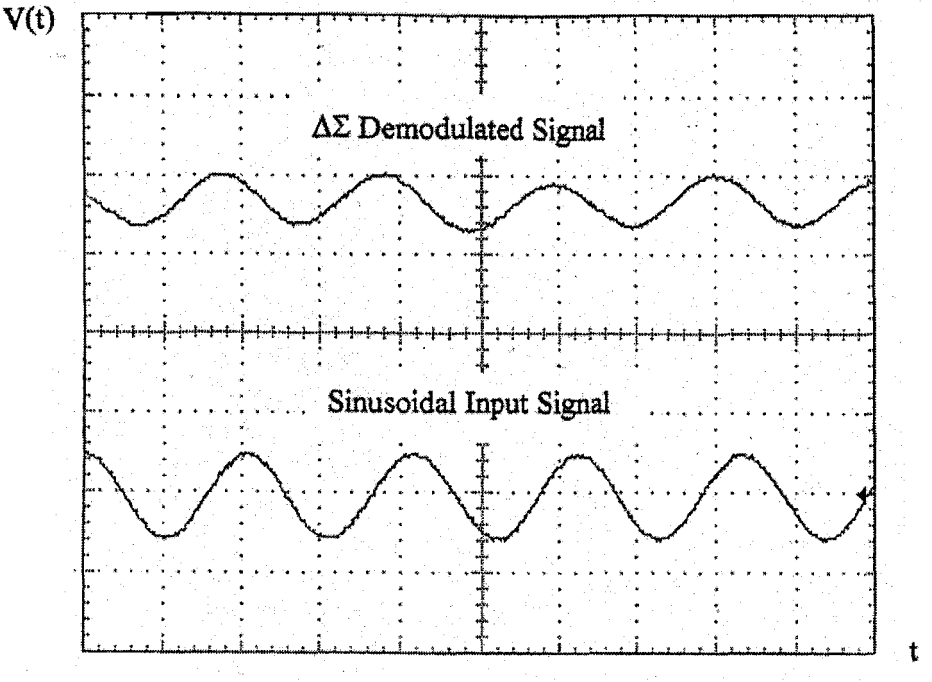


Figure 4. Results of Realization

The results are identical. The photo in Figure 5 presents the prototype board of our digital version of a DSM decoder.

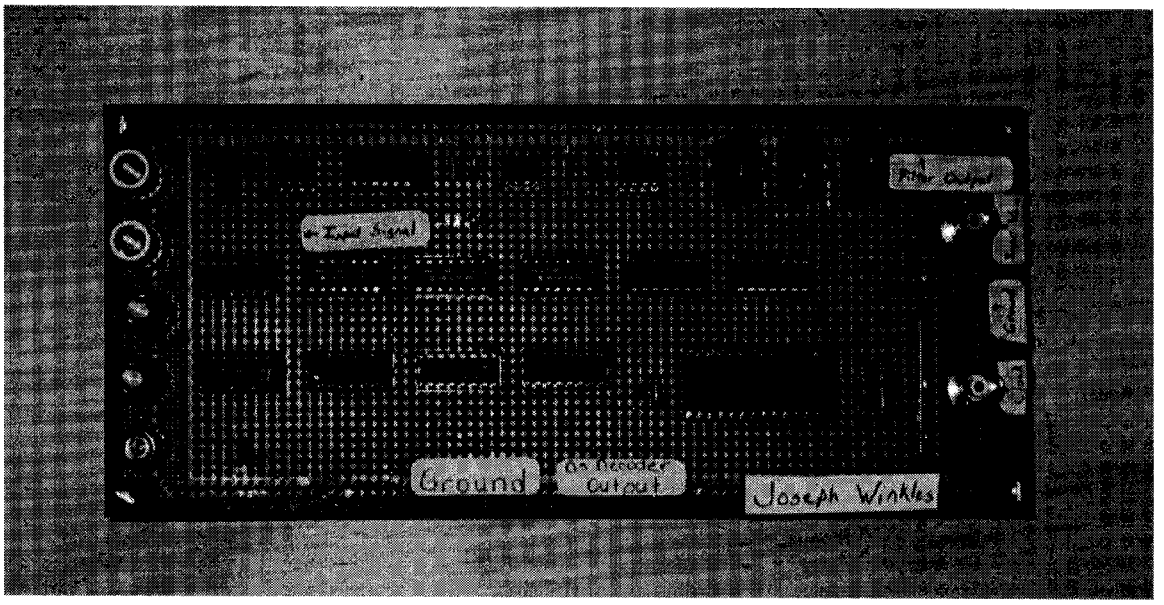


Figure 5. Filter Design Prototype

3. CONCLUSION

The proposed structure has been shown to be useful for the implementation of the conventional delta-sigma decoder. The digital realization consists of an N bit shift register, an up-down counter, and some logic. Its output can be A/D converted, or can be used for signal processing in direct form. It can be used as a delta-sigma PCM converter as well.

4. REFERENCES

[1] J. Candy, G. Temes. "Oversampling Delta-Sigma Data Conversion." IEEE Press, 1992.

[2] Dj. Zrilic, G. Petrovic, B. Yuan. "A Simplified Realization of a Delta-Sigma Decoder." Electronics Letters Vol.33No.18 (1997): 1515-16.

[3] L. Kaysing, J. Winkles, Dj. Zrilic. "Hardware Implementation of a Delta-Sigma Decoder-Analog Approach," paper published in this proceeding.

[4] Electronics Workbench
Interactive Image Technology Corporation
700 King Street West, Suite 815
Toronto, Ontario, Canada MSV ZY6

HARDWARE IMPLEMENTATION OF A DELTA-SIGMA DECODER-ANALOG APPROACH

Loren Johns-Kaysing, Joseph Winkles, Dj. Zrilic, G. Petrovic
NASA Center for Autonomous Control Engineering
New Mexico Highlands University
Department of Engineering
E-mail: loren_k@edison.nmhu.edu, Fax: (505) 454-3306
Las Vegas, NM 87701

Abstract: A simplified realization of the ordinary sigma-delta decoder is described. Analog realization consists of an N bit shift register, subtracter and integrator. Based on theoretical finding in reference [2], proposed realization requires only subtraction of first and last bits from the shift register.

Key Words: analog, delta, sigma, decoder, demodulator

1. INTRODUCTION

In general, a delta-sigma decoder is a linear or nonlinear low-pass filter whose role is to aggregate the "useful" signal spectral components and at the same time remove higher order spectral components of quantization noise. The basic principles of delta-sigma modulation (DSM) have been known for more than thirty years [1], and here we just briefly illustrate the main principle of DSM. Figure 1 shows a block diagram of a first order DSM system.

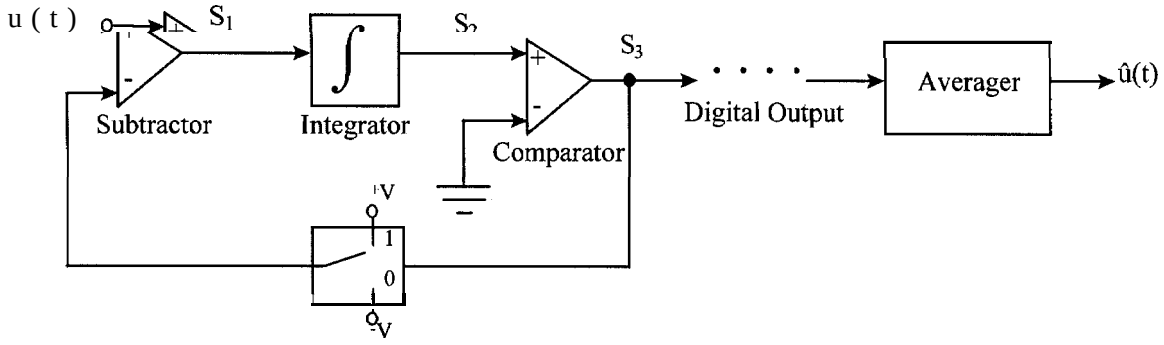


Figure 1. A first order delta-sigma modulator and demodulator.

The function of the first order DSM is illustrated in Table 1.

t	u(t)	s,	S*	S_3	û (t)	u(t) - û(t)
o	0.5	0	0	0	0	0.5
1	0.5	0.5	0.5	1	0.25	0.25
2	0.5	-0.5	0	1	0.5	0
3	0.5	-0.5	-0.5	-1	0.25	0.25
4	0.5	1.5	1	1	0.5	0
5	0.5	-0.5	0.5	1	0.5	0
6	0.5	-0.5	0	1	0.5	0
7	0.5	-0.5	-0.5	-1	0.5	0
8	0.5	1.5	1	1	0.5	0

Table 1.

In the table it is assumed that u(t) = 0.5 for all t. The output of the demodulator (averager) is

$$\hat{u}(t) = \frac{\Sigma_{t=0}^{3} S_3 (t - i)}{4}$$

which is the average of the current and the three previous valued of S_3. It is assumed that the initial valued of S_3 are $S_3(-1) = S_3(-2) = S_3(-3) = O$. We can see from Table 1 that after some time O(t) = u(t). That is, the original signal is fully recovered. There are many ways to implement a sigma-delta decoder (averager). Most probably, the simplest delta-sigma decoder is an ordinary finite impulse response filter (FIR) with uniform coefficients. In this paper, we implemented a novel approach proposed by Zrilic et al. in reference [2].

2. IMPLEMENTATION CONSIDERATION

Based on concept proposed in [2], transfer function of simplified DSM decoder is given as

$$H(D) = \left(\frac{1}{N}\right)\left(\frac{1-D^N}{1-D}\right) = H_1(D)H_2(D) \quad (1)$$

where

$$HI(D) = 1 - D^N \text{ and } HZ(D) = N^{-1}(1 - D)^{-1},$$

N is the length of shift register and D is unit delay. Figure 2 presents a block diagram of the delta-sigma decoder according to equation 1.

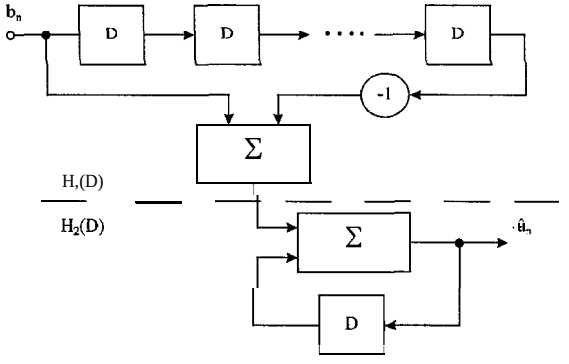


Figure 2. An implementation of the delta-sigma decoder according to equation 1.

There are different ways to realize the proposed structure from Figure 2. Here we propose analog approach. In reference [3] of this proceeding a fully digital approach will be presented.

3. ANALOG APPROACH

A simplified block diagram of this approach is given in Figure 3.

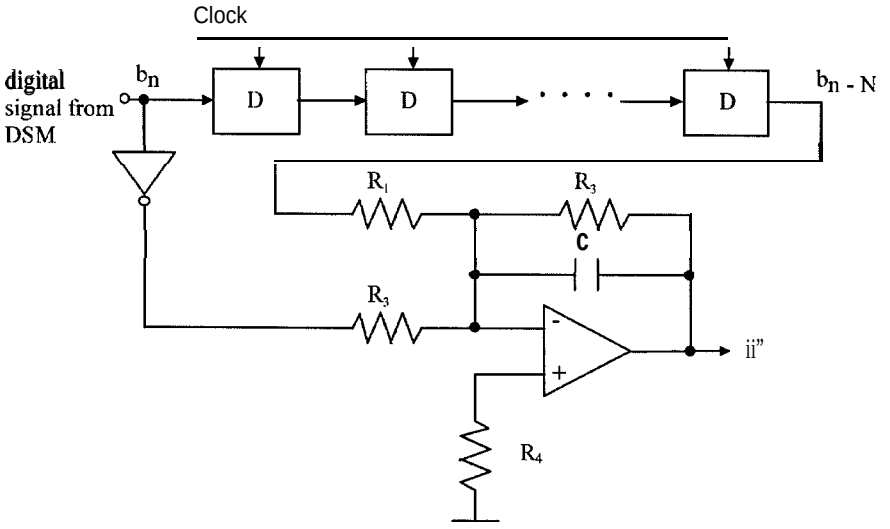


Figure 3. Proposed analog approach.

To verify correctness of this approach, simulation work was performed first. Figure 4 presents electric diagram of simulations performed with WORKBENCH [4]. Using this software, we were able to verify work of DSM, shift register (in our example, N = 24), summing circuit and integrator. Analog low pass filter of the 3rd order was used as reference only, Figure 4.

Figure 5 presents simulation results when input signal was sinusoid of frequency $f_n = 1$ Hz, and sampling frequency off, =50 Hz. We can see that the output signal $\hat{u}(t)$ is identical to the original u(t) and reference $\hat{u}'(t)$ except that it is phase shifted. The simulation diagram from Figure 4 is the realization diagram as well. Figure 6 presents the results of the realization taken from the scope. We can see pretty good agreement between theory, simulation and practice.

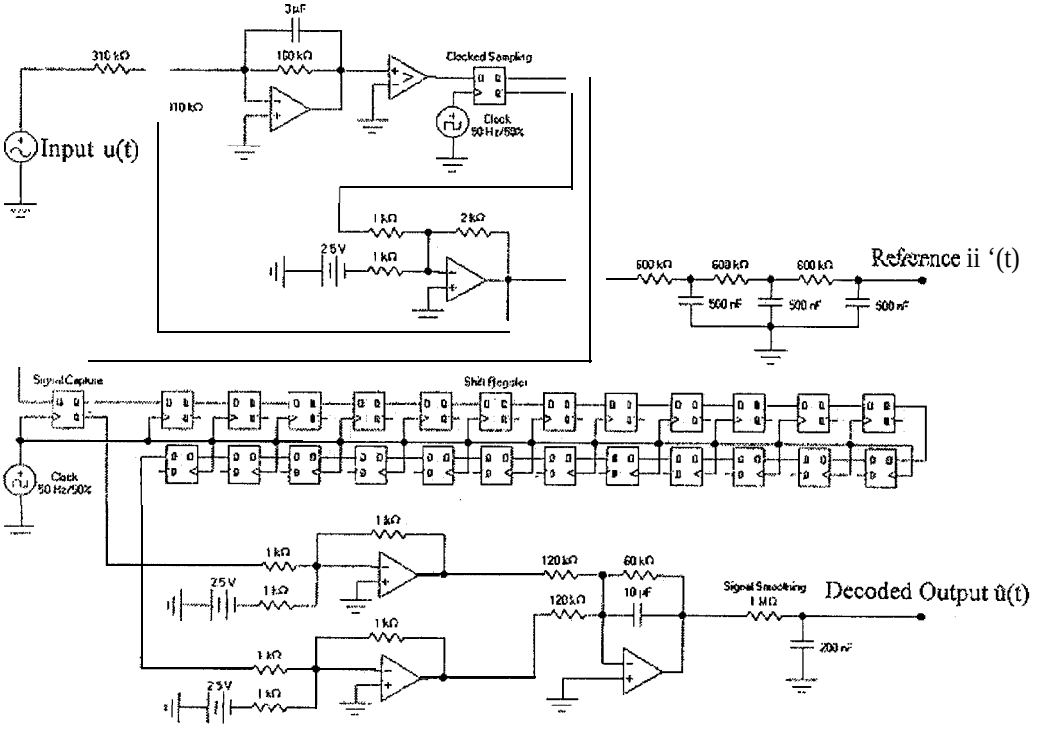


Figure 4. Simulation and Realization Diagram

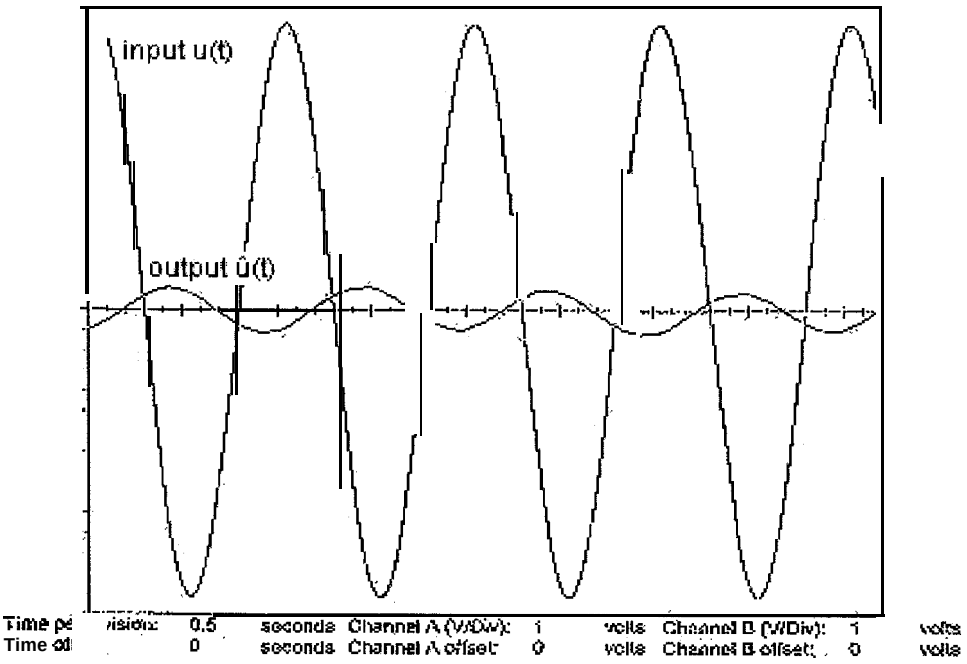


Figure 5. Simulation Results

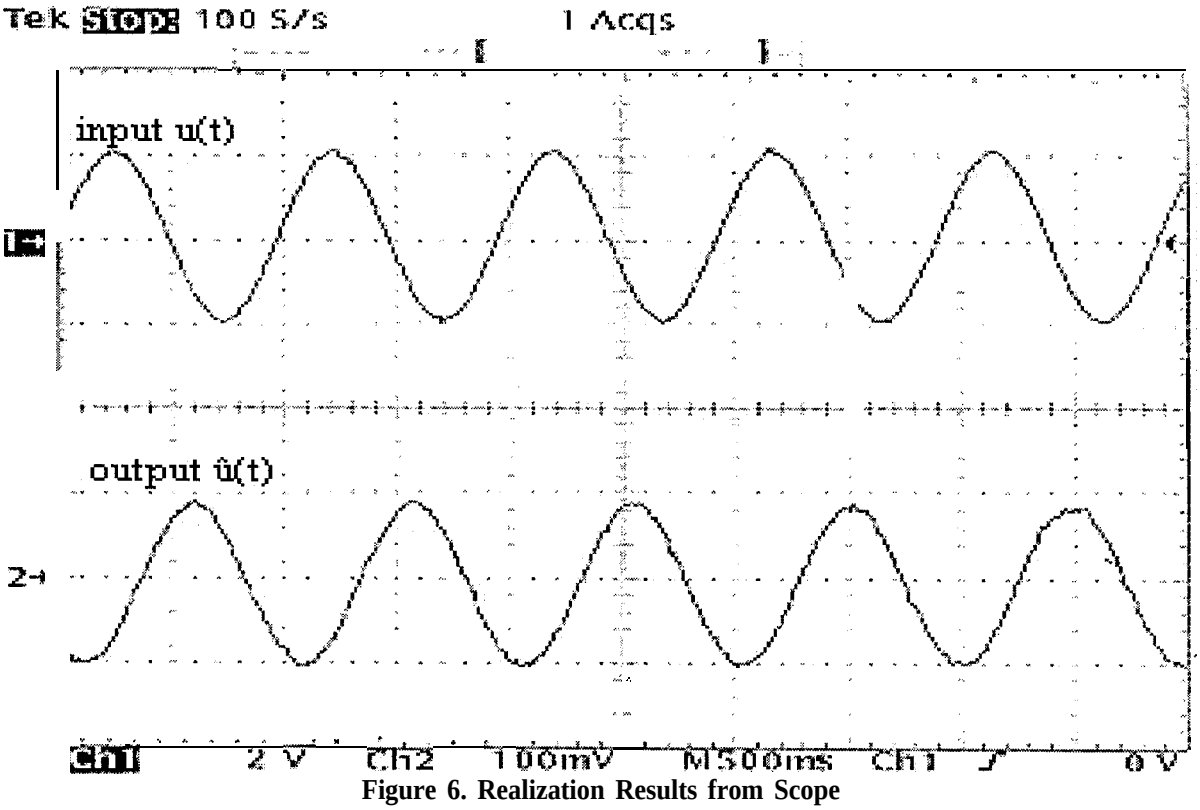


Figure 6. Realization Results from Scope

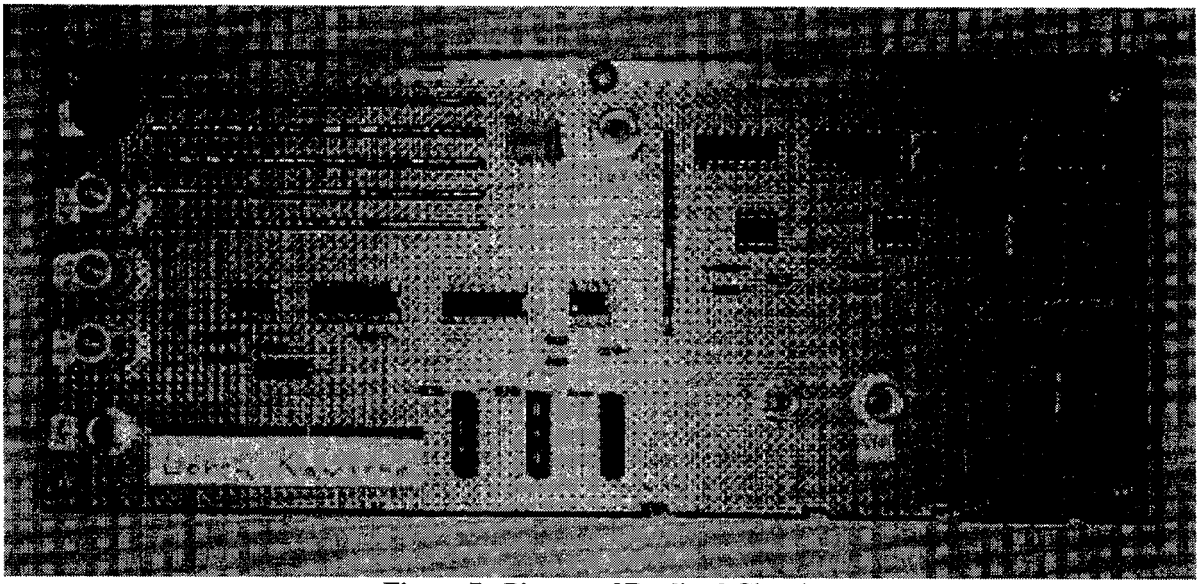


Figure 7. Picture of Realized Circuit

4. CONCLUSION

In this paper, we have simulated and implemented the analog structure of DSM decoder proposed in reference [2]. It has been proven that only the first and last digit of the N bit shift register need to be processed in order to get demodulated output. Simulation and realization results have proven the correctness of the proposed approach.

5. REFERENCES

[1] J. Candy, G. Temes. "Oversampling Delta-Sigma Conversion." IEEE Press, 1992.

[2] Dj. Zrilic, G. Petrovic, B. Yuan. "A Simplified Realization of a Delta-Sigma Decoder." *Electronic Letters* Vol. 33 No. 18 pp 1515-1516

[3] J. Winkles, L. Kaysing, Dj. Zrilic. "Hardware Implementation of a Delta-Sigma Decoder-Digital Approach.", paper published in this proceeding

[4] ELECTRONIC WORKBENCH
Interactive Image Technology Corp.
700 King Street W, Suite 815
Toronto, Ontario, Canada MSV ZY6

98 URC079

Visualizing and Classifying Satellite Imagery with Neural Networks

Brian S. Penn
Pan-American Center for Earth & Environmental Studies
University of Texas at El Paso
El Paso, TX 79968

915-747-7497
bpenn@geo.utep.edu

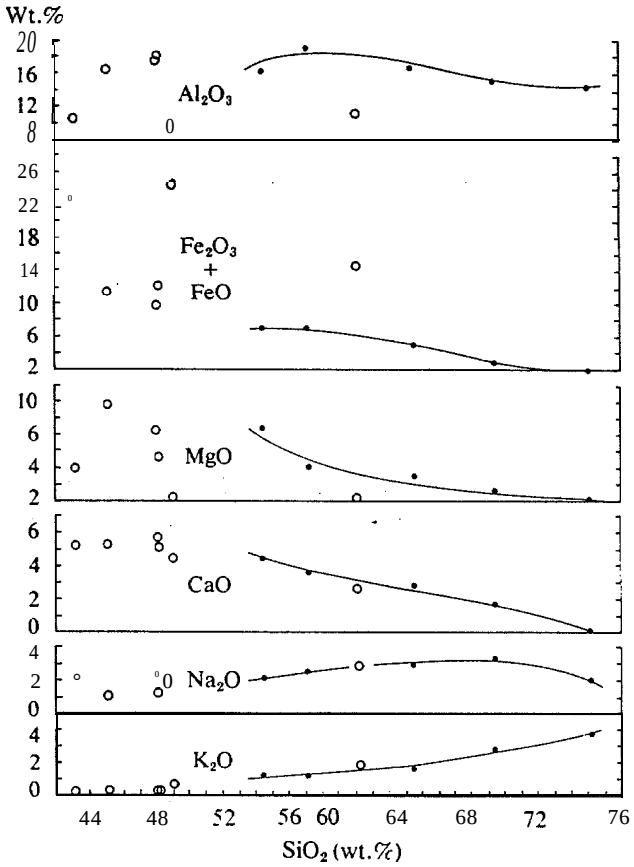
INTRODUCTION

Circa 1980 satellite technology such as NASA's Landsat 4 & 5's Thematic Mapper (TM) imagery is restricted in spectral, spatial, and radiometric resolution to <1 μm, 30 m, and 8-bits; respectively. At the time of their launches, these multispectral sensors were "state-of-the-art" with seven bands, or slices, of the electromagnetic spectrum. Plans for new satellites include resolutions of 10 nm spectrally, 1 m spatially, and 12-bits radiometrically. These types of sensors will gather Gigabytes of data. Current analytical approaches, especially those requiring human interaction, are inadequate to process and analyze these quantities of data as evidenced by the fact that greater than 95% of the more than 3 million Landsat images have not been seen by human eyes (Augusteijn and Dimalanta, 1992).

Consider the volume and complexity of TM data alone. Each full scene covers an area 185 km x 175 km; requiring over 290 Mbytes of storage space. Each 8-bit data band can be viewed as a 2-dimensional gray-scale matrix. Three of these 8-bit data bands can be combined on an RGB monitor to produce a color image. Each colored gun (red, green, or blue) displays the data over 256 discrete values. The resulting image may be true color depending on the bands chosen. When wavelengths outside the range (.4 -.7 μm) of human vision are displayed the image is referred to as false-color. Even with this information only 3/7s of the data are displayed at onetime. While approaches such as Principle Component Analysis (PCA) can reduce the dimensionality of the data, it is unclear what the resulting data means because the original data is lost in the PCA process.

When considering hyperspectral data (>20 bands), this problem grows exponentially. When faced with several hundred bands of data, does viewing the data in 3-dimensions justify collecting so much data? This problem is also encountered when analyzing multi-dimensional geochemical data.

Major element analysis of geochemical data is often overlooked because of the seemingly high dimensionality of the data, i.e., 11 components (SiO_2, TiO_2, Al_2O_3, Fe_2O_3, FeO, MgO, CaO, MnO, K2O, Na_2O, P_2O_3). The traditional approach to viewing these components is via Harker Diagrams where SiO_2 is on the abscissa (or x-axis) and one or more of the other elements is graphed on the ordinate axis. Other graphical approaches assume a pre-existing petrogenetic relationship among the data such as an AFM diagram (see figure 1). This often leads researchers to the wrong petrogenetic conclusions because their data is not from a mafic environment. Such



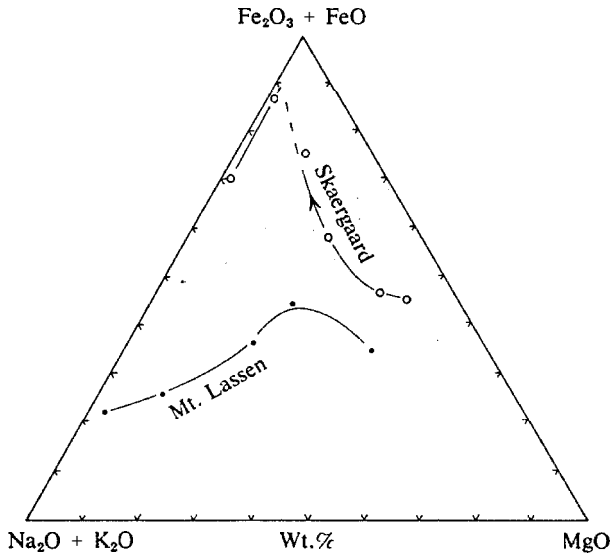


Figure 1. A) Harker variation diagrams for samples from Mt. Lassen in the Cascade **Mountains** and B) an **AFM** (Alkali **($Na_2O + K_2O$), Iron($FeO + Fe_2O_3$),** and magnesium **(MgO)** diagram for Mt. **Lassen** and the **Skaergaard** intrusion in Greenland **(after** Best, 1982).

is the case with the Spanish Peaks in South-Central Colorado. The Spanish Peaks are a Middle-Oligocene to Early-Miocene intrusive center with compositions ranging from alkalic to sub-alkalic. The earliest intrusive are alkaline. The second phase of intrusive activity involved sub-alkaline rocks. This was followed by a second phase of alkaline intrusive activity. Thermodynamics and experimental petrology preclude an evolutionary process capable of yielding this diverse suite of mineralogies, By analyzing the major element data in multiple dimensions, Penn (1994) was able to clearly establish previously unrecognized relationships among the data even though the Spanish Peaks had been extensively studied for almost a century. The approach used by Penn was rather laborious involving repeatedly graphing one chemical component against another, and watching for consistencies among samples. An automated approach akin to this is employed in *N-d visualizer* in ENVI, a hyperspectral imaging processing package from Research Systems, Inc. developed by Boardman and Kruse (1994). *N-d visualizer* enables the user to project the resulting classifications into 2 or 3 dimensions. While N-d *visualizer* represents a significant advance in terms of viewing multi- and hyper-spectral data, it is still quite labor intensive and often subjective. What is needed is an automated process that is both objective and maintains the topology of the original data. A relatively recent innovation in computer science and artificial intelligence is the development of self-organizing systems with these attributes.

SELF-ORGANIZING MAPS

Neural Networks (NNs) represent a shift in the research of how to get computers to think like human beings. From the 1940s - 1960s research in artificial intelligence focussed on developing systems loosely based on neurons in the human brain. This line of research showed promise until Minsky and Papert (1 969) demonstrated that NNs were unable to distinguish non-linear relationships. It was nearly 20 years before research using NNs began again with vigor. Minsky and Papert had shown a weakness of one particular approach using NNs, but less than five years later Werbos (1974) demonstrated a method for circumventing this problem.

During the 1970s and 1980s when NNs were viewed as a deadend, several researchers continued working with them including Hopfield, Tuevo Kohonen. Kohonen developed a very simple kind of NN (Kohonen, 1986). Kohonen's NN is composed of a single layer (either 1 or 2 dimensions) of nodes. Each node is composed of a vector with a magnitude of one and a particular direction inn-dimensional space. After normalization, multi-dimensional data is presented to each of the individual nodes. Using a "winner-take-all" competition, the node whose vector most closely matches the input data is found. This winning vector incorporates, or adjusts, its vector weights to match the input data, Vectors in the nodes surrounding the winning node are modified so that they are made to look less like the input vector. In this manner each node in the NN internally develops the ability to recognize vectors similar to itself This is referred to as self-organization, i.e., no external information is supplied to lead to correct classification. A Kohonen NN is called a self-organizing map for this reason.

One chief feature of the Kohonen NN is its ability to preserve the topology of the data presented to it. Topology is the study of the geometry of data sets. If a process preserves the topology, then upon completion the original relationships between the data points remains intact. This is exactly the desired result when working with multispectral data.

PURPOSE

The purpose of this research is to establish a quantitative relationship between hyperspectral data (collected either with a handheld spectroradiometer, airborne, or a satellite-based spectrometer) and geochemical measurements. Previous hyperspectral research focussed on utilizing measurements to classify/identify rock compositions at the pixel and subpixel level. This research will enable researchers to relate spectroscopy measurement to geochemical compositions. This approach will corroborate geochemical analyses and eventually provide a fast, low-cost substitute for large-scale exploration efforts.

Figure 2 (top) shows the results from a classification of the Jet Propulsion Laboratory (JPL) mineral spectra library JPL1sli.dat with a l-dimensional 10 node Kohonen NN. This particular library contains 159 reflectance spectra for many known minerals over the range of .4 - 2.5 μm. These spectra were normalized and presented to a l-dimensional Kohonen NN. Figure 2 (bottom) shows the results of classifying 109 major element geochemical analysis withal-dimensional 10 node Kohonen NN. These two resultant classifications are not related.

DISCUSSION

This paper is essentially a report of work in progress. It has been demonstrated that Kohonen NNs are capable of classifying diverse suites of multidimensional data including geochemical data and mineral spectra. Clearly the geochemical classification is based solely on major element abundances. The next phase of this research is to analyze the classification of the mineral spectral data in detail to ascertain the basis for the classification, i.e., are they mostly based on response or absorption features, or some combination of the two. At this time, analysis of the hyperspectral data for the samples collected from the Spanish Peaks region has not been done. It is expected that groupings found within the geochemical analyses and the spectral signatures will be very similar.

REFERENCES

Augusteijn, M. F., and Dimalanta, A. S, 1992. Feature Detection in Satellite Images using Neural Network Technology. *Telematics and Informatics*, *vol. 9,* no. 3/4: 131-144.

Best, M. G. 1982. Igneous and Metamorphic Petrology, W.H.Freeman and Company, New York, 630 p.

Boardman, J. W., and Kruse, F.A. 1994, Automated spectral analysis: A geologic example using AVIRIS data, north Grapevine Mountains, Nevada in Proceedings, Tenth Thematic Conference on Geologic Remote Sensing, Environmental Research Institute of Michigan, Ann Arbor, MI, p. 1407-418.

Kohonen, T. 1986, Representation of Sensory Information in Self-Organizing Feature Maps and Relation of these Maps to Distributed Memory Networks, *Proceedings of SPIE Advanced*

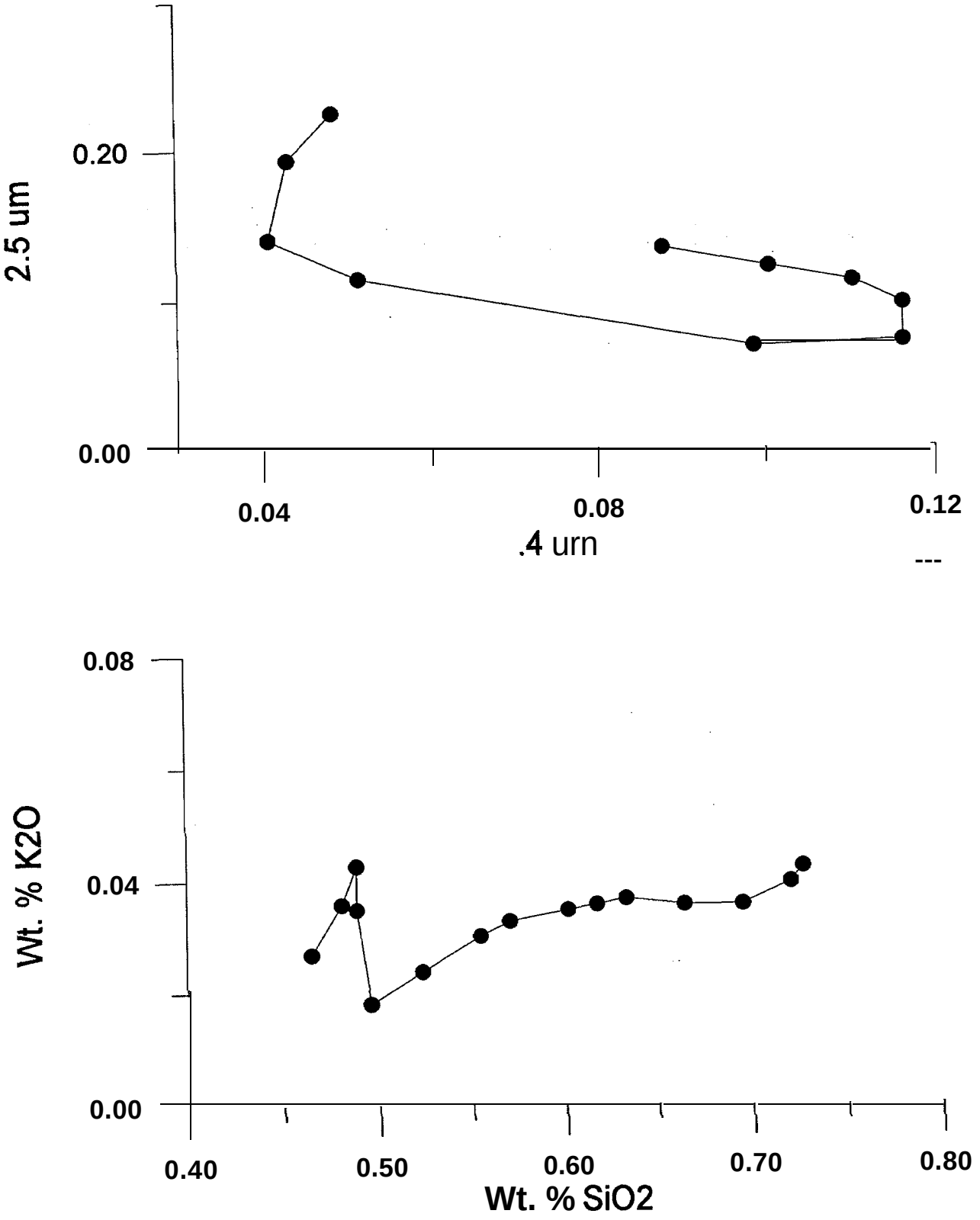


Figure 2. Top - Groupings for JPL1sli.dat dataset of spectra from Kohonen Self-Organizing Map with 10 nodes. Bottom - Groupings for 109 major element samples for the Spanish Peaks Intrusive Rocks of South-Central Colorado from Kohonen Self-Organizing Map with 15 nodes.

Institute on Hybrid and Optical Computing, vol. 634.

Minsky, M. L., and Papert, S. 1969. *Perceptions: An Introduction to Computational Geometry*, MIT Press, Cambridge, MA.

Penn, B. S. 1994. An Investigation of the Temporal and Geochemical Characteristics, and the Petrogenetic Origins of the Spanish Peaks Intrusive Rocks of South-Central Colorado, unpublished Ph. D. Thesis, Colorado School of Mines, Golden, Colorado, 199 p.

Werbos, P. J. 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph. D. Thesis. Harvard University.

Visualization in a Mach 2 Jet of Mach Wave Radiation

Edward Barber, G. Buzyna, A. Krothapalli
Fluid Mechanics Research Laboratory
Center for Nonlinear and Nonequilibrium Aeroscience
Florida A&M University
Tallahassee, FL 32310

Introduction

Since Lighthill introduced his general theories on sound generated aerodynamically (1952,1954), later to become the field of Aeroacoustics, people have been seeking ways to reduce jet noise. The latest driving force behind reducing jet noise is the interest in the development of a High Speed Civil Transport project (HSCT). This vehicle would be a supersonic successor to the present commercial airliner. In order to make the HSCT a viable project, the aircraft noise must be reduce to acceptable limits. There are many noise factors, such as the structural noise and engine combustion noise, but the noise produced by the turbulent jet, as studied by Lighthill (1954) and Ffowcs-Williams (1963) is our primary concern.

It is generally accepted that there are three major components of jet noise, see Tam (1995): 1) screech tones; 2) broad band shock associated noise; and 3) turbulent mixing noise (inclusive of Mach wave radiation). The first two components of noise exist because the jet is not isentropically expanded, meaning the jet is operated off design and shock cells are present in the jet core. The mixing noise component is always present, and the high frequency Mach wave radiation generated is quite prevalent in the first few jet diameters. It has been seen that the frequency of jet noise shifts from a high frequency to a lower frequency as one moves in the downstream direction. Some authors, for example Lighthill (1993), propose this shift is due to random, non-linear interactions ("bunching") mainly in the Mach direction, 6.

$$\theta = \cos^{-1}\frac{1}{M}$$

where M is the Mach number of the jet. This "bunching" would tend to imply that the high frequency waves coalesce to form higher intensity, lower frequency waves away from the nozzle exit. This "bunching" process initial occurs in a conical manner, but as more and more interactions occur the waves take on a more spherical nature. In this study we propose that the low frequency waves of a jet (specifically Mach 2) are a direct result of large scale turbulent structures in the jet, as opposed to "bunching". The idea is to block the high frequency Mach wave radiation from propagating downstream, and see if the low frequency waves still persist. The idea of the large scale structures generating the noise is consistent with Tam (1995), where he stated that fine scale turbulence contributes to background noise, while large turbulent structure dominate this noise component. Since this study focuses on the mixing noise component, the shock associated noise is eliminated by operating the nozzle at the nearly ideally expanded condition, corresponding to a nozzle pressure ratio (NPR) of 7.82. This study is novel in the fact that the jet is produced using a shock tunnel similar to the methods used by Oertel (1975) in the 1960's. And also the noise generated will be visualized, similar to Lowson & Ollerhead (1968) and Powell (1952), using the Schlieren technique.

Facility

The experiments for this study were carried out on the shock tunnel located at the Florida A&M University - Florida State University College of Engineering's Fluid Mechanics Research Laboratory (FMRL), see Figure 1.

Figure 1: Tunnel Schematic

The regions in the shock tunnel are usually denoted by specific numbers. The high pressure section is region four (4), the low pressure section is region one (1), the area behind the contact surface is region three (3), the region between the contact surface and incident shock wave is region two (2), and the area between the reflected shock and the end of the tunnel is denoted as region five (5). Figure 2 shows the trends for the different regions in the tunnel.

Figure 2: Shock tunnel Regions

The FMRL shock tunnel has a 6.67 cm inner diameter and a 8.89 cm outer diameter. The tunnel consist of a 2 m driver and a 6 m driven section as seen in Figure 1. A Bauer Model 1280-100 four stage reciprocating compressor compresses air into a series of four (4) storage tanks. The compressor can compress the air to a maximum pressure of 2000 psig. A Tescom Model 44-1115-24 pressure regulator feeds a steel braided flexible hose that supplies the driver section of the shock tunnel with a pressure ranging from 0-2000 psig.

The high pressure driver is separated from the low pressure driven section by a non-metallic (low to moderate pressure) or metallic (high pressure) diaphragm. The primary diaphragm is ruptured using a pneumatic plunger, and once ruptured, a shock wave is produced that propagated down the tunnel increasing both the temperature and the pressure inside. The wave reflects off of and ruptures the secondary diaphragm (made of paper) further heating and raising the pressure of the air inside. This twice heated air is the supply for the Mach 2 nozzle attached to the end of the tunnel. The driven section has several PCB piezo-electric pressure transducer mounted in the tunnel walls for tracking the shock wave produced monitoring its speed and the tunnels dynamic pressure, see figure 3.

The nozzle attached to the shock tunnel facility for this test was a Mach 2.0 axi-symmetric nozzle. The nozzle was designed using the method of characteristics, and had an exit diameter of 2.913 cm. Once the nozzle flow is generated a Schlieren system employing a EG&G Flashtube, with a duration 2.5 microseconds, a pair of 12" flat mirrors, and a pair of 12 _" parabolic mirrors are used to capture the flow field.

Figure 3: Pressure Trace

Results

Starting Process

One of the most important parameters in shock tunnel testing is the useful test time, or the time region 5 exist undisturbed. This region is denoted by the flat plateau in figure 3. Prior to achieving this steady pressure, the nozzle experiences unsteady start up with the pressure varying until the flow is choked. This is evident by the dip in the pressure trace after the initial shock. This unsteady start-up process can be seen in Figure 4.

Figure 4: Nozzle Starting Process

The process of how the jet evolves in time can be seen and the initial shock created by impulsively starting the jet can be seen propagate out of the images as not to bias the steady data.

Free Jet

Once the unsteady nozzle start-up is complete, the jet is given several milliseconds to evolve downstream (25-30 jet diameters). This fully developed jet is ideally expanded. The pressure in region 5 is approximately 115 psia, and the twice shock heated air has a stagnation temperature of 580 K. The time of the image acquisition can be seen by the spike in the upper curve of figure 3. A sequence of 3 images at three (3) different locations, at the same operating condition, are combined to make a composite image spanning 25 diameters.

Figure 5: Free Jet

In the first few diameters of the jet, high frequency Mach wave radiation can be seen radiating from the jet. It is interesting to note that at 8 diameter downstream the observable frequency has lowered and the angle of the radiation seems to have changed. A correlation can be seen between the frequency and size of the turbulent jet structures and the noise radiated. As the structures coalesce downstream, their frequency lowers as too does the acoustic waves visualized.

Isolated Jet

In an effort to give validity to the noise correlating to large structures, a radial disc, with a central hole, is inserted to block the downstream propagation of Mach wave radiation, but allows the jet to flow freely. This experiment was conducted with the disc at two (2) downstream locations, 10 cm and 20 cm, from the nozzle exit.

a) 10 cm position

b) 20 cm position

Figure 6: Isolated Jet

In figure 6a, the Mach wave radiation can be seen reflecting off the disc in the upstream direction, but as speculated some low frequency waves are still seen downstream. Some of the downstream waves in the free jet are missing in the isolated jet, but the waves of higher intensity persist. This would suggest that the high amplitude, low frequency waves are generated by the large scale structures and some of the lower amplitude waves seen downstream can be attributed to upstream waves, or "bunching".
In Figure 6b, the low frequency waves can be seen as far downstream as 17 jet diameters. Also if the waves ate traced backwards, the wave angle seem to have the sources existing beyond 10 jet diameters downstream.

Conclusion

From the experimental results, it appears there is a combination of both "bunching" and correlating with large scale structures to produce the low frequency noise downstream. The high frequency noise is attributed to Mach wave radiation, but the low frequency noise is composed of the noise generated by large turbulent structure, with a contribution from the upstream high frequency waves. The dominant, or highest amplitude, waves arise from the large structures. This would explain why most noise reduction techniques that affect the exit region of the nozzle yield only minimal reductions in the overall sound pressure level (OASPL) of the jet. It appears further investigation needs be made into the role of Mach wave radiation in OASPL. More importantly, more research into isolating the low frequency component of jet noise is needed, and then techniques to alter the jet farther (8-10 jet diameters) downstream can be studied to yield significant reductions in the OASPL making programs such as the HSCT a reality.

Bibliography

[1]Ffowcs-Williams, J.E. "The Noise from Turbulence Convected at High Speed," *Philosophical Transactions of the Royal Society Ser. A , 255,* Apr. 1963, pp. 469-503

[2]Lighthill, M.J. "On Sound Generated Aerodynamically: I. General Theory," *Philosophical Transactions of the Royal Society Ser. A , 211,* Mar. 1952, pp. 564-587

[3]Lighthill, M.J. "On Sound Generated Aerodynamically: II. Turbulence as a Source of Sound," *Philosophical Transactions of the Royal Society Ser. A , 222,* Feb. 1954, pp. 1-32

[4]Lighthill, M.J. "Some Aspects of the Aeroacoustics of High-Speed Jets," *[CASE Report. 93-20,* May 1993, pp. 1-43

[5]Lowson, M.V. & Ollerhead, J.B. "Visualization of Noise from Cold Supersonic Jets," *J. Acoust. Sot.* Am. , 44, 1968, pp. 624-630

[6]Oertel H. "Jet Noise Research by means of Shock Tubes," *Shock Tube and Shock Wave Symposium , 1975,* pp. 488-49;

[7]Powell, A "on the Noise Emanating from a Two-Dimensional Jet Above the Critical Pressure," *The Aeronautical Quarterly ,4,* Feb. 1953, pp. 103-122

[8]Tam, C.K.W. "supersonic Jet Noise," *Annual Rev. Fluid Mech., 1995,* pp. 17-43

Turbulence Distortion of Combustion Energy Transfer Rates

Frantz Camulaire, U.P. De Silva, T.Mochena, and J.A. Johnson III
Laboratory for Modern Fluid Physics
Center for Nonlinear and Nonequilibrium Aeroscience
Florida A&M University, Tallahassee, Florida

Introduction The significance of turbulence within the velocity field of an internal combustion (IC) engine is well established from qualitative optical diagnostics. However, very little use is made of the unique possibilities which turbulence might afford for reaction control and performance manipulation. In fact, despite dedicated efforts of many scientists throughout the years, turbulence continues to persist as one of the least understood aspects of combustion in science and technology. More specifically, Lancaster et al. reported that the velocity field in an IC engine is turbulent and non-stationary. The turbulence is important because of its effect on the flame speed, and in turn the combustion rate. Whether a system is in chemical equilibrium depends on whether the time constants of the controlling chemical reactions are short compared with time scales over which the system conditions (temperature and pressure) change. Chemical processes in engines are often not in equilibrium. Experimental observations show that turbulence in a nonequilibrium process can distort the reaction rate and that multi-step relaxation processes can afford selective turbulence distortion of reactive processes. In principle, selective turbulence distortion allows an independent manipulation of emissions and efficiency in the internal combustion engine using additives. The main objective therefore, will be to determine the circumstance under which the selective distortion might be implemented.

This paper presents Hot Wire Anemometry (HWA) measurements made in a motored IC engine sampled at 100KHz sampling rate. Several computational tools have been utilized to analyze the statistical characteristics of the evolution of turbulence. The power spectrum is used to display the frequency composition of the time variation of the dynamical variables and to derive the power law behavior whereby the spectral indices can be calculated based on the different slope values. Following Hinze (1975), the expected spectral power distribution of the in-cylinder flow field should exist in two characteristic regions. The first one, in which the flow exhibits large eddies, will exist mainly at low frequency (100 to 1,000 Hz). For the second one, in which the flow exhibits small eddies, the power spectral distribution will exist mainly in the region of high frequency (1,000 to 100,000 Hz). This region, the Kolmogorov -5/3 power domain, is also characteristic of an isotropic flow field at high Reynolds number where the turbulent power, E, can be expressed as $E = kf^{-5/3}$, where k is the turbulent dissipation power. In addition, the attractor description of the turbulence (which is a direct result of the modern computer technology and a relatively new development in science) is used in analyzing the flow characteristics by using chaotic dimension analyses. These turbulent indicators such as chaotic dimension, D_2, and spectral index, n in $P \sim \omega^{-n}$, turbulent intensity, I, as well as the Reynolds number, Re, are computed for several speeds ranging from 300-1000 rpm and also for different radial positions in the engine combustion chamber. The research is aimed at explicitly determining the applicability of the dominant mode approach to the dynamics of a motored engine and to prepare explicitly for the coupling of turbulence and nonequilibrium thermodynamics in internal combustion engine dynamics.

The measurement technique Hot Wire Anemometry (HWA) provides a powerful tool for the investigation of in-cylinder turbulent flow field in a motored engine. A vast literature already exists on this subject. Ever since Semenov's pioneering use of a hot wire anemometer to measure the fluid motion in a motored reciprocating engine, researchers have been routinely utilizing this diagnostic tool for its relatively high sampling rate capabilities.

Experimental Set-up and Procedures The general layout of the Internal combustion engine coupled with the dynamometer configuration is shown in Fig. l(a). A 15 HP, single cylinder, four stroke reciprocating internal combustion engine has been modified from commercial to research grade. The engine is motored by a 20 HP, 250 VDC armature, 92 amp, General Electric dynamometer model TLC-15 with a 15 kVA isolation transformer. The dynamometer control system is comprised of a four quadrant D.C. control system and the Dyn-Loc III digital controller. The top view of the engine's cylinder head and the Hot Wire Anemometry (HWA) system are shown in schematic form in Fig. l(b). The TSI's model 1214 streamlined probe is inserted into one of the four ports in the cylinder head. The probe was traversed in the combustion chamber using a Lintech Model 41583, three-axis- traverse-system with one micron resolution.

The velocity measurements were taken using a TSI IFA-300 Constant Temperature Hot Wire Anemometry system which consists of a built-in signal conditioner and a thermocouple circuit, as well as an optional integrated calibration system. In this system a high sampling rate of 300 KHz is facilitated by the use of SmartTune automatic bridge optimization technology. The pressure measurements were taken by the Kistler model 4075A20 piezoresistive pressure transducer and were captured using a four-channel 500-MHz Tektronics Model TDS744 color digitizing oscilloscope with a maximum record length of 130K data points per channel. A two channel, Tektronix Model 5110 programmable pulse generator was utilized to trigger the oscilloscope and the IFA-300 external trigger board, whereby capturing both the pressure and the velocity simultaneously. These measurements were taken for several crank shaft speeds ranging from 300-1000 rpm and also for different radial positions (0-70 mm) in the engine combustion chamber using a 100KHz sampling rate.

The HWA measures the fluid velocity by sensing the changes in heat transfer from a small electrically-heated element (approx. 4 microns in diameter) exposed to the fluid. In the "constant temperature anemometer" the cooling effect produced by the flow passing over the element is balanced by the electrical current to the element, so that the element is held at a constant temperature. The change in current due to a change in flow velocity shows up as a voltage at the anemometer output. The anemometer is linked to a personal computer, where the data are displayed, analyzed, and stored.

Data Analysis The rapid development of computational methods has meant that a continuous signal x(t) is very often measured by sampling and discretizing. Therefore, an experiment generally provides a discrete sequence of real numbers x_j regularly spaced at time intervals of At. Fig. 2(a) illustrates a sample of the discretized velocity measurements (x_j) from the motored I.C. engine sampled at 100 KHz, 800 rpm, 20 mm from the cylinder wall, using 32,000 data points (N). As we implement spectral analysis techniques with a computer, it is necessary to transform the sampled time data lengths into the frequency domain via the discrete Fourier transform. We define the Fourier transform of the discrete velocity sequence (x_j) to be the operation creating a corresponding discrete series X_k such that:

$$X_k(f_k) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} x_j \exp\left(-i\frac{2\pi jk}{N}\right)$$

where: k=1,.....N, $i = \sqrt{-1}$ pure imaginary, $f_k = k\Delta f = k\left(\frac{1}{N}\right)$

For convenience At is taken as the time counter. The graph representing X_k^2 as a function of the frequency f is called the power spectrum. In the turbulent state, the power spectrum falls off as f-n, where the spectral index n, can be obtained from the slope of the plot of log (X_k^2) versus log(f) (referred in general as the power law). See Fig.2(b) for a sample of the full-cycle power spectrum. An overall turbulent behavior is observed along with a distinction between the low frequency and high frequency spectral index n as anticipated in the discussion above.

In order to give a description of the degree of complexity in the flow field, it is necessary to determine a chaotic dimension analysis. The chaotic dimension (D_2) gives lower limit on the minimum number of dimensions required to describe the evolution of the system. There are several different ways of computing chaotic dimensions (or fractal dimensions of a chaotic/turbulent system) from time series data. The Grassberger and Procaccia (G-P) method applies to the computation of dimension, D_2, for time series data such as $U_i = U(i\Delta t)$ where i runs from O to some large number Npts, which is the number of points in the data set. U_i data are embedded in an M-dimensional phase space yielding N vectors,

$$\text{`}j = \left(U_{jp}, U_{jp+q}, U_{jp+2q}, \ldots\ldots, U_{jp+(m-1)q}\right)$$

Besides the parameters M and N, this also introduces as free parameters, the integers p and q, which describe the time spacing between the beginning points of these vectors and the time spacing between component elements of an individual vector. Each pair of vectors in this space has a distance between them, which is chosen hereto be the Euclidean norm,

$$r_{jk} = \left[\left(u_{jp} - u_{kp}\right)^2 + \left(u_{jp+q} - u_{kp+q}\right)^2 + ... + \left(u_{jp+(M-1)q} - u_{kp+(M-1)q}\right)^2 \right]^{1/2}$$

the G-P correlation function is defined theoretically as $C(r) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j=1}^{N} H\left(r - \left|r_{jk}\right|\right)$ where $H\left(r - \left|r_{jk}\right|\right)$ is the Heaviside function. C(r) could be understood as the number of pairs of points (j,k) whose r_{jk} *is* less than or equal to r; and chaotic dimension, D_2, could be defined by scaling as C(r) - r^{D2} as r-> O. The chaotic dimension can be obtained from the slope of the plot of log Np(r) versus log(r), where Np(r)=C(r)/0.5N^2. The scaling of 0.5N^2 is introduced so that Np always reaches the same value, Np=1, when r is large enough to make C(r) include all possible pairs. See Fig. 3(a) for a sample of the normalized plot of the G-P correlation function, log Np(r) versus log(r).

Results and Conclusions The chaotic dimensions, D_2, and the spectral indices, n, were calculated for several crank shaft speeds ranging from 300 - 1000 rpm and also for different radial positions (O - 70 mm) in the engine combustion chamber. Fig. 3(b) shows a typical evolution of turbulent parameters during an engine cycle. We have restricted ourselves to the expansion and the compression phases because, during those phases, our measurement of the local pressure and temperature allow a complete determination of the local Reynolds number. Figs. 4(a) and 4(b) show the results for the expansion stroke where we measure rms velocity fluctuations as a function of Reynolds number for fixed locations (changing RPM) and at fixed values of RPM (changing locations). We have performed similar comparisons for our measurements of power spectral index. Figs. 5(a) and 5(b) show the results for the expansion stroke where we measure spectral power index from velocity fluctuations as a function of Reynolds number for fixed locations (changing RPM) and at fixed values of RPM (changing locations). The overall results with respect to the turbulence parameters for the compression stroke are the same as those for the expansion stroke.

There are several clear results from the measurements to date. First, variations in the correlation chaotic dimension are found during a given engine cycle, which are unique and reproducible on a cycle-to-cycle basis. The variations are both significant and systematic. Secondly, it is observed that cycle-to-cycle variations in turbulent intensit y do not produce important variations in the parameters which measure degree of complexity in the turbulence. This robustness in the turbulence parameters is also found when comparisons are made over a range of engine rpms. In addition , it is now easy to speculate that, when combustion is turned on in such an engine, the turbulence will have the mode or modes associated with this motored behavior as a background to any other turbulent effects which are specific to the combustion environment. This provides immediate opportunities for analyses of the turbulence and turbulence-based manipulations in an internal combustion engine using mode-mode coupling.

References

1. Lancaster, D.R., Krieger, R.B., Sorenson, S.C, and Hull, W.L., "Effects of Turbulence on Spark-Ignition Engine Combustion," SAE Paper
2. Glover, A.R., Hundleby, G.E., and Hadded, O., SAE Paper no. 880379, 1988
3. Catania, A.E., and Mittica, A., Asme, 85-DGP-12, 1985
4. Corcione, F.E., Bertoli, C., Police, G., and Valentine, G., International Symposium COMODIA, 1985 pp. 273, 283
5. Hinze, J.O., Turbulence, McGraw Hill, New York 1975
6. Semenov, E. S., 1958: Device for measuring the turbulence in piston engines. Instrum, Exp. Tech. 1, 102-111.
7. De Silva, U.P., Reynolds Stresses & Turbulent Closure in Supersonic Shear Flows, *PhD Dissertation, 1995* pp.70-76.

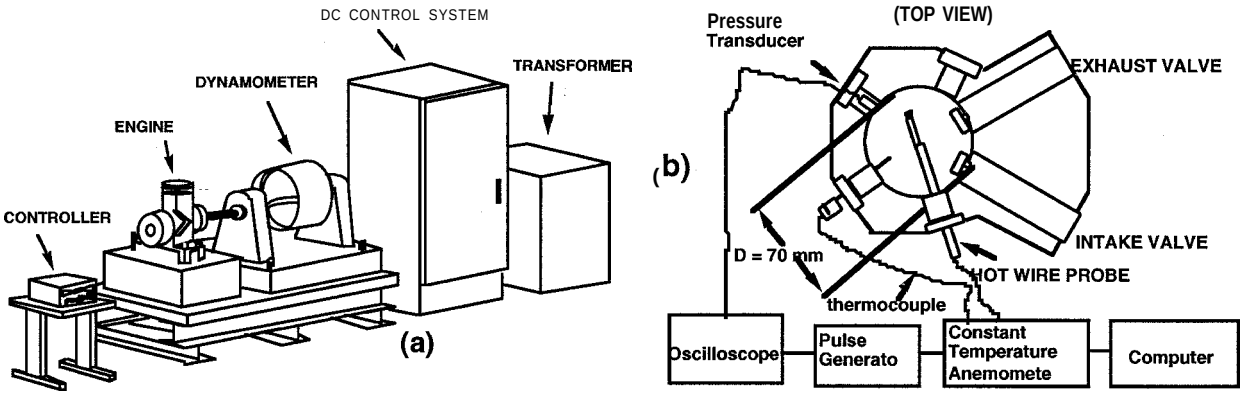


Figure 1. (a) General Layout of the Internal Combustion Engine-Dynamometer Configuration. (b) Schematic of the Cylinder Head with Hot Wire Probe

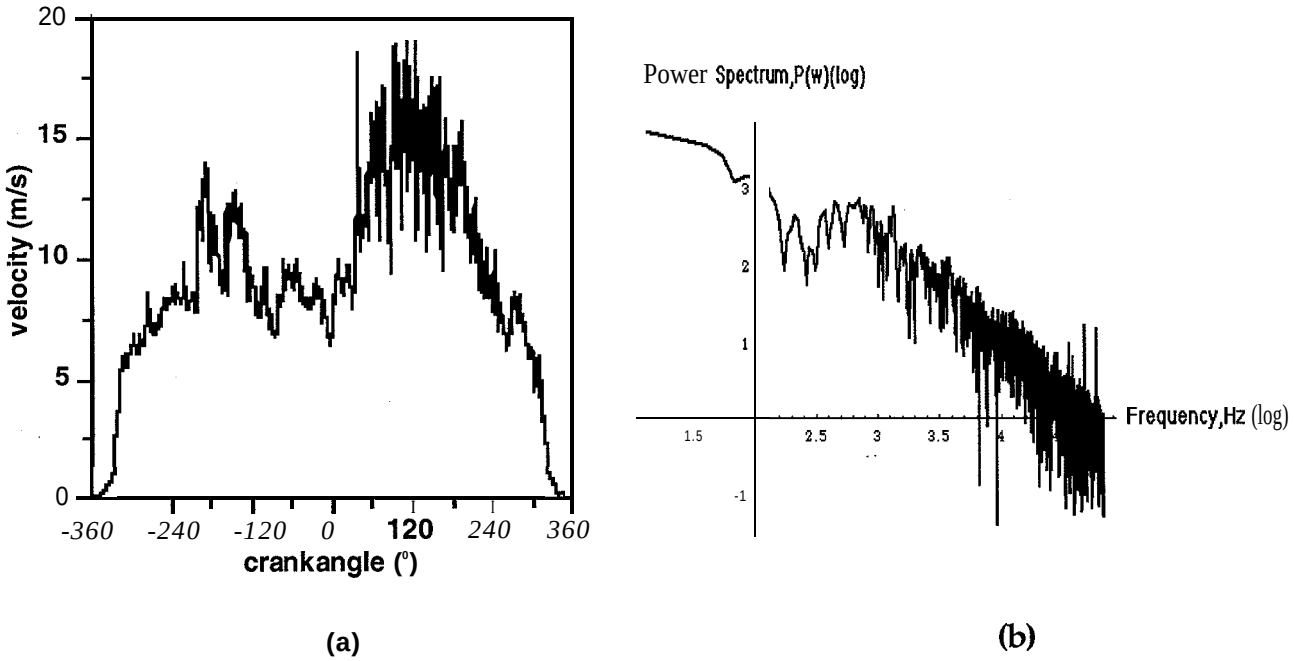


Figure 2. (a) Complete Set, Velocity Data, 800 RPM, x = 20 mm. (b) Power Spectrum of Velocity. (One cycle, 15,000 pts, 800 RPM, 100 KHz Sampling Rate at x = 20mm)

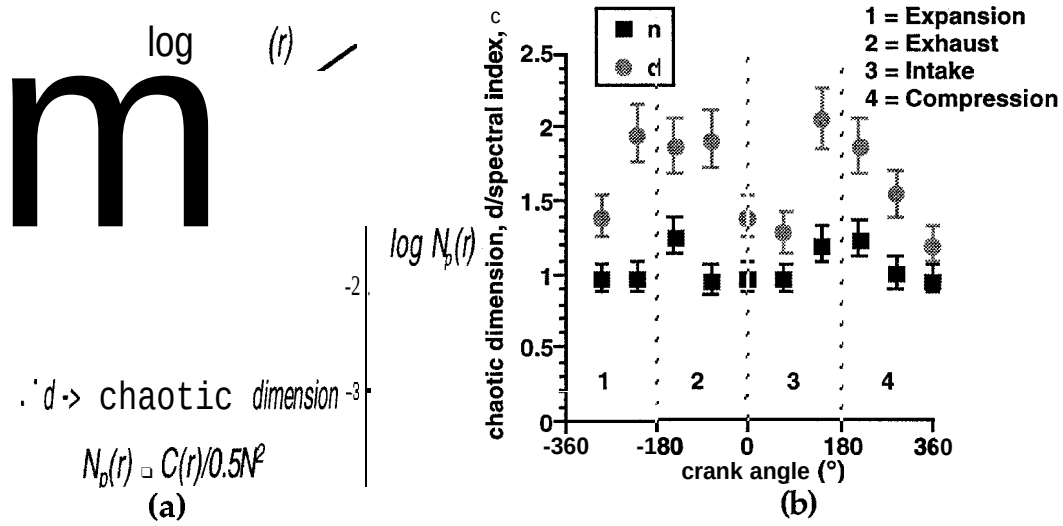


Figure 3. (a) Normalized Plot of Correlation Integral. (b) Turbulence Indicators in the Velocity. (800 RPM, x =20mm, 100 KHz Sampling Rate)

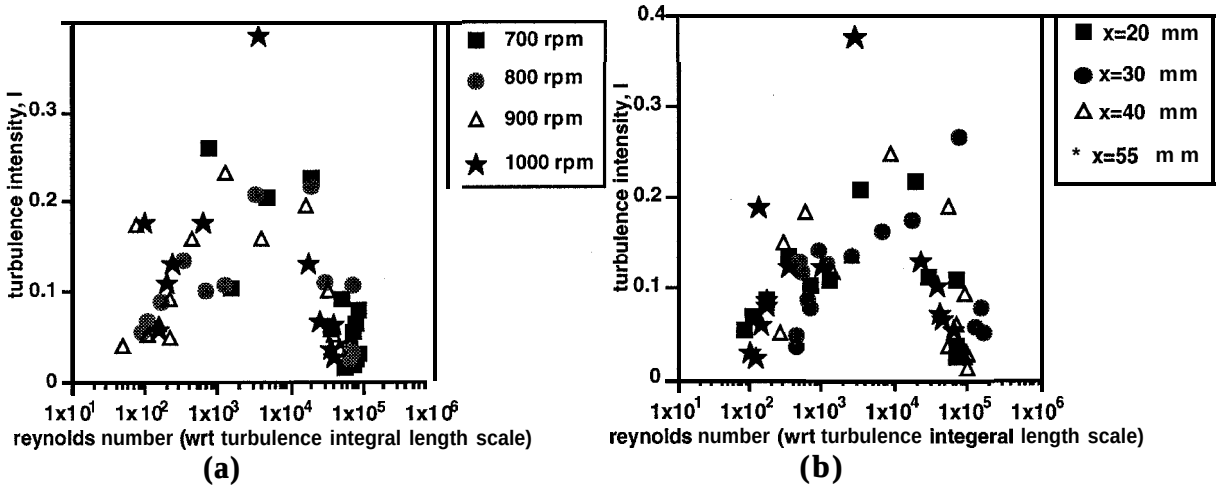


Figure 4. (a) Turbulence Intensity from Velocity Fluctuations vs Reynolds No., Changing RPM. (Expansion Stroke, 100 KHz Sampling Rate, x = 20 mm) (b) Turbulence Intensity from Velocity Fluctuations vs Reynolds No., Changing Location. (Expansion Stroke, 800 RPM, 100 KHz Samp. Rate, 15 Kpts/cycle)

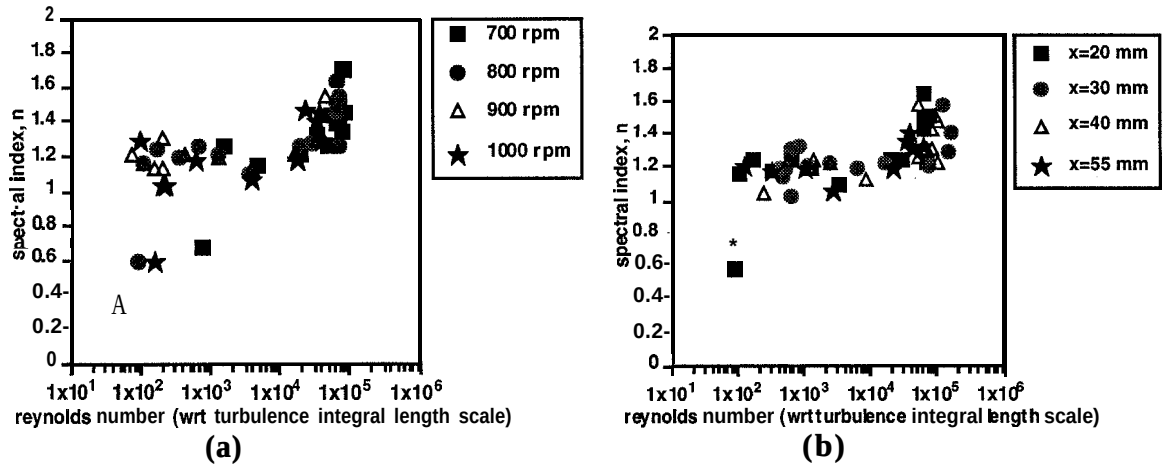


Figure 5. (a) Power Spectral Index from Velocity Fluctuations vs Reynolds No., Changing RPM. (Expansion Stroke, 100 KHz Samp. Rate, x = 20 mm) (b) Power Spectral Index from Velocity Fluctuations vs Reynolds No., Changing Location. (Expansion Stroke, 800 RPM, 100 KHz Samp. Rate, 15 Kpts/cycle)

Laser Pencil Beam Based Techniques for Visualization and Analysis of Interfaces Between Media

Grigory Adamovsky
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio

Sammie Giles, Jr.
University of Toledo
Toledo, Ohio

Abstract

Traditional optical methods that include interferometry, Schlieren, and shadowgraphy have been used successfully for visualization and evaluation of various media. Aerodynamics and hydrodynamics are major fields where these methods have been applied. However, these methods have such major drawbacks as a relatively low power density and suppression of the secondary order phenomena. A novel method introduced at NASA Lewis Research Center minimizes disadvantages of the "classical" methods. The method involves a narrow pencil-like beam that penetrates a medium of interest.

The paper describes the laser pencil beam flow visualization methods in detail. Various system configurations are presented. The paper also discusses interfaces between media in general terms and provides examples of interfaces.

1. Introduction

Flow visualization methods have been used for years to evaluate parameters of various flows. Descriptions of these methods, which utilize interferometers, Schlieren, and shadowgraphs, can be found in a number of references [1-3]. The methods are based on the accumulative variations either in the phase of an optical beam propagating through a medium or in its derivatives. With an interferometer, the changes in the density of the medium along the beam propagation are evaluated. On the other hand, Schlieren methods are employed to detect changes in the first derivative of density distribution. Finally, shadowgraphs are used to visualize the second derivatives of the density. It is obvious that for very slow variations in the density interferometric systems are the most appropriate. Shadowgraphs are the best for visualization of flows with very rapid changes of density. For instance, shocks, which are characteristic of supersonic flows, are observed the best by shadowgraphs. These three basically different flow visualization methods have one thing in common. They all usually require a large diameter collimated beam. This requirement leads to a relatively low spatial power density and consequently low contrast. To compensate for a low contrast the power of the light source has to be increased. Another problem arises from the fact that a large diameter collimated beam suppresses the second order phenomena that accompany the wave propagation. An example of such phenomena is light diffraction from inhomogeneities.

Inhomogeneities in the medium cause the beam to deviate from its original path. The deviation is often accompanied by such optical phenomena as distraction and interference. The pattern of transmitted light carries information about inhomogeneities in the medium. Furthermore, information about the medium itself may be extracted from the pattern. The issue is closely related to inhomogeneities. Interfaces can be abrupt or distributed. An interface maybe described mathematically by a derivative of a particular physical property with respect to space. An abrupt interface is the boundary between two distinct fluids. The two fluids then exhibit different physical properties across the interface. The spatial derivatives of the physical properties, in this case, have singularities at the interface. An air bubble in water forms an abrupt interface. The physical properties across a distributed interface change gradually and the corresponding spatial derivatives are finite. The type of interface determines the dominant optical phenomenon that accompanies the beam propagation through the medium. Thus, evaluation of the resultant optical pattern may help to retrieve information about the interface itself.

2. Theory and Analysis

Propagation of electromagnetic waves is governed by the four coupled Maxwell's equations. Solutions of these equations that also satisfy appropriate boundary conditions describe analytically the phenomenon of wave propagation. For nonmagnetic dielectric spatially inhomogeneous media with no electric charges or currents two second order partial differential equations result. These equations describe the propagation phenomena in terms of either electric or magnetic fields. Just one of these equations is sufficient for analytical and computational purposes. The equation presented in terms of the electric field vector in the Cartesian coordinate system is given by:

$$\nabla^2\bar{E}+\bar{\nabla}\bar{\nabla}\left[\bar{E}\cdot\frac{\bar{\nabla}\varepsilon}{\varepsilon}\right]=\varepsilon\mu\frac{\partial^2\bar{E}}{\partial t^2}$$

The second term on the left side of this equation reflects a presence of inhomogeneities in the medium. If& does not vary in space the equation assumes the form of the wave equation. Spatial variations in the permeability ε could be presented mathematically and inserted in the above equation. That could lead to further simplifications and to more manageable partial differential equations. The incident electromagnetic field also plays a role in the simplification process. For instance, assume that the incident field is a plane wave that propagates in the Y direction. We can decompose this incident plane wave into two mutually orthogonal linearly polarized waves. Select these waves in such manner that the electric field of one of them is transverse to the plane of incidence (the XY plane). This is a TE wave with the electric field vector in the Z direction and the magnetic field vector in the X direction. The other wave, with the magnetic field vector being transverse to the XY plane, is a TM wave with the electric field vector in the -X direction and the magnetic field vector in the Z direction. Assume also that the medium has one-dimensional variations in the permittivity that occurs in the X direction. This describes a case when the one-dimensional inhomogeneities occur in a direction normal to the direction of the wave propagation. Then the following partial differential equation for the magnetic field vector of the TM wave results:

$$\left(\nabla_{xy}^2-\frac{1}{\varepsilon}\frac{d\varepsilon}{dx}\frac{\partial}{\partial x}+k_0^2\varepsilon\right)H_z^{TM}=\left(\nabla_{xy}^2-\frac{d(\log\varepsilon)}{dx}\frac{\partial}{\partial x}+k_0^2\varepsilon\right)H_z^{TM}=0$$

A similar expression could be derived for the electric vector of the TM wave.

In general, an inhomogeneous medium of interest maybe treated as a combination of homogeneous media with transition regions between them. Those transition regions are interfaces. The interfaces could be either abrupt or distributed. In case of an abrupt interface between two homogeneous media solutions of the wave equation are sought separately in each of the media. The solutions are then coupled on the interface using matching conditions for normal and tangential components for the electric field and magnetic fields. For a distributed interface the process is more complicated. Depending on the "profile" and geometry of the interlace and a mutual orientation of the interface and the incident EM field, the governing equations could take various degrees of simplification [4,5]. In some simple cases an exact analytical solutions may be obtained. However, in the most cases the equations cannot be solved exactly. Even when the solution can be written in an algebraic or closed form, the practical impact of having such solution is minimal unless there are numerical values associated with the solution.

The arguments presented in the last paragraph emphasize the importance of modeling and numerical computations of the electromagnetic wave propagation through media with interfaces. The in depth analysis of wave propagation through inhomogeneous media, various modeling approaches, and numerical methods may be found in the literature [6, 7]. One of the reported methods includes computing the passage of a Gaussian beam through an inhomogeneity and then propagating the resultant wavefront into the far field using the Fresnel diffraction equation [8]. Such numerical methods as the FD-TD [9] integral [10], and a ray optics approximation [11] have been proposed to compute propagation through the inhomogeneity with a shock-like profile of the refractive index. Computational results of a Gaussian beam propagating through an inhomogeneity of a cylindrical shape are shown in Fig. 1. The results were obtained using the finite difference-time domain (FD-TD) method. Among other candidate methods an anomalous diffraction approximation could also be used if variations in the refractive index across the interface are very small [12]. The use of laser pencil beams with the Gaussian profile has certain advantages. The obvious one is very

high power density that a laser has. Another advantage comes from the fact that the laser power density is confined, especially within the Rayleigh zone, to a profile described by a Gaussian. When such incident beam strikes an interface or inhomogeneity, the resultant diffracted and scattered waves propagate beyond its spatial domain defined by a Gaussian. Separation of the scattered field from the total one results and a more detailed structure of the pattern can be observed and evaluated *[3].

3. Experiments and Applications

A simple setup has been constructed to demonstrate the principle of pencil beam propagation through a medium with various interfaces. It consisted of a laser beam striking the interface at a grazing incidence. A conventional shadowgraph could be added for visualizing the interface. Descriptions of such experimental setup may be found in the References sited above. Schematic of the setup is given in Fig. 2. It depicts a Gaussian beam striking an interface between two media at a grazing incidence. This configuration permitted observation of various phenomena associated with light propagation through normal and bow shocks. In both cases the beam splitting and spreading were seen. Large angle scattering on a bow shock was also observed. An example of a laser beam splitting by a bow shock is shown in Fig. 3. The bow shock was obtained in a supersonic tunnel by placing a blunt body in the flow. The laser beam was sent through a transparent section of a tunnel normal to the flow direction. The picture shows the pattern seen by a camera (Fig. 3a) and the relative distribution of light intensity (Fig. 3b). An example shown in Fig. 4 displays intensity distributions across two patterns resulted from a laser beam propagating through a water chamber with a thermal gradient. One pattern is generated by a negative temperature gradient inside the chamber (19.9 ‘Cat the top and 59.9 “Cat the bottom). The other pattern corresponds to a case with a positive gradient (56.6 “C at the top and 20.3 °C at the bottom).

In addition to flow evaluation a similar system has been used to study abrupt interfaces produced by air bubbles in water. The bubble is generated in test chamber 14 filled with distilled water. The laser pencil beam penetrates the chamber through transparent walls, grazes the air-water interface, and upon the exit from the chamber produces a pattern. The water temperature is controlled using two thermostatic circulators. One circulator maintains a constant temperature at the top of the surface inside the chamber and the other at the bottom. Thus, a constant temperature is maintained inside the chamber or a temperature gradient can be introduced. Different patterns have been observed under different thermal conditions inside the chamber.

The system configuration used to analyze experimentally the mentioned above phenomena has utilized a stationary laser beam. In order to enhance capabilities of the pencil beam method a scanning mechanism may be added to allow the laser beam to change its position in space. Various scanning techniques have been reported. Various mechanical, electro- and acousto-optical scanners 15 as well as their applications to flow visualization [16] have already been discussed.

Another embodiment shown in Fig. 5 utilizes a spectral scanner. Major components in spectral scanners are a tunable light source and an optical dispersive element. Examples of the dispersive elements are dispersion prism and diffraction gratings. These components are installed in the transmitting part of the sensing system, which also includes a controller. The tunable source generates a narrow beam of light (i.e., a pencil beam) whose optical frequency changes in time in a prescribed manner. It is a known fact that the direction of alight beam after interaction with a fixed dispersive element depends on the optical frequency of the light. This space-frequency or space-wavelength scanning generates a “rainbow” with the difference that each “color” appears in its place at a given time. The dispersive element is placed in the focal plane of a collimating lens. This arrangement converts a cone of light of different colors into a series of mutually parallel pencils of light of corresponding colors.

The pencil beam may contain several individual beams with different optical frequencies (wavelengths). The wavelengths may be cooperatively or independently changed in a time-prescribed manner. Thus, the fixed dispersive element produces several “rainbows”. The term “fixed” is used to indicate spectral scanning by a stationary dispersive element in contrast with other embodiments in which the light beam is physically translated or otherwise manipulated.

The optical dispersive element may also replace a reflecting mirror or prism in the angular scanner. Such a hybrid system combines a spectral scanner with an angular electromechanical one. The hybrid scanner may also employ a plurality of optical beams with different wavelengths. These optical beams strike the angular scanner that has its reflecting element, mirror or prism, replaced by the dispersive element. A multiplicity of spectral cones or “rainbows” will result.

Acknowledgment

The authors would like to acknowledge the technical support that Mr. B. M. Floyd (Cortez III Service Corp.) has provided on various stages of this project.

References

[1] A. H. Shapiro, The ***Dynamics and Thermodynamics of Compressible Fluid Flows***, Ronald Press, New York, 1953.
[2] W. Merzkirch, ***Flow Visualization,*** Academic Press, Orlando, FL, 1987.
[3] L. A. Vasil'ev, ***Schlieren Methods,*** Israel Program for Scientific Translations Ltd., 1971, (Translation from Russian of *Tenevye Methody*, Izdatel'stvo "Nauka", Moskva, 1968), Distributed by Keter Inc., New York.
[4] G. Tyras, ***Radiation and Propagation of Electromagnetic Waves,*** Academic Press, New York, 1969.
[5] A. Ishimary, ***Electromagnetic Wave Propagation, Radiation,*** and Scattering, Prentice Hall, New Jersey, 1991.
[6] J. R. Wait, ***Electromagnetic Wines in Stratified Media***, Franklin Book Company, Elkins Park, PA, 1993 (Authorized reprint from revised edition, Pergamon Press, Oxford, England, 1972).
[7] W. C. Chew, ***Wines and Fields in Inhomogeneous Media,*** Oxford Univ. Press, 1996.
[8] J. Panda and G. Adamovsky, "An Experimental Investigation Of Laser Light Scattering by Shock Waves," 33rd AIAA Aerospace Science Meeting and Exhibit (January 9-12, 1995, Reno, Nevada), Paper # AIAA 95-0518.
[9] G. Adamovsky and N. Ida, "Laser Beam Propagation through Inhomogeneous Media with Shock-Like Profiles: Modeling and Computing", in ***Optical Technology in Fluid, Thermal, and Combustion Flows III***, S. S. Cha, J. D. Trolinger, and M. Kawahashi, eds., Proc. SPIE 3172, pp. 530-539 (1997).
10 S. I. Hariharan and D. K. Johnson, "Transmission of Light Waves Through Normal Shocks," Appl. Opt. 34, pp. 7752-7758, 1995.
[11] J. Panda and G. Adamovsky, "Laser Light Scattering by Shock Waves," Phys. Fluids 7, pp. 2271-2279, 1995.
12 M. Kerker, ***The Scattering of Light and Other Electromagnetic Radiation,*** Academic Press, New York, 1969.
13 P. Langlois and R. A.Lessard, and A. Boivin, "Real-Time Curvature Radii Measurements Using Diffraction Edge Waves," Appl. Opt. 24,8, pp. 1107-1112 (1985).
14 N. Rashidnia, 'Instabilities Around a Bubble Due to Combined Maragoni and Buoyancy Effects", presented at 31st National Heat Transfer Conference, August 5-9, ' 1996 Houston, Texas; published in *AIChE Symposium Series, No. 310,* vol. ***92, American*** Institute of Chemical Engineers, New York, pp. 110-118,1996.
[15] K. Gustafsson and B Hok, "Light Scanning Techniques - a Critical Review", International Journal of Optoelectronics, 4, (l), pp. 33-52 (1989).
[16] G. Adamovsky and D. K. Johnson, "Optical Techniques for Shock Visualization and Detection," in ***Optical Techniques in Fluid, Thermal, and Combustion How, S. S.*** Cha and J. D. Trolinger, eds., Proc. SPIE 2546, pp. 348-357, 1995.

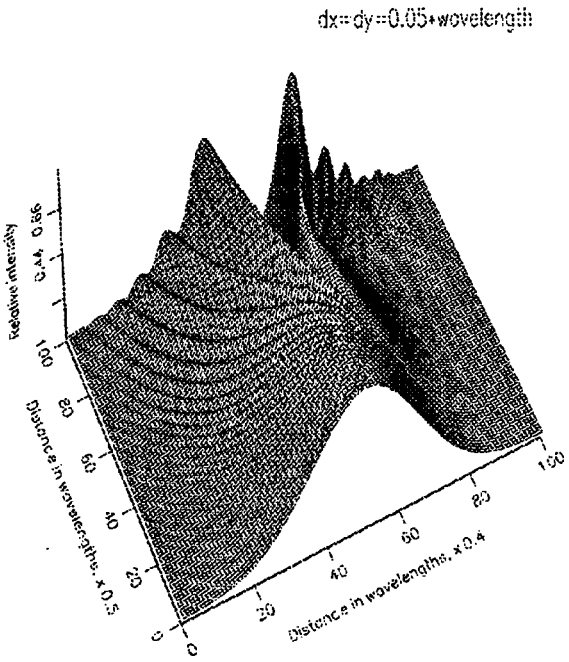


Fig. 1: Propagation of a Gaussian beam through inhomogeneous media under a grazing incidence (computational results),

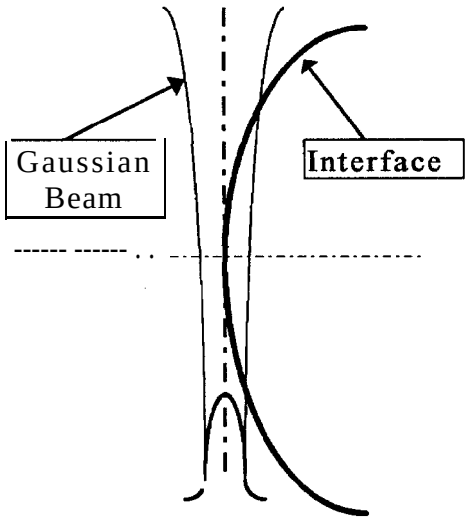
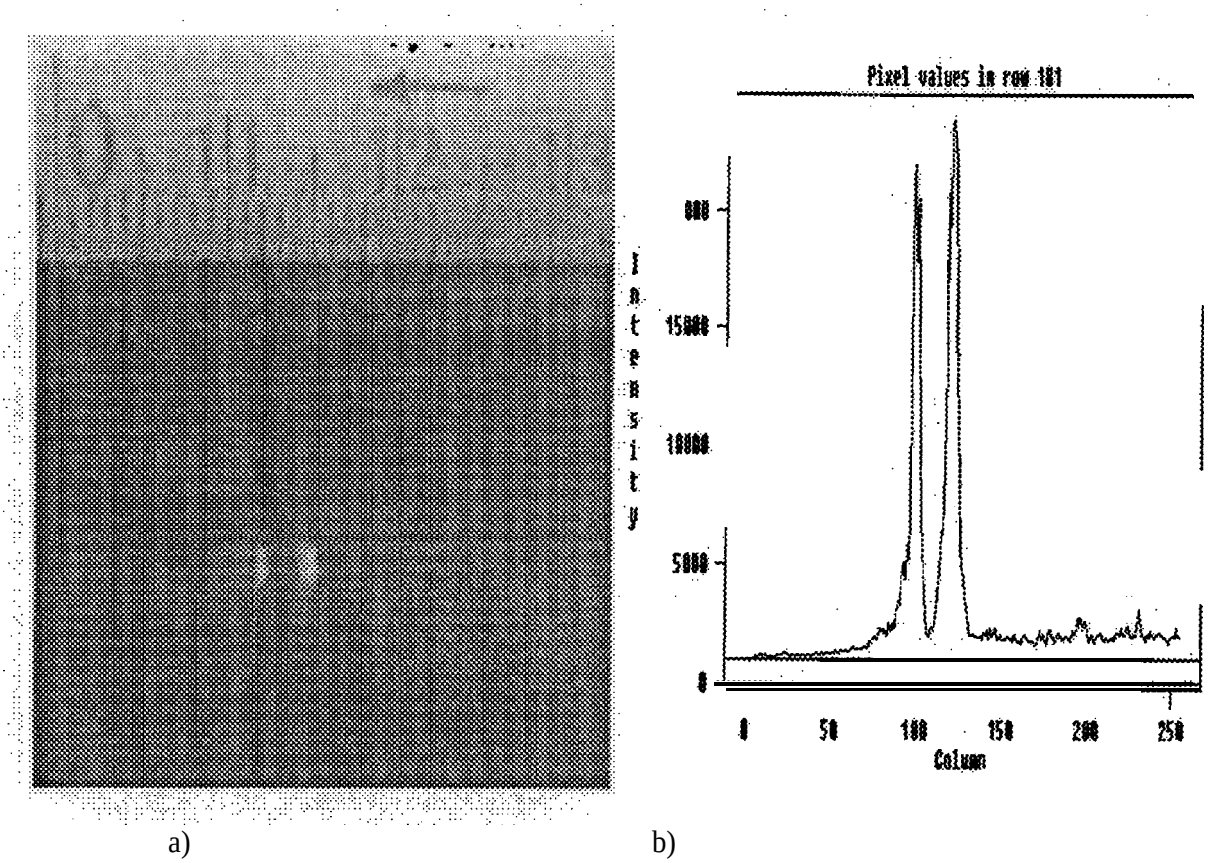


Fig. 2: Schematic diagram of a Gaussian beam striking an interface at a grazing incidence.



a) b)

Fig. *3:* Laser beam splitting by a bow shock:
a) pattern observed by a camera;
b) intensity distribution across the pattern.

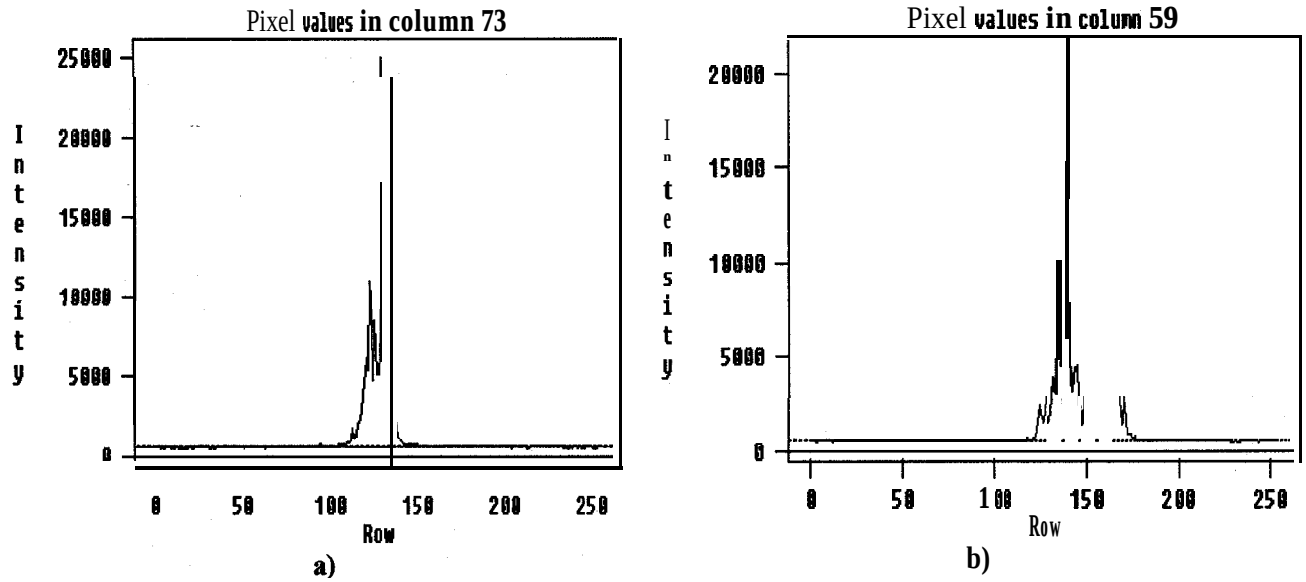


Fig. *4:* Diffraction patter&of a Gaussian beam propagated through diffused interfaces generated by two different temperature gradients:
a) negative temperature gradient (top temperature: 19.9 ‘C, bottom temperature: 59.9 ‘C);
b) positive temperature gradient (top temperature: 59.6 ‘C, bottom temperature: 20.3 “C).

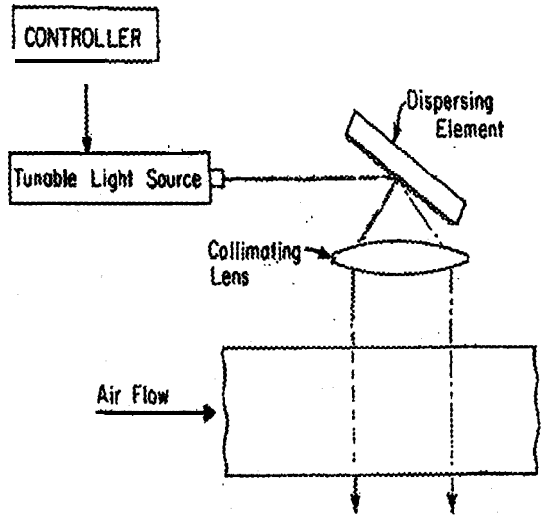


Fig. *5:* Diagram of a spectral scanner.

ROBUST CONTROL DESIGN FOR FLEXIBLE STRUCTURES WITH IDENTIFIED PARAMETER UNCERTAINTY

Samir S. Ahmad, Jiann-Shiun Lew, and Lee H. Keel
NASA Center for Automated Space Science
Tennessee State University
Nashville, TN 37203-3401

ABSTRACT

This paper presents a robust control synthesis for large flexible structures with identified parameter uncertainty. The identified parameters are obtained by applying system identification algorithms to analyze the time-domain response data. The uncertainty of the identified parameters can be from environmental changes, equipment noise, and unmodeled dynamics, etc. The system with parameter uncertainty is modeled as an interval transfer function, which is a model set of bounded transfer function coefficients. The robust control algorithm is based on a μ-synthesis control technique for the interval transfer function. The effect of the identified parameter uncertainty on the performance of the designed controller is discussed. A 20-bay truss structure is used for the study.

KEYWORDS: Robust control design, System identification, Identified parameter uncertainty.

INTRODUCTION

In the last decade, the robust control design of large space structures has received special attention in control community. Various approaches have been proposed [1-3]. In this paper, we address the problem of vibration suppression of large flexible structures with identified parameter uncertainty.

The first step of the robust control design is to model the system with uncertainty. From many sets of tested data under various uncertain circumstances, the system identification algorithms [4] are used to process these sets of time-domain responses data to obtain many identified models. Each identified model is expressed as a transfer function to represent one set of data. Each transfer function coefficient has variation. Then an interval transfer function is constructed to tightly enclose all the identified models as the bounded transfer function coefficients.

The second step is to design a robust controller that robustly stabilizes the entire interval transfer function and provide the satisfactory performance. For the robust control design, a μ-framework is formed for the identified interval transfer function. This μ-synthesis technique, coupled with the interval system framework, has an obvious advantage oven a conventional one. This interval system enables us to reduce the structural uncertainty block to 4 [3] instead of 2n for an nth order interval plant . This significantly reduces the complexity of the synthesis and the computational effort of design and analysis.

The effect of the identified parameter uncertainty on the performance of the designed controller is discussed. In this paper, the interval system techniques are used for the analysis of the open

loop and closed-loop interval transfer functions [3]. A 20-bay truss structure with identified parameter uncertainty will be used to demonstrate and verify the algorithm.

MODELING OF TRUSS STRUCTURE WITH UNCERTAINTY

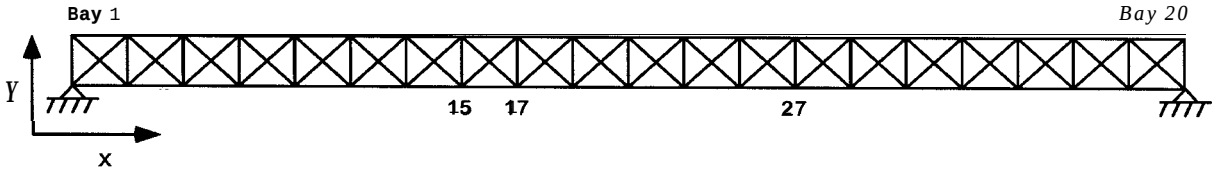


Figure 1: Twenty bay truss.

The finite element model [5] of a 20-bay truss structure as shown in Figure 1 is the system used for the study. This structure is fixed to the ground at both ends. The structure has 101 aluminum tube elements with cross-sectional area of 5.97×10^{-5}, and each nondiagonal element is one-meter long. The mass of the fastener, which is used to connect elements, at each node is 0.60374Kg. Two 10Kg added masses are located at nodes 15 and 27, respectively. The natural frequencies of the healthy structure of the first three modes are 4.88, 14.30, and 38.11 Hz, respectively. In this paper, the finite element model of these first three modes with the assumed damping is used. The damping ratios of the first, second, and third modes are 1.2%, 1%, and 0.6%, respectively. In this study, an actuator located at node 17 is used to excite the vibration in the y-direction and an accelerometer at this node is used to measure the vibration. The time domain response data are simulated by the state space model

$$x(k+1) = \mathrm{Ax(k)} + Bu(k) + \mathrm{v(k)} \tag{1}$$

$$y(k) = Cx(k) + Du(k) + w(k) \tag{2}$$

where $[A, B, C, D]$ represents the discrete state space model of the first three modes, u is the actuator input, y is the acceleration output. The variables v and w are process and measurement noise, respectively, and are assumed Gaussian, zero mean signals, The process noise and measurement noise are used to represent system uncertainty. The identified parameters are obtained by applying the system identification software, which is developed in NASA Langley Research Center [4], to analyze the random input response data. The results based on two different noise levels are presented. For the low noise case, the standard deviation of the measurement noise is 2% of the magnitude of the noise free output. The standard deviation of the random process noise is 1% of the magnitude of the corresponding state. The high noise level is 5 times higher than the low noise for either process or measurement noise. For each noise level, there are 50 sets of input/output response data to represent the system under various uncertainties. The l^{th} identified model is obtained from the l^{th} set of data as

$$g^l(s) = \frac{b^l_6 s^6 + b^l_5 s^5 + b^l_4 s^4 + b^l_3 s^3 + b^l_2 s^2}{a^l_6 s^6 + a^l_5 s^5 + a^l_4 s^4 + a^l_3 s^3 + a^l_2 s^2 + a^l_1 s^1 + a^l_0} \tag{3}$$

From all the identified models, an interval transfer function is constructed by using the bounds of each transfer function coefficient as [3]

$$P(s)=\frac{[b^-_6,b^+_6]s^6+[b^-_5,b^+_5]s^5+[b^-_4,b^+_4]s^4+[b^-_3,b^+_3]s^3+[b^-_2,b^+_2]s^2}{[a^-_6,a^+_6]s^6+[a^-_5,a^+_5]s^5+[a^-_4,a^+_4]s^4+[a^-_3,a^+_3]s^3+[a^-_2,a^+_2]s^2+[a^-_1,a^+_1]s^1+[a^-_0,a^+_0]} \qquad (4)$$

Table 1 shows the bounds of the identified interval transfer functions for both low noise and high noise cases.

Table 1: Bounds of the identified interval transfer functions.

Intervals	Low Noise	High noise
$[a_5^-, a_5^+]$	[5.2937, 5.5 109]	[4.8842, 5.9703]
$[a_4^-, a_4^+]$	$[6.6349\text{x}10^4, 6.6395\text{x}10^4]$	$[6.6231\text{x}10^4, 6.6471\text{x}10^4]$
$[a_3^-, a_3^+]$	$[1.7383\text{x}10^5, 1.8273\text{x}10^5]$	$[1.5977\text{x}10^5, 1.9818\text{x}10^5]$
$[a_2^-, a_2^+]$	$[5.2462\text{x}10^8, 5.2545\text{x}10^8]$	$[5.2325\text{x}10^8, 5.2721\text{x}10^8]$
$[a_1^-, a_1^+]$	$[4.3820\text{x}10^8, 4.8088\text{x}10^8]$	$[3.8923\text{x}10^8, 5.6006\text{x}10^8]$
$[a_0^-, a_0^+]$	$[4.3516\text{x}10^{11}, 4.3632\text{x}10^{11}]$	$[4.3332\text{x}10^{11}, 4.3825\text{x}10^{11}]$
$[b_6^-, b_6^+]$	$[7.0179\text{x}10^{-2}, 7.0518\text{x}10^{-2}]$	$[6.9589\text{x}10^{-2}, 7.1292\text{x}10^{-2}]$
$[b_5^-, b_5^+]$	$[1.4925\text{x}10^{-1}, 4.0482\text{x}10^{-1}]$	$[-3.4782\text{x}10^{-1}, 8.7853\text{x}10^{-1}]$
$[b_4^-, b_4^+]$	$[3.6017\text{x}10^3, 3.6548\text{x}10^3]$	$[3.4863\text{x}10^3, 3.7510\text{x}10^3]$
$[b_3^-, b_3^+]$	$[3.4145\text{x}10^3, 8.6224\text{x}10^3]$	$[-5.3099\text{x}10^3, 2.0496\text{x}10^4]$
$[b_2^-, b_2^+]$	$[1.9123\text{x}10^7, 1.9628\text{x}10^7]$	$[1.8147\text{x}10^7, 2.0826\text{x}10^7]$

FRAMEWORK FOR ROBUST CONTROL DESIGN

This section presents how to form the structural uncertainty of the interval transfer function for the μ-synthesis. Both numerator and denominator of the interval transfer function are interval polynomials

$$\delta(s) := a_0 + a_1 s + a_2 s^2 + \cdots + a_n s^n, \qquad a_i \in [a_i^-, a_i^+], \ \forall\ i \qquad (5)$$

The four Kharitonov polynomials associated with this interval polynomial are [3],

$$\begin{aligned} K^*(s) &= a_0^- + a_1^- s + a_2^+ s^2 + a_3^+ s^3 + a_4^- s^4 + a_5^- s^5 + \cdots, \\ K^2(s) &= a_0^- + a_1^+ s + a_2^+ s^2 + a_3^- s^3 + a_4^- s^4 + a_5^+ s^5 + \cdots, \\ K^3(s) &= a_0^+ + a_1^- s + a_2^- s^2 + a_3^+ s^3 + a_4^+ s^4 + a_5^- s^5 + \cdots, \\ K^4(s) &= a_0^+ + a_1^+ s + a_2^- s^2 + a_3^- s^3 + a_4^+ s^4 + a_5^+ s^5 + \cdots \end{aligned} \qquad (6)$$

The interval polynomial can be written as [3]

$$\delta(s) = K^c(s) + \lambda_1 K^e(s) + \lambda_2 K^o(s), \qquad (\lambda_1, \lambda_2) \bullet [-1,1] \times [-1,1] \qquad (7)$$

With

$$K^c(s) = \frac{1}{4}[K^1(s) + K^2(s) + K^3(s) + K^4(s)]$$

$$K'(s) = \frac{1}{2}[K^2(s) - K'(s)], \quad K^o(s) = \frac{1}{2}[K^3(s) - K^1(s)] \qquad (8)$$

The preceding interval transfer function can be written as

$$P(s) = \{\frac{N^c(s) + AN'(S) + A, N''(s)}{D^c(s) + \lambda_3 D^e(s) + \lambda_4 D^o(s)}, \quad (\lambda_1, \lambda_2, \lambda_3, \lambda_4) \in [-1,1] \times [-1,1] \times [-1,1] \times [-1,1]\} \qquad (9)$$

In this paper, a weighting transfer function W_p is designed for loop-shaping the closed-loop transfer function for the vibration suppression. The designed feedback loop of the system with structural uncertainty is presented as the block diagram shown in Figure 2. To have suitable control force and performance under sensor noise, W_s and W_c *are* also included in the problem formulation as shown in Figure 2. All these performance specifications are accounted in the μ-framework [1].

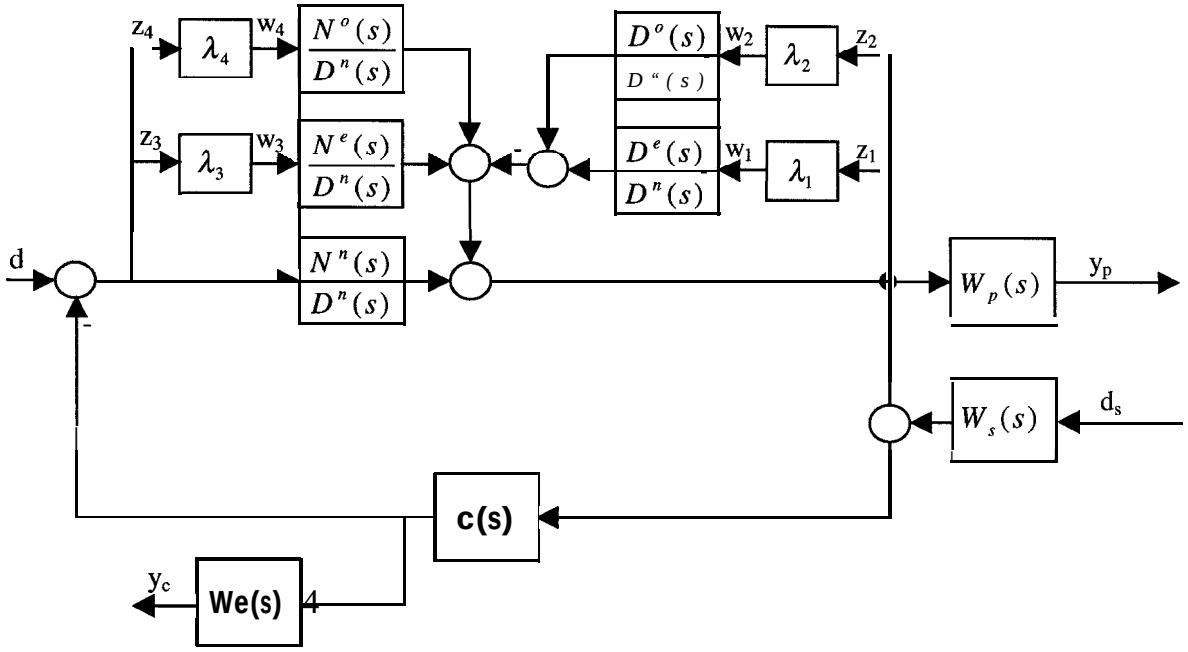


Figure 2: Diagram of feedback loop system with uncertainty.

RESULTS

Table 2: Ranges of damping ratios of identified interval models

	1st mode %	2nd mode %	3rd mode %
Low noise Max	1.307	1.096	0.649
Min	1.096	0.891	0.553
High noise Max	1.690	1.420	0.830
Min	0.832	0.537	0.381

Table 3: Minimum damping ratios of closed-loop interval systems.

	1st mode %	2nd mode %	3rd mode %
Low noise	53.506	4.556	3.850
High noise	8.257	1.456	1.350

The results of the design based on the preceding μ-framework to the identified two interval transfer functions are presented in this section. Table 2 shows the ranges of the damping ratio, which are obtained by applying interval system techniques [3], these two identified interval transfer functions. For each mode the range of the damping ratio of the high noise case is

between 4 to 5 times larger than that of the low noise case. Figures 3 depict the magnitude envelopes of the closed-loop (solid lines) and open-loop (doted line) interval systems. For the low noise case, a significant amount of damping has been achieved for every mode; this implies that the vibration suppression of all the modes under uncertainty. For the high noise case, the damping of each mode also increases to a certain level. Table 3 shows the minimum damping ratio, which represents the worst performance, of each mode of the close-loop interval transfer functions. The damping ratio of the first mode for the low noise case increases tremendously and it is more that 5 times larger than that for the high noise case. This implies that the performance of the design based on the identified interval transfer function of high noise case reduces significantly due to model uncertainty.

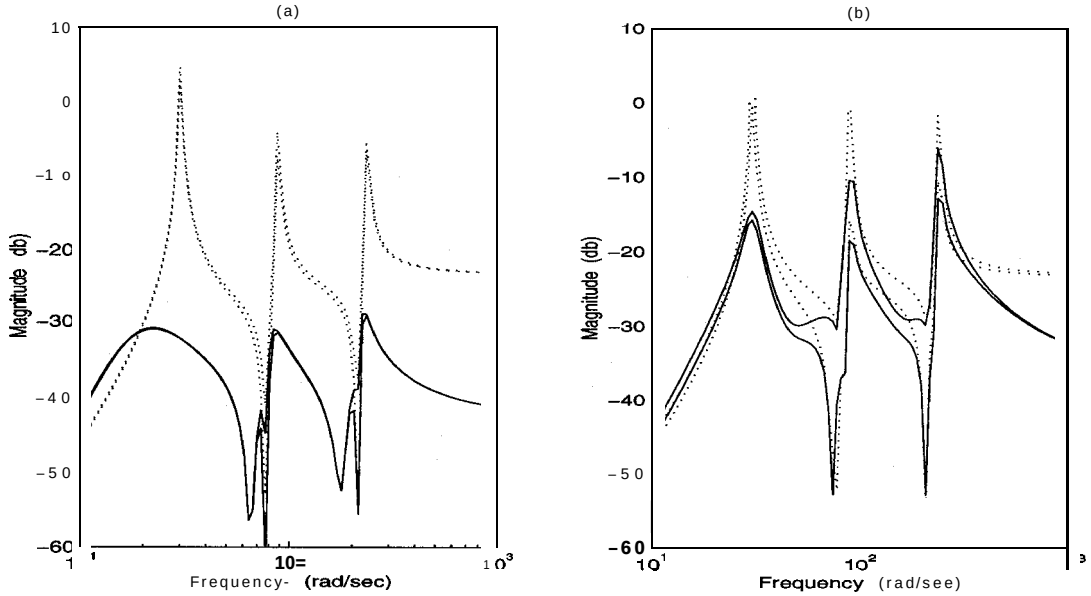


Figure 3: Magnitude envelopes of the open-loop and closed-loop interval systems: (a)low noise case, (b)high noise case.

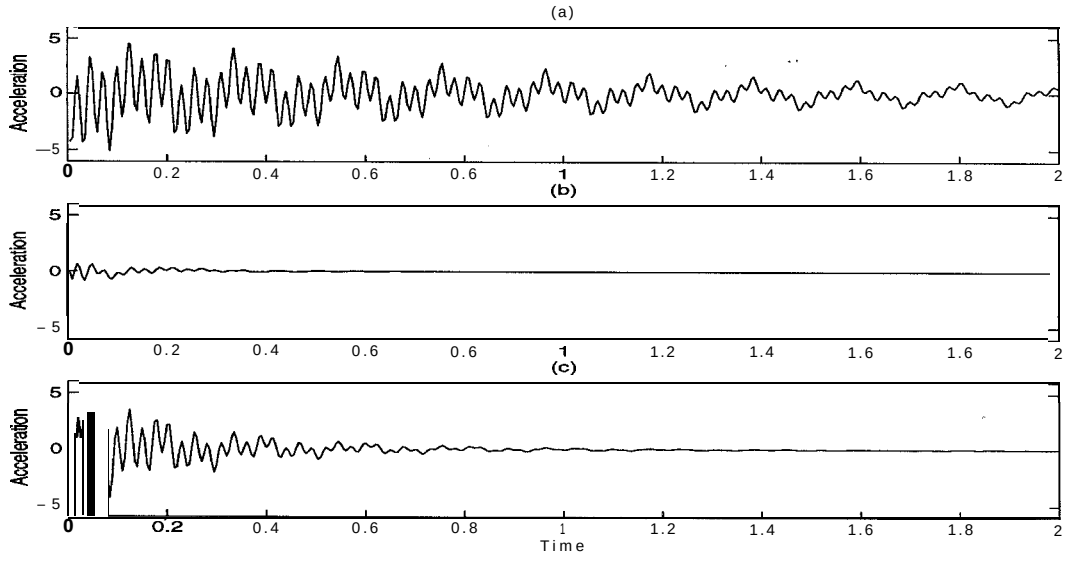


Figure 4: Impulse responses of system without uncertainty: (a)open-loop system, (b)close-loop system with low noise controller, (c) close-loop system with high noise controller.

The impulse responses of the open-loop system and the closed-loop system for the noise-free system without uncertainty are shown in Figure 4. The impulse responses of the closed-loop systems decay to zero around 0.3 and 1.2 seconds, respectively. This also verifies the performance sacrifice due to the robustness condition. The results of the impulse responses of all the identified models are similar to the ones in Figure 4.

CONCLUSIONS

A robust control synthesis is applied for the vibration suppression of large flexible structures with identified parameter uncertainty. The system identification algorithms are used to obtain identified models to generate an interval transfer function to represent a system with identified parameter uncertainty. The design is based on the loop-shaping of a μ -framework of the identified interval transfer function. Unlike the standard μ -synthesis, the size of the structural uncertainty block remains to be 4 regardless of the number of parameters. The interval polynomial techniques are used for the analysis of the designed controller. The results show that the large parameter uncertainty sacrifices the performance significantly. This algorithm of the combination of system identification and μ -synthesis in conjunction with the interval polynomial techniques provides an effect tool for robust control and analysis.

ACKNOWLEDGMENTS

This research is supported in part by NASA Grant NCC5-228. This support is greatly appreciated.

REFERENCES

1. Balas, G. J., and Doyle, J., "Robustness and Performance Tradeoffs in Control Design for Flexible Structures," *IEEE Transactions on Control Systems Technology,* Vol. 2, No. 4, 1994, pp. 352-361.

2. Lim, K. B., "Robust Control Design Framework for Substructure Models," *Journal of Guidance, Control, and Dynamics,* Vol. 19, No. 1, 1996, pp. 181-190.

3. Bhattacharyya S. P., Chapellat, H., and Keel L. H., *Robust Control: The Parametric Approach,* Prentice Hall PTR, Englewood Cliff, New Jersey, 1995.

4. Juang, J.-N., Phan, M., and Horta, L. G., "User's Guide for System/Observation/Controller Identification Toolbox," NASA TM 107566, 1992.

5. Craig, R. R., *Structural Dynamics: An Introduction to Computer Methods,* Wiley, *New* York, 1981.

98 URC084

Elements of Robust Control: With Examples

Willie Tillery* CR. Ashok kumar† A. Homaifar‡
North Carolina A&T State Univeristy
Greensboro NC 27411

Abstract

In this paper, we provide a simple procedure to understand the contemporary robust control lit erature. In particular, we focus on uncertainty types, and by examples, we discuss the evolution of H_∞ design from the Small Gain Theorem. The paper is organized in single input single output setting for better understanding of the subject.

1 Introduction:

One of the most important tasks in the analysis and design of control systems is mathematical modeling of the systems. Controllers are normally designed for a nominal model with an assumption that there are no modeling uncertainties. This basic approach was adopted in early stages to stabilize simple systems. However, such controllers may or may not be capable of stabilizing the system when modeling uncertainties are considered.. Systems with uncertainties are generally complex to characterize mathematically. In fact, with such uncertainties, an entirely different controller my be needed in the presence of uncertainties.

In this paper, we address the above uncertainties and their classifications using mathematical models. Secondly, we present the evolution of robust control design procedures. We organize the paper as follows. In section 2, we present modeling uncertainty types. In section 3, we present the H_∞ design procedure which depends on the basic small gain theorem. In section 4, we provide simple examples to understand uncertainty types and H_∞ design example. The class room example [1], namely, the sun-seeker system used in spacecraft attitude control problems is considered for illustration..

2 Classification of modelling uncertainties:

From the undergraduate course material [1], it is clear that the transfer functions G(s) and state space models [A,B,C,D] represents simple linear differential equations. These equations are obtained by linearizing nonlinear differential equations with second and higher order effects neglected. Therefore, there is an uncertainty type existing while linearizing nonlinear systems. These uncertainties are called parametric uncertainties. Besides, uncertainties that are not predictable also exists. For inst ante, in automotive cruise control problem,wind gusts

*Senior, Dept. of Electrical Engineering
†Postdoctoral Research Associate, Center for Aerospace Research
‡Associate Professor, Dept. of Electrical Engineering

ranging from 15mph to 80mph, is known to introduce errors in cruise speed. Likewise, we experience turbulence that is neither measurable precisely not predictable during flight cruise. These uncertainties are called exogenous inputs which can not be accounted exactly in design model. Other than the exogenous inputs, computational time delay by digital computers, flexibility effects of wing and fuselage in aircraft, solar panel vibrations in spacecraft etc., are often treated as unmodelled dynamics.

2.1 Modelling uncertain ies by mat hemat its:

In this paper, we restrict our discussion to linear time invariant continuous time state space and transfer function models. For a state vector x(t) $\in R^n$ (meaning, time dependent real numbers in a column vector with n-elements), the state-space model is given by

$$\dot{x} = Ax + bu \qquad x(t = 0) = x_0 \tag{1}$$

where $A \in R^{n \times n}$, $b \in R^{"}$ are time invariant matrices. u(t) is the scalar control input used for controlling an output variable $y(t) = Cx$ (t), which is a linear combination of state vector elements. Here C is a time invariant row vector with n elements. Equation (1) is a typical mathematical model representing a physical system. Such systems in transfer function setting with zero initial conditions is given by

$$\frac{y(s)}{u(s)} = G_0(s) = C(sI - A)^{-1}b \tag{2}$$

Within this setting we now categorize modelling uncertainties.

2.2 Exogenous Inputs:

Suppose the system experiences disturbance $w(t)$ input at input node, then equation (1) modifies to:

$$\dot{x} = Ax + bu + \Gamma w \tag{3}$$

where Γ is an n-vector provided by modelling experts. If w(t) appears at output, then $y(t)$ is written as

$$y(t) = Cx + Dw \tag{4}$$

Note that the associated transfer function with D = O, now changes to

$$G(s) = Go(s) + \Delta G(s) \tag{5}$$

where $\Delta G(s) = \Gamma(sI - A)^{-1} b \frac{w(s)}{u(s)}$, is the modelling uncertainty.

2.3 Real Parametric Uncertain ies

In mechanical, electrical circuits, electro-mechanical systems etc., mass, inertia, stiffness resistors, capacitors and so on, contribute to the matrices A and b. In reality, these parameters significantly change either due to wear ant tear, or due to the linearization procedure discussed in the above section. Therefore, the error matrices ΔA and Δb are introduced. The system with parametric uncertainties is given by

$$\dot{x} = (A + \Delta A)x + (b + \Delta b)u$$
$$y = (C + \Delta C)x \quad (6)$$

Note that AA, $\Delta b, \Delta C$ can be either time varying or time invariant. Here we assume that they are time invariant. In this setting, we also observe that the basic properties such as controllability and observability with $(A + AA, b + \Delta b, C + \Delta C)$ are now different.

2.4 Unmodelled Dynamics

In exogenous inputs and real parametric uncertainties, one important assumption that we noticed is that x(t) $\in R^n$ is unaltered. We also observe that the resulting equations are ordinary differential equations. In some cases, the dynamics is governed by the partial differential equations. The vibrational problems in mechanical systems are typical examples. In such problems, the elements in state vector $x(t)$ are denumerably large, with components $x_1 \cdots x_\infty$. Suppose we assume only three coordinates xl, x_2, x_3, for design of a control system. Then the dynamics related to $x_4 \cdot$" . x_∞ is called the unmodelled dynamics.

2.5 Representation of uncertainties for design:

Interestingly, it has been shown that majority of the control problems can be put under two categories, namely, the additive and multiplicative uncertainties.

2.5.1 Additive uncertainty:

Given nominal plant model Go(s) and a stable transfer function r(s), the class $\mathcal{A}(G_0, r)$ is the set of all perturbed plants G(s) such that

$\mathcal{A}1$: Go(s) and G(s) have the same number of open right half plane poles and none on the $j\omega$-axis.

$\mathcal{A}2$: $|G_0(j\omega) - G(j\omega)| < |r(j\omega|$ 'v' w

2.5.2 Multiplicative uncertain y:

Given nominal plant model Go(s) and a stable transfer function r(s), the class $M(G., r)$ is the set of all perturbed plants G(s) such that

Ml: Same as $\mathcal{A}1$

$\mathcal{M}2$: $G(s) = G_0(s)[1 + l(s)]$ and $|l(j\omega| < |r(j\omega)| \forall$ w. Here, we observe that the error enters as a kind of percentage error and $l(s) = [\frac{G(s)}{G_0(s)} - 1]$.

In both additive and multiplicative uncertainties, note that the perturbed model G(s) is well defined with respect to r(s) and $l(s)$. Therefore, the problem statement is stated as below: Given a *nominal system Go(s) and a perturbed system G(s) or uncertainty profile r(s)* or l(s), *is there a controller C(s) robustly stabilizing the perturbed model ?.*

3 H_∞ theory:

In this section we present the origin of robust control. Within the robust control context, the following three problems arise

1. It maintains closed loop stability even with plant uncertainties

2. It maintains closed loop performance even with plant uncertainties

3. It maintains stability over large uncertainty profile and performance over a smaller uncertainty profile

Within this setting, H_∞ control theory allows one to design robustly stabilizing controller (Problem 1) for any of the following three types of plant uncertainty. (i) Additive, (ii) Multiplicative and (iii) Stable Factors (not discussed here). Given an unity feedback configuration with perturbed plant G(s) and controller C(s), the fundamental result from small gain *theorem* is stated as below:

Theorem 1: Given GO(S), stable r(s), and a controller C(s) which stabilizes GO(s), then C(s) stabilizes all the perturbed systems G(s) in a set $M\ (GO, r)$ if and only if

$$\left|\frac{G(j\omega)C(j\omega)}{1+G(j\omega)C(j\omega)}\right| \leq \frac{1}{|r(j\omega)|} \tag{7}$$

We shall now recast the above theorem in design setting:

3.1 Design Objective:

Choose C(s) such that the magnitude of the complementary sensitivity function $T = \frac{G(s)C(s)}{1+G(s)C(s)}$ is always less than the reciprocal uncertainty r(s).

3.2 H_∞ norm:

If $f(s)$ is a stable transfer function, its H_∞ norm denoted by $||f||_\infty$ is defined as

$$||f||_\infty = \sup_\omega |f(j\omega)| \tag{8}$$

Note that in single input setting, $||f||_\infty$ is the peak of the magnitude plot for f(s). If the peak does not exits, and if it reaches some constant value $\alpha(\omega_0)$ for all $\omega_0 \geq$ w, then the maximum does not exist. In such cases, $\sup_\omega |f(j\omega)|$ assumes the constant value $\alpha(\omega_0)$.

Our objective is now to demonstrate how H_∞ is used in single input settings. This exhaustive literature in multiinput multioutput setting is overwhelmingly simplified in reference [2] for understanding purpose.

4 Numerical example:

We consider the sun-seeker control systems to illustrate uncertainty types. Here, we assume the plant G:(s) $= \frac{2500}{s(s+25)}$ and a static controller C(s) = K. As we see from the open loop dynamics, the poles –25 and O are sufficiently separated to each other. Thus, the design model $G_0'(s)$ can be simplified to $Go(s) = \frac{2500}{s}$. In which case, the uncertainty $\Delta G_u(s) = Go(s) - G_0'(s)$ due to unmodelled dynamics is given by

$$\Delta G_u(s) = \frac{2500(s+24)}{s(s+25)}$$

In order to illustrate uncertainty due to exogenous inputs, the state space model for GO (s) = $C(sI - A)^{-1}b$ may be computed as below:

$$\begin{array}{rcl} \dot{x} & = & 2500u \\ y & = & x \\ u & = & -Kx \end{array} \quad \Rightarrow \quad \begin{array}{l} A = 0 \\ b = \mathbf{2500} \text{ and} \\ C = 1 \end{array}$$

If the model above expereince an exponentially decaying disturbance w = e^{-30t}, then the state equation becomes

$$\dot{x} = 2500u + w$$

The perturbations $\Delta G_w(s) = C(sI - A)^{-1}\frac{w(s)}{u(s)}$, given by:

$$\Delta G_w(s) = \frac{1}{s(s + 30)u(s)}$$

is the uncertainty due to exogenous input. Similarly, for parametric uncertainty, we may assume Δb = – 100 and the corresponding transfer function due to parametric uncertainty is $\Delta G_p(s) = C(sI - A)^{-1}\Delta b$, given by:

$$\Delta G_p(s) = \frac{-100}{s}$$

Given these uncertainties, our objective is to find a stabilizing controller K such that the perturbed models $G_u(s)$ = Go(s)+ $\Delta G_u(s)$, $G_w(s)$ = Go(s)+ $\Delta G_w(s)$ and $G_p(s)$ = Go(s)+ $\Delta G_p(s)$ are all stable in the presence of uncertainties $\Delta G_u(s)$, ΔG_w (s) and $\Delta G_p(s)$. For this purpose, we need to choose a stable transfer function $r(s)$ whose reciprocal magnitude plot is greater than $|KG_u(j\omega)|$, $|KG_w(j\omega)|$, and $|KG_p(j\omega)|$. That is, we use theorem 1 to find the gain K, which satisfies the condition

$$\left|\frac{KG_u(j\omega)}{1 + KG_u(j\omega)}\right| \left|\frac{KG_w(j\omega)}{1 + KG_w(j\omega)}\right| \left|\frac{KG_p(j\omega)}{1 + KG_p(j\omega)}\right| < \frac{1}{|r(j\omega)|}$$

Here two problem arise. First, selection of a *stable* $r(s)$ overbounding the uncertainties $\Delta G_{u,w,p}$ is difficult. Secondly, selecting a controller K minimzing the peak of the magnitude of the complementary transfer T(s) is also a difficult problem.

5 Conclusions:

The purpose of this paper is to learn the contemporary robust control literature in single input single output settings. The problems are well exposed, and H_∞ theory is understood to be a tool for robustly stabilizing a given uncertain system. In multiinput multioutput setting, reference [3] will be used for further understanding of the subject. Since H_∞ is an optimization problem, some of the issues addressed to seek an optimal controller may be effectively solved by using Genetic Algorithm. For this purpose, the class room project [4] will be explored in H_∞ control setting.

6 References:

1. *Linear Systems and Control*, class notes, Department of Electrical Engineering, NC A&T State University

2. Huibert Kwakernaak, *Robust Control and* H_∞ Optimization- Tutorial *Paper*, Automatica, 1993

3. John C. Doyle, Keith Glover, Promod P. Khargonekar, Bruce A. Francis, *State-Space Solutions to Standard* H_2 *and* H_∞ *Control Problems*, IEEE Transactions on Automatic Control, 1989

4. *Genetic Algorithm*, class notes, Department of Electrical Engineering, NC A&T State University

ADVANCEMENTS ON THE CONTROL OF THE LONGITUDINAL MOTTION OF THE SPACE SHUTTLE DURING RE-ENTRY

Roland Reynard Marsh Marwan Bikdash, Ph.D. Abdollah Homaifar, Ph.D.
NASA CAR, Center for Aerospace Research
Department of Electrical Engineering
North Carolina A&T State University
Email:
roland@ncat.edu
bikdash@ncat.edu
homaifar@ncat.edu

Abstract

A controller is used to serve as expert guidance in an environment where precision is necessary. In the case of a space shuttle orbiter, millions of dollars of equipment and human life dictate that accurate measurements be taken to ensure the safe arrival of the space orbiter during re-entry into the Earth's atmosphere. In previous work, several Sugeno fuzzy approximators used for the guidance of the orbiter re-entry were designed. The primary concern was to maintain the drag acceleration and the heating loads within design constraints to ensure safe arrival. As an extension of this work we will examine the effects of wind gusts on the performance and robustness of the fuzzy engines designed in there-entry system. For more realistic evaluation, we will include the effect of wind in the simulation. These gusts will be modeled as Rayleigh-type perturbations of the Mach number. We will also discuss the effect of multiple trajectory tracking. This is helpful in providing several acceptable options for choosing a re-entry path in real time.

Introduction

An important aspect of space travel is the effective execution of recentering the earth's atmosphere. The procedure is very sensitive to outside disturbances, so particular attention must be paid to effects such as wind perturbations and path choosing. Since the space shuttle orbiter acts like a glider, controlling the drag controls the rate of deceleration. This control is achieved by a series of banking maneuvers. Any slight outside fluctuation, like a wind gust, can cause the shuttle to be blown off path. That is why it is important to have a system that will be able to handle such a disturbance and maintain the original course.

Plotting many trajectories creates an envelope that outlines the boundaries of acceptable flight trajectories. The previous work developed the method for plotting an acceptable flight path. An acceptable trajectory is one that follows a smooth path. A "bumpy" path is one that has a lot of fluctuations over the given time frame. This bumping around in the atmosphere causes great strain on the orbiter. By having an envelope of trajectories, the shuttle can be ensured a safe path into its landing pattern. It is also important to have an acceptable entry angle. An angle that is too steep will cause to the orbiter to exceed its heat load constraints. An angle that is too shallow will cause the orbiter to act as a

missile crashing to Earth. The purpose of this paper is to discuss the results of testing the effects that the newly introduced wind disturbances had on the guidance laws created earlier, as well as, a brief discussion of multiple trajectory tracking.

Wind Perturbations

To serve the purposes of our experiments, the wind is represented by the equation

$$1.\quad dV = -D - \mathrm{g}\sin\gamma + 18.2x * randn(1)$$

where dV is the time derivative of the velocity of the orbiter, D is the drag, g is the acceleration due to gravity, γ is the bank angle, x represents the fraction of the wind to be calculated, and *randn is* a Gaussian random number generator. For this paper, the percentages of the wind that were used were 0.01, 0.02, 0.04, and 0.08. The wind will either work against the orbiter or with it depending on the random number generator. This coincides with the arbitrary nature of atmospheric wind. The first test performed was in an ideal or no wind environment. This served as the benchmark of our experiments. All subsequent experiments were compared to the ideal to determine the usefulness of the simulator. The effectiveness of the guidance laws was determined by the smoothness of the curve and its relation to the benchmark. The smoother the curve, the better the performance. The graph below illustrates the benchmark for this set of experiments.

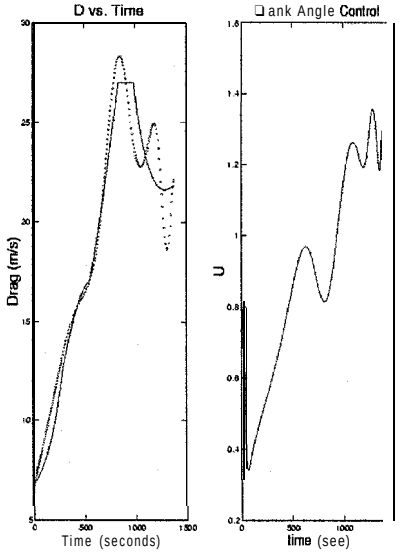


Figure 1.

The left portion of the figure represents the drag versus time graph. The dotted line is the actual trajectory taken, while the solid line represents the reference data. Under these test

conditions, this was considered an acceptable benchmark. The right portion of the figure shows the bank angle control for the ideal case. The bank angle is important because it serves as the only true way for the orbiter to control its rate of deceleration. The next step in the process was to introduce a minimal wind disturbance to see how the guidance laws would react. The first test was to use a x value of 0.01. This would represent a very small wind disturbance. Figure 2 illustrates the results of that test.

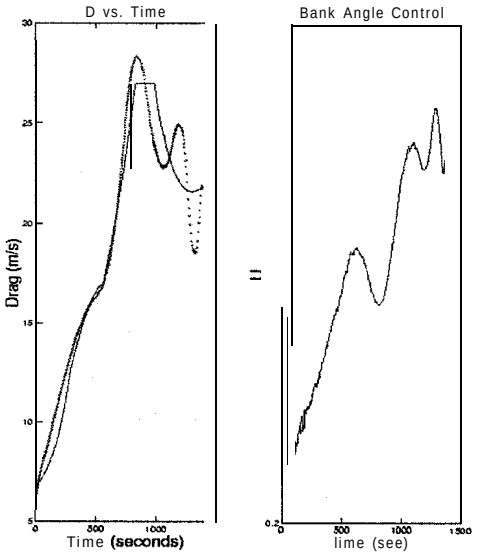


Figure 2.

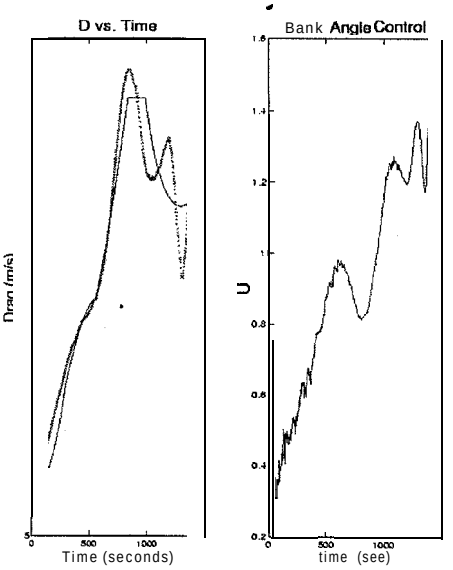


Figure 3.

The actual curve shown here was very similar to that of the ideal case. This confirmed an acceptable simulation of the orbiter reentry. The right portion of the figure does show that there was some trouble in controlling the bank angle. Our second wind simulation was with a slightly stronger wind disturbance. The idea here was to find a threshold for the simulator's control. Figure 3 shows that the controller experienced a bit more difficulty with the bank angle control than in the first try. The orbiter did manage to complete the reentry process, but its path was not as smooth as the ideal or even the previous try. This was obviously not a threshold since the path shown still qualified as an acceptable path, albeit, a rough one. The next attempt was with a x value of 0.04. The logic here was to double the previous value to test the simulator's performance. As seen in Figure 4, the path was quite rough. The bank angle control was also erratic. The orbiter was able to able to reach its endpoint even though the path fluctuated to a degree. What ended as the final attempt, the x value was again doubled. The result was an excessively fluctuating path, as well as, an erratic bank angle control. The orbiter also missed the prescribed endpoint due to the variation of its path. There was no need to continue with the testing since the orbiter guidance proved to be unsuccessful in the last try.

Multiple Trajectory Tracking

The idea of multiple trajectory tracking is to provide an envelope of different trajectories. These trajectories are centered on an initial flight path angle. The purpose is to always move toward this initial flight path angle for any given entry angle. This allows the orbiter a small degree of freedom in choosing an angle for reentry. It also serves to correct the path of the orbiter if it should be blown off course.

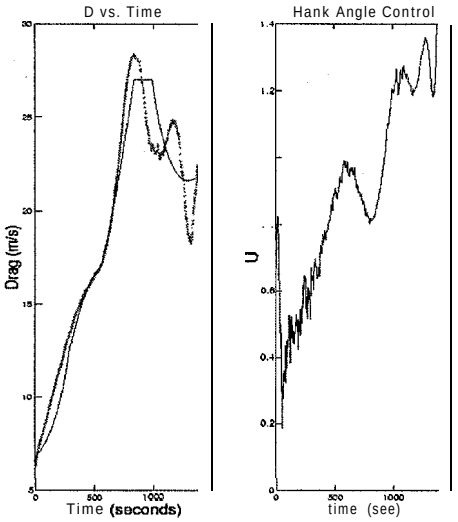


Figure 4.

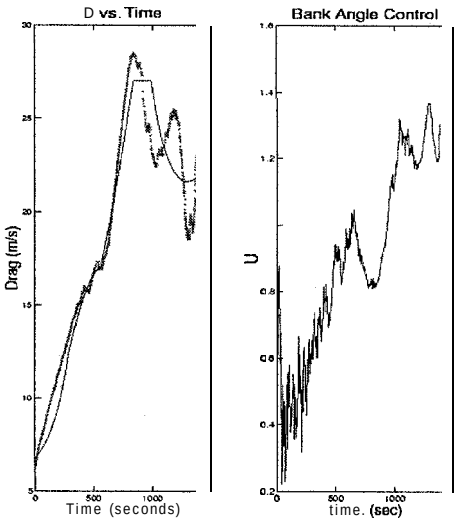


Figure 5.

Conclusion

The purpose of the series of experiments was simply to analyze and discover the realm of acceptability of the orbiter guidance laws developed earlier. It was proven that the laws behaved well in environments with slight wind. The results became erratic as the wind disturbances increased. Future care must be given to the effectiveness of the guidance laws when dealing with harsh atmospheric environments. The focus must be on the bank angle control. If this control changes too excessively, the orbiter will violate it dynamic constraints and the results would be disastrous. This is not to say that the guidance laws introduced previously are erroneous. They were not devised to handle strong wind gusts. With some fine-tuning, the guidance laws can serve as an expert controller. Also bear in mind that the determination for the wind gust was very general. This too could hinder the guidance laws from working properly.

References

1) Chaudhuri, Sam P. and Chen Ying. "Fuzzy Logic Controller Design: Target Tracking System and Automobile Control System", *Electro/94 International. Conference Proceedings,* Combined Volumes, pp. 383-392. May 1994.
2) Goldman, Jeffrey. "Path Planning Problems and Solutions", *Proceedings of the IEEE 1994 National Aerospace and Electronics Conference NEACON,* Vol.1,pp. 105-108, May 1994.
3) Sartor, Ken, Bikdash, Marwan, and Homaifar, Abdollah. "Fuzzy Guidance of the Shuttle Orbiter During Atmospheric Reentry"
4) Sartor, Kenneth J. W., Guiding the Longitudinal Motions of the Space Shuttle Orbiter During Atmospheric Reentry Using Sugeno Fuzzy Approximations, North Carolina A&T State University, 1996.

Explorers of the Universe

Marino C. Alvarez, Michael R. Busby, Goli Sotoohi
Center of Excellence in Information Systems
Tennessee State University

William J. Rodriguez
Astronomy and Physics Teacher
University School of Nashville
Nashville, Tennessee

Lee Ann Hennig, Jerry Berenty
Astronomy and Physics Teachers
Thomas Jefferson High School for Science and Technology
Alexandria, Virginia

Terry King
Astronomy Teacher
Hunters Lane High School
Nashville, Tennessee

Doreen Grener, John Kruzan
Earth Space and Physics Teachers
Wellington School
Columbus, Ohio

Abstract

The Explorers of the Universe is a multifaceted scientific/literacy project that involves teachers and their students with problem oriented situations using authentic materials. This paper presents examples of self-directed cases researched by high school students and the metacognitive tools they use in the planning, carrying out, and finalizing their reports.

Introduction

Engaging students in ways that demand decision and meaning-making are necessary if we are to educate for the future. Information that is isolated into compartmentalized units of study serves to stifle the interest and curiosity of high school science and mathematics teachers and their students, When this type of learning environment exists, students are placed into situations that demand memorizing this information for later retrieval or using formulae to solve mathematical problems with prescribed solutions for checking one's answers. This type of learning reduces knowledge into capsules that are sorted by topic with the ultimate purpose of retrieval by testing. Neglected are the processes for thinking about learning in meaningful contexts and the application of newly learned information.

This paper focuses on ways that teachers and their students are becoming "communities of thinkers" (Alvarez, 1966a, 1966b, 1997). Within these "communities" teachers and students are learning and thinking together about the subject discipline and the other content areas that relate to this discipline. In these classrooms, teachers are inviting students to participate by offering lessons and assignments that require critical thinking (thinking about thinking in ways to bring about change in one's experience) and imaginative thinking (exploring future possibilities with existing ideas) rather than emphasizing rote memorization of facts.

Communities of Thinkers

The Explorers of the Universe is an interdisciplinary project that engages teachers and their students in thinking about learning (http: //coe2.tsuniv.edu/explorers). This project is designed to encourage "communities of thinkers" in unique and multifaceted ways. One method is by analyzing and reporting data received from variable stars that teachers and students investigate by remotely controlled automatic photoelectric telescopes (Alvarez, 1995; Alvarez & Rodriguez, 1995). These telescopes are housed at the Fairborn Observatory in Washington Camp, Arizona, and controlled by astronomers at Tennessee State University in Nashville via the Internet.

Another approach is through affiliated projects at NASA Goddard Space Flight Center. Teachers and students in the Explorers of the Universe project have access to two projects Mars Orbital Laser Altimeter (MOLA) and Vegetation Climate Lidar (VCL) Mission to Mars -2000. Students in the middle and secondary grades will be investigating self-directed cases relating to these two projects. In the VCL mission, these students will be carrying on a longitudinal study, beginning with the initial stages of the planning processes, through launch 2000, and subsequent data gathering and analyses. The MOLA project has students using the mapping topography data received from the Mars Global Surveyor and developing analogies to their respective terrain locations.

A community of thinkers is defined as an active group of students and teachers striving to learn more about a discipline by engaging in critical and imaginative thinking (Alvarez, 1996a, 1995). During this inquiry, the teacher thinks about the facts and concepts that students need to understand, supplementary reading materials and artifacts that need to be provided, ways to incorporate other subject disciplines into the inquiry, and selects from an array of teacher-directed/teacher-assisted strategies to facilitate student thought. Likewise, the student becomes an active thinker in the learning process by engaging in the inquiry, relating prior knowledge and world experience both informal and formal, selecting from an array of student learning strategies, and working with the teacher toward extending meaning and understanding of the subject matter. Within our community of thinkers, teachers and students ask questions, seek answers, and reflect on their thoughts and feelings as they engage in action research case-based investigations.

Developing a community of thinkers focuses on the kinds of thought processes needed by the teacher and students to achieve learning outcomes. Thinking of ways to achieve learning outcomes differs from focusing on ways that learning outcomes can be achieved. The former is process-oriented; the latter product-oriented (Alvarez, 1995). In an effort to increase learning efficiency, we focus on the processes of thinking: selecting, eliminating, searching, manipulation,

and organizing information. Emphasis is placed on thinking as a process involving a sequence of ideas moving from some beginning thought, through a series of a pattern relationships, to some goal or resolution (Alvarez, 1995).

Metacognitive Tools for Learning

Concept maps and Interactive Vee Diagrams are two metacognitive tools that students use in their self-directed case-based investigations. These tools are used by students to regulate and monitor their own thinking. Concept mapping is intended to assist students to see relationships of central and subordinate ideas (see Novak, 1990; Novak & Gowin, 1984). Students are using concept maps to visually portray their case-studies that integrate the disciplines. For example, Katy Quinn, Mazi Abdolrasulnia, and John Farrell constructed their concept map to depict the relationships between Algol and Greek and Arabic mythologies. Vee diagrams are used to aid students to understand the structure of knowledge. They are used in their investigations by teachers and their students to plan, carry out, and finalize their case-based investigations. The Vee heuristic was developed by Gowin (198 1) to depict the important epistemological elements that are involved in the construction of knowledge or when making new meanings.

These learning tools serve to promote self-analysis and to enhance critical thinking by having one think about the ideas and their relationship to each other. When constructing and sharing these maps and vees with others, an individual is able to rethink and reflect upon the ideas visually portrayed on the concept maps and vee diagrams.

We have developed an ***Interactive Vee Diagram*** that is located on the Explorers of the Universe web site. The Interactive Vee Diagram appearing on the web site is in the same format as advocated by Gowin (1981). The elements arrayed on the Vee are identical to those that appear on the paper version. The only addition is a space for students to write suggestions or problems they may be encountering in their analysis of their research or the planning of their research study. The Interactive Vee is menu driven and asks students their name, school address, and e-mail address. Also included are instructions for entering information on the Vee. Other features include a link to a paper describing the uses of the Vee, an action research plan, a visual of a Vee with explanations of the epistemic elements, a movie clip of a student discussing his Vee with the teacher, and an animation of the components of the Vee. There is also a share folder for interactive correspondence. The Interactive Vee Diagram is restricted to those teachers and students affiliated with the Explorers project. Teachers give their students passwords for entry and students submit their information to the university educator at TSU's Center of Excellence in Information Systems. Depending on the nature of the project, TSU educators and astronomers (Greg Henry, Dr. Joel Eaton, and Dr. Frank Fekel) advise students by submitting their comments electronically on their respective Vee's. Students working on similar case-investigations can collaborate with other students at other schools and share their Vees with them by giving them their password for access.

Teacher/Student Investigations

Bill Rodriguez and Lee Ann Hennig have written ***Finding Periods in Variable Star Data,***

Explorers of the Universe Technical Manual 101-97. They have compiled sample data and presented challenging situations for students to apply their mathematics and scientific knowledge. Greg Henry, TSU astronomer, served as a consultant and provided the sample data for analyses. Students in both of these teachers' classes field-tested the manual and made suggestions for revision during its initial stages of development. The manual is used by students in the Explorers of the Universe project to better understand the techniques necessary when evaluating variable star data.

Bill Rodriguez and his astronomy students in grades 9-12 are working on several projects. Noah Charney, Vineet Misrah, and Adam Doochin are analyzing data from HR 1817. John Gillmor and Sarah Miller are analyzing Cepheid data. They are discovering the inverse-square relationship between distance and intensity of light using light probes. They then apply this information to the period-luminosity relationship to determine distances to Cepheid variable stars. Alicia Wright and Sedef Everest are making a case-study on "What Are Variable Stars?" They are using a broad lens and a multiple perspective to discuss how variable stars are classified and then are analyzing data from variable star types that are not already under analysis by others in the class. Ben Bredesen and Gabe Forsythe are using 35mm photography to determine the mass of Jupiter and are making an animation of the motion of the moons around Jupiter. Eugenio Cilentio has developed a manual on astrophotography. He will use his manual to photograph variable stars and carry out photometry with these stars.

Terry King and his high school students are investigating global warming. Angelis Bigott, J. Lynn Michael, Ryan Porter, Breanne Walker and Jessica Wood are investigating two causes that affect global warming. One is the combustion of fossil fuels which increases the atmospheric concentration of certain greenhouse gases. These include carbon dioxide, methane, ozone, halocarbons and nitrous oxide. These gases allow sunlight to penetrate but restrict the amount of heat radiated back into space. They are also studying aerosols, minute particles, that tend to get into the atmosphere blocking or reflecting light. Although the production of greenhouse gases may have a profound effect on the Earth's climate, it is not clear at the present if humans or natural causes are responsible for global warming.

Lee Ann Hennig's high school astronomy students are studying variable stars. Regan Blackwood, David Ellis, Jana Harris, William McDaniel, and Patrick Meade are engaged with various projects affiliated with variable stars.

Doreen Grener and John Kruzans' high school students are using the Technical Manual 97-101 to become acquainted with the procedures necessary for collecting and analyzing variable star data. Doreen's students, Andrew Kocsis and Matt Dickens are investigating climate temperatures on earth and in space. John Kruzan, physics teacher, is also involving his students in affiliated projects. Mike Larson, Llateefa McGhee, and Justin Mulhotra are applying their knowledge of physics to problems associated with space.

Allen Keel is a 10th grade student at Martin Luther King, Jr. Magnet School, Nashville, Tennessee. He is studying the effects of the sun's magnetic fields on the earth's climate. Allen is working with TSU astronomer Greg Henry and is correlating group sun spot numbers with the

historical temperatures and precipitation data for various geographic locations. Allen is using the www to find data for comparison and using MATLAB language to write the program for analysis.

Conclusion

Students in our project are learning and thinking with technology in meaningful ways (see Alvarez 1977b). The challenges of using authentic data and applying science and mathematics principles in problem-oriented situations enables students to use their critical and imaginative thinking skills. These students actively engage in "real life" projects and collaborate with their peers, teachers, and scientists in genuine circumstances. They submit their drafts of case reports for peer review, and publish their reports on the WWW for review by faceless and unknown persons throughout the world. E-mail communications to other teachers, scientists, and students augment their understanding with facts and ideas. These literacy, technology, and investigative scientific skills are enhanced by thinking and innovative teachers who are curious about their discipline and are willing to negotiate the curriculum with their students.

Students are learning about science and mathematics facts and concepts. They are also relating what they are learning to other subject areas that they have studied. Of more importance is that the students are beginning to look at how they think, learn, and ask questions. They have analyzed sources carefully and tried to identify new sources of information. Conceptual learning and decision-making and application of written discourse and scientific inquiry across disciplines through an emergent curriculum using innovative technology can create a community of thinkers. In the communities created, knowledge is shared and ideas are valued.

Acknowledgments

1. This paper is supported by the Center of Excellence in Information Systems at Tennessee State University, and by NASA through the Tennessee Space Grant Consortium NGT -40021, Network Resources and Training Site (NRTS) NCC5-96, and NASA Center for Automated Space Science NCC5-228.

References

Alvarez, M.C. (1997a). Communities of thinkers: Investigating interactive scientific literacy environments. In J. Willis, J.D. Price, S. McNeil, B. Robin, & D.A. Willis (Eds.), ***Technology and Teacher Education Annual, 1997*** (Vol. II, pp. 1236-123 9). Eighth International Conference of the Society for Information Technology and Teacher Education (SITE). Charlottesville, VA: Association for the Advancement of Computing in Education (AACE).

Alvarez, M.C. (1997b). Thinking and learning with technology: Helping students construct meaning. *NASSP Bulletin, 81, (592), 66-72.*

Alvarez, M.C (1966a). A community of thinkers: Literacy environments with interactive technology. In K. Camperell, B.L. Hayes, &R. Telfer (Eds.), ***Literacy: The information highway to success*** (pp. 17-29). American Reading Forum, vol. 16. Logan, UT: Utah State University.

Alvarez, M. C. (1966b). Explorers of the Universe: Students using the world wide web to improve their reading and writing. In B. Neate (Ed.), ***Literacy saves lives*** (pp. 140-145). Winchester, England: United Kingdom Reading Association.

Alvarez, M.C. (1995). Explorers of the Universe: An action research scientific literacy project. In K. Camperell, B.L. Hayes, &R. Telfer (Eds.), ***Linking literacy: Past, present, and future*** (pp. 55-62). American Reading Forum, vol. 15. Logan, UT: Utah State University.

Alvarez, M.C. & Rodriguez, W.J. (1995). Explorers of the Universe: A pilot study. In W.M. Linek * E.G. Sturtevant (Eds.), ***Generations of literacy*** (pp. 221-236). The Seventeenth Yearbook of the College Reading Association.

Gowin, D.B. (1981). ***Educating.*** Ithaca, NY: Cornell University Press.

Novak, J.D. (1990). Concept maps and vee diagrams: Two metacognitive tools to facilitate meaningful learning. ***Instructional Science, 19, 29-52.***

Novak, J.D., & Gowin, D.B. (1984). ***Learning how to learn. New*** York: Cambridge University Press.

Rodriguez, W. J., & Hennig, L.A. (1997). ***Finding Periods in Variable Star Data, Explorers of the Universe Technical Manual 101-97***. Nashville, TN: Center of Excellence in Information Systems.

ANALYSIS OF PROTEINS INVOLVED IN BIODEGRADATION OF CROP BIOMASS

Kamau Crawford and Audrey Trotman, Center for Food& Environmental Systems for Human Exploration of Space, Tuskegee University, Tuskegee, AL 36088, U.S.A.

ABSTRACT

The biodegradation of crop biomass for re-use in crop production is part of the bioregenerative life support concept proposed by the National Aeronautics and Space Administration (NASA) for long duration, manned space exploration. The current research was conducted in the laboratory to evaluate the use of electrophoretic analysis as a means of rapidly assaying for constitutive and induced proteins associated with the bacterial degradation of crop residue. The proteins involved in crop biomass biodegradation are either constitutive or induced. As a result, effluent and cultures were examined to investigate the potential of using electrophoretic techniques as a means of monitoring the biodegradation process. Protein concentration for optimum banding patterns was determined using the Bio-Rad Protein Assay kit. Four bacterial soil isolates were obtained from the G.W. Carver research Farm at Tuskegee University and used in the decomposition of components of plant biomass. The culture, WDSt3A was inoculated into 500 mL of either Tryptic Soy Broth or Nutrient Broth. Incubation, with shaking of each flask was for 96 hours at 30°C. The cultures consistently gave unique banding patterns under denaturing protein electrophoresis conditions, The associated extracellular enzymes also yielded characteristic banding patterns over a 14-day period, when native electrophoresis techniques were used to examine effluent from batch culture bioreactors. The current study evaluated sample preparation and staining protocols to determine the ease of use, reproducibility and reliability, as well as the potential for automation.

INTRODUCTION

The degradation of crops for re-use in a crop production sub-system, is a facet of the Bioregenerative Life Support System (BLSS) which is currently being studied by the U.S. National Aeronautics and Space Administration (NASA) for long-duration space missions. A space mission constitutes a closed environment, thus it becomes necessary to be aware of all components in the system. Biodegradation of crop biomass in a closed environment is quite susceptible to both the introduction of foreign agents and" to spontaneous mutations of microorganisms because of the multitude of the various interacting components in the system. The frequency of spontaneous mutations with large effects, is dependent on natural selection, and the degree of varying organisms in the system increase the chances of mutation (1). Mutations of any of the organisms in the system can lead to potentially adverse effects on humans. Mutated microorganisms may be parasitic and virtually undetectable. Such organisms in the human habitat also have the potential for transmission into the bioreactor system, and could have possible deleterious effects, As a result, tracking the enzymes in the system becomes an important parameter to be monitored (2). Changes in the concentration of enzymes can be used for the biological monitoring of microorganisms (2). Tracking the microorganisms and enzymes in the system monitors changes in the system early, alerting the individuals in the space station of the possible danger they are in. Bacteria, the most numerous of system components (3), are very important in crop degradation in a closed environment and can be easily monitored by tracking the protein enzyme products of the bacteria, Enzyme products are representative of the microorganisms in the system, since microorganisms only secrete specific enzymes. Some of the enzymes are constitutive, constantly being produced by the bacterial cells. Each species of bacteria produce unique sets of constitutive proteins. other enzymes are induced, only being produced in the presence particular substrates whether it is starch, cellulose, hemicellulose, etc.

(3). The constitutive proteins are found in the system with or without substrate (biomass in BLSS) in the system, while the induced proteins are found only with biomass in the system. With these theories in mind, a study was conducted to evaluate the use of electrophoretic analysis as the means of tracking the proteins in the system. Electrophoresis involves the separation of proteins to form banding patterns based on the mobility of the ions in a charged field. The technique involves a gel in which the protein is placed. The gel sits in a buffer solution between opposing negatively and positively charged electrodes (4). Because of the varying shape, size (ranging from 6,500-200,000 Dalton), and charge differences of proteins, each enzyme has different migration rates. The varying migration rates give unique bands that can be consistently reproduced. Electrophoresis has the ability to simultaneously measure the disappearance and appearance of products and has been shown to be effective in detecting mutations (5). It also has the potential to be automated since automated versions already exist which use computer assisted technology (6). The objective of the study was to evaluate the use of electrophoresis to rapidly assay for both the constitutive and induced proteins involved in bacterial degradation of crop biomass, using denaturing and native techniques, respectively.

MATERIALS AND METHODS

Biological degradation

Agents of biodegradation. Bacterial soil isolates were obtained from soil collected from a research field in Tuskegee AL, where sweetpotato had been grown for many years. Three bacterial cultures -WDSt3A, WLLC3, and WLPr1 - which were taken from the George Washington Carver research farm were isolated. WDST3A and WLLC3, both *Serratia marcessans* were selected for starch-degrading and cellulose-degrading capabilities, respectively. WLPr1, *Enterobacter cloacae*, was chosen for protein-degrading abilities. A fourth culture, American Type Culture Collection (ATCC) 11172, *Pseudonomas putida*, was selected for its cellulose-degrading ability.

Inoculum. One mL of each culture; and a mixed culture, composed of 0.1 mL of each culture, was inoculated into 100 mL of a sterile Tryptic Soy Broth (TSB) solution. TSB (Difco) was sterilized by autoclaving (121°C, 15 psi, 20 min). Incubation of inoculum was on an orbital shaker at 30°C for 40 hrs.

Each culture (25 mL) was transferred to 50 mL centrifuge tubes. Bacterial cells were washed by centrifugation (3000 rpm, 4°C, 10 tin) in Phosphate Buffer Saline (PBS) (pH 7.3) containing the following per liter: 8.0g NaCl; 0.34g KH_2PO_4; 1,21 g K_2HPO_4. This procedure was repeated twice. After the final wash, the pellet was suspended in 5 mL PBS.

Carbon source. Finely ground sweetpotato biomass was used as the carbon source. Portions consisting of 0.0 1g biomass were transferred into separate 50 mL centrifuge tubes. The biomass was sterilized by autoclaving (12 1 "C, 20 tin). To each tube with sweetpotato biomass was added to 15 mL of Basal culture medium (pH 7.0) containing the following salts per liter: 1.0g K_2HPO_4; 0.50g $NaNO_3$; 0.50g $MgSO_4 \cdot 7H_2O$; 0.50g KCl; 0.01g $FeSO_4 \cdot 7H_2O$.

Inoculation. A 0.1 mL portion of the designated inoculum was transferred to the sweetpotato biomass + culture medium in each tube. Incubation of the inoculated sample was in an orbital incubator/shaker at 30°C. A representative sample for each of the 5 inocula was removed and frozen on days 0,1,4,7, and 14. At each sampling time, an uninoculated treatment was also removed to serve as a control. There were two replicates for each treatment.

Protein Concentration Assay

For each sample, 100 mL was measured into each test tube and 5 mL of Protein assay dye reagent was added using the BIORAD Protein Assay Kit. IgA and Bovine Serum albumin (BSA) were used as standards. Of the standards 20, 40, 60, 80, and 100°/0 solutions were prepared. Absorbance was measured at 595 nm. Samples and standards were incubated at room temperature for -5 minutes (no longer than 45 minutes) prior to measuring absorbance. Results from the standards were used to plot a standard curve. The standard curve was used to determine protein concentration of samples.

Electrophoresis

Sample Preparation. From each sample, 1 mL (containing both biomass and effluent) was pipetted into 1.5 mL microcentrifuge tubes, The microcentrifuge tubes were centrifuged at 14,000 rpm, 4°C for 5 minutes (Eppendorf). The effluent was separated from the pellets and transferred to different microcentrifuge tubes.

One part (150 µL) of the pellet was removed in PBS by adding 150 µL PBS to each of the pellet samples, vortexing, and pipetting 150 µL back out. This part was then transferred to the upper section of filter/microcentrifuge tubes. One part 80°/0 ethanol was then added to the pellet and allowed to sit for an hour, after which the microcentrifuge tube was centrifuged at 14,000 rpm, 4°C for 10 min. The portion that passed into the base of the filter/microcentrifuge tube was saved. The filtered portion was then separated into two aliquots by transferring 20 µL to separate microcentrifuge tubes. Of the two aliquots, one was used for native gel electrophoresis and the other for denaturing gel electrophoresis.

One part (150 µL) of the effluent of each sample was removed and placed into the upper sections of the filter/microcentrifuge tubes. The microcentrifuge tube was then centrifuged at 14,000 rpm, 4°C for 10 min. The portion passed into the base was collected and further separated into two aliquots by transferring 30 mL into separate tubes. Of the two aliquots one was used for native electrophoresis and the other for denaturing electrophoresis.

Native sample preparation, Native sample buffer was prepared by adding the following to de-ionized H_2O to make 8 mL: 1.0 mL 0.5 M Tris-HCl; 0.8 mL Glycerol; 0.4 mL O. 5°/0 bromophenol blue. For each sample (approx. mL), 80 mL of the native buffer was added to each tube and mixed. Ten mL of the molecular weight standard (Broad Range SDS Page Bio-Rad Control #76587) was added to 200 mL native buffer immediately prior to loading. The gel used was a 4-15% Tris-Glycine Bio-Rad ReadyGel. Running buffer was a 1X solution of Tris-HCL running buffer.

Denaturing sample preparation. Denaturing sample buffer was prepared by adding the following to de-ionized H_2O make 8 mL: 1.0 mL 0.5 M Tris HCl; 0.8 Glycerol; 1.6 mL 10% SDS; 0.4 mL Beta-mercaptoethanol; 0.4 mL 0.4% bromophenol blue. For each sample 80 mL of the denaturing buffer was added to each tube and mixed. The microcentrifuge tubes were capped and placed into a rack, which was immersed in 95°C water for 4 minutes. This was done immediately prior to loading the samples in the well. Ten mL of the molecular weight standard (Promega Midrange Molecular Weight Standard) was added to 40 mL of the denaturing sample buffer. Running buffer used was a 1X solution of Tris/Glycine/SDS running buffer.

Running Procedure. The electrophoresis technique used was the SDS-page gel electrophoresis method ( ). Gels used were 4- 15°/0 Tris Glycine BIORAD ReadyGel Polyacrylamide gels for separation in range of 40,000-200,000 Dalton. An EC-135 Power supply was used and the gels

were ran at a constant voltage of 200 volts for approximately 45 minutes at room temperature. In loading, 10 mL was added to each well. The samples were added to wells 2 through 9, while the standard was added to wells 1 and 10.

Optimal Band Determination. Different concentrations/ dilutions of the sample were examined for optimum band definition. The different dilutions tested were the following:

Undiluted sample
1:2 1 mL sample + l. 0 mL PBS
1:5 1 mL sample + 4.0 mL PBS
1:10 1 mL sample + 9.0 mL PBS
1:100 100 µL sample + 9.9 mL PBS

Each of these dilutions was added to the gels in both 10 and 15 mL sample per well. Optimal banding pattern was assessed by the clearness and definition of the bands,

Band Visualization. Gels were obtained from chambers of holding vessel, and placed in a fixative enhancer solution for 20 min. Fixative enhancer solution was composed of the following mL per liter: 500 mL methanol; 100 mL acetic acid; 100 mL acetic acid. Gels were gently agitated in solution for approx. 20 minutes. Gels were then rinsed in de-ionized H2O for 20 min. Gels were stained and developed with gentle agitation in staining solution for approx. 20 minutes, Staining solution was prepared by adding the following to de-ionized H2O to make 100 mL: 5.0 mL Silver Complex Solution; 5.0 mL Reduction Moderator Solution; 5,0 mL Image Development Reagent; 50 mL Development Accelerator Solution. Once stained, the gels were placed in a 5% acetic acid solution to stop the reaction and photographed,

RESULTS AND DISCUSSIONS

Electrophoresis

Optimal Band Determination. It was determined from the distinct definition of the bands that with 10 µL of the sample per well, the 1:2 dilution was optimum, With 15 µL of the sample added per well, 1:10 dilution was better. Determination was made according to personal visualization of the bands, Because of the good band quality, 10 µL sample per well was chosen because of the ease of loading.

Protein Concentration. The protein concentration was determined in order to have adequate amounts of protein for electrophoresis. The concentrations of each culture is as follows:

WLPr1 0,20
WDSDT3A 0.05
WLLC3 0.54
ATCC1 11720.065

Electrophoresis Analysis. Analysis showed that banding patterns were unique from culture to culture, and were consistently reproducible over the 14 day period. Patterns increased in intensity as the days increased. Denaturing gels tended to have more defined banding patterns for both the pellets and the effluents of the cultures, while native gels showed more blurred definition. This lack of definition is a general characteristic of native electrophoresis. Slight changes in any conditions such as pH, temperature, or protein concentration can easily effect native electrophoresis as opposed to more extreme changes in denaturing techniques. For the bacterial culture, ATCC 11172, the general banding patterns were consistent at all sampling times. A similar result was found for all single culture inoculations. Generally, no bands were visible on Day O, however, as the experiment progressed, the band intensity increased. This can be explained

by the general nature of enzymes, especially for induced enzymes. As time progressed, bacteria release more enzymes into the system, and as a result, electrophoretic analysis detected more enzyme products. Bacteria released a wider variety of induced proteins, which are produced depending on substrate. The gels for mixed culture inoculum, showed more smearing than when single culture inoculation was used. This is to be expected because the interaction of various microorganisms in a mixed system results in a competition of proteins of similar size and charge. Although not distinct, it still showed much of the unique banding patterns of the single cultures.

CONCLUSION

Mixed culture more nearly represents the actual conditions that will be seen in a closed environment. Further study is being completed to sharpen the definition and improve the techniques of electrophoretic analysis of mixed culture systems. The larger degree of enzyme products in the system, most likely causes "overlapping" of similar products in electrophoresis. Future studies will be to focused on the characterizing the enzymes systems based on electrophoretic and DNA analyses. Characterization of the microorganisms based on specific enzymes products involved can lead to replenishment of essential bacteria that have been depleted; thus improving enhanced degradation in a closed environment.

ACKNOWLEDGEMENT

This research was conducted with grants from the United States Aeronautics and Space Administration-NCC 9-51 and the Department of Agriculture/CSREES. The authors are grateful for support of this work by the G. W. Agricultural Experiment Station. Ms. Geralda Parvilus, Graduate Student and Dr. Phillip Loretan, Project Coordinator.

REFERENCES

(1) Garcia-Dorado, A, 1997. The Rate and Effects Distribution of Viability Mutation in Drosophila: Minimum Distance Estimation. ***Evolution, 51(4),*** pp. 1130-1139.

(2) Gray, M., A. Charpentier, K. Walsh, P. Wu, and W, Bender. 1991. Mapping Mutations in the Drosophila ***rosy*** Locus Using Denaturing Gradient Gel Blots. In: *Genetics,* 127:139-149. Genetics Society of America.

(3) Paul, EA., and F.E. Clark. 1996. Soil Microbiology and Biochemistry, pp. 71, 65. San Diego: Academic Press.

(4) Rybicki, Ed, and Maud Purves. "SDS Polyacrylamide Gel Electrophoresis (SDS-Page)." Dept. Microbiology, University of Cpae Town. Online. Internet. 8 Nov. 1997. Available FTP: http: //www. uct. ac. za/ microbiology/ sdspage.html.

(5) Payne, Jerry F., and Linda L. Fancey. 1982. "Effect of Long Term Exposure to Petroleum on Mixed Function Oxygenases in Fish: Further Support for use of the Enzyme System in Biological Monitoring. ***Chemosphere,*** Vol. II, No. 2, pp. 207-2 13. Pergamon Press.

(6) Hakim, B. H., E. Delatour, and M. Hanss. 1985. Automation of an Alternating Field Microelectrophoresis Apparatus. *Rev. Sci. Instrum*, 56(3), pp.462-467. American Institute of Physics. March.

PROTEIN KINASES POSSIBLY MEDIATE HYPERGRAVITY-INDUCED CHANGES IN F-ACTIN EXPRESSION BY ENDOTHELIAL CELLS.

Felisha D. Love[§], Caroline D. Melhado[§], Francis N. Bosah[‡],
Sandra A. Harris-Hooker[‡], and Gary L. Sanford[§]

Departments of Biochemistry[§] and Medicine[‡]
Morehouse School of Medicine, Atlanta, Georgia 30310-1495

INTRODUCTION

Basic cellular functions such as electrolyte concentration, cell growth rate, glucose utilization, bone formation, response to growth stimulation, and exocytosis are modified in microgravity (1). These studies indicate that microgravity affects a number of physiological systems and included in this are cell signaling mechanisms. Rijken and coworkers (2) performed growth factor studies that showed PKC signaling and actin microfilament organization appears to be sensitive to microgravity, suggesting that the inhibition of signal transduction by microgravity may be related to alterations in actin microfilament organization. However, similar studies have not been done for vascular cells. Vascular endothelial cells play critical roles in providing nutrients to organ and tissues and in wound repair.

The major deterrent to ground-based microgravity studies is that it is impossible to achieved true microgravity for longer than a few minutes on earth. Hence, it has not been possible to conduct prolonged microgravity studies except for two models that simulate certain aspects of microgravity. However, hypergravity is quite easily achieved. Several researchers have shown that hypergravity will increase the proliferation of several different cell lines while decreasing cell motility (3) and slowing liver regeneration following partial hepatectomy (4), These studies indicate the hypergravity also alters the behavior of most cells. Several investigators have shown that hypergravity affects the activation of several protein kinases (PKs) in cells (5,6). In this study, we investigated whether hypergravity alters the expression of f-actin by bovine aortic endothelial cells (BAECs) and the role of PK's (calmodulin II dependent, PKA and PKC) as mediators of these effects.

METHODS

BAEC's were obtained from NIGMS/Corielle Cell Repository and maintained in Dulbecco's Modfied Eagle Medium (DMEM) supplemented with 1X antibiotic and 10% fetal bovine serum (FBS). Cells were subculture at 1 X 105 cells in slide flaskettes. Cells were subjected to hypergravity treatment (centrifugation at 6G) in the presence or absence of protein kinase inhibitors (PKIs) - KT5720 (PKA) and cherylerythrine (PKC),were added singly to the medium of subconfluent cultures (2μL PKI/mL) for 24-72 hr. Stationary control cultures were treated similarly but without centrifugation.

Cultures were assessed for f-actin expression as follows: after 24,48 or 72 hr, cultures were rinsed with phosphate buffered saline (PBS) and fixed with 1% glutaraldehyde/0 .02 °/i Triton X- 100/2.5°/0 formalin in PBS for 30 sec then fixed in 1% glutaraldehyde/2.5% formalin in PBS for 24 hr.

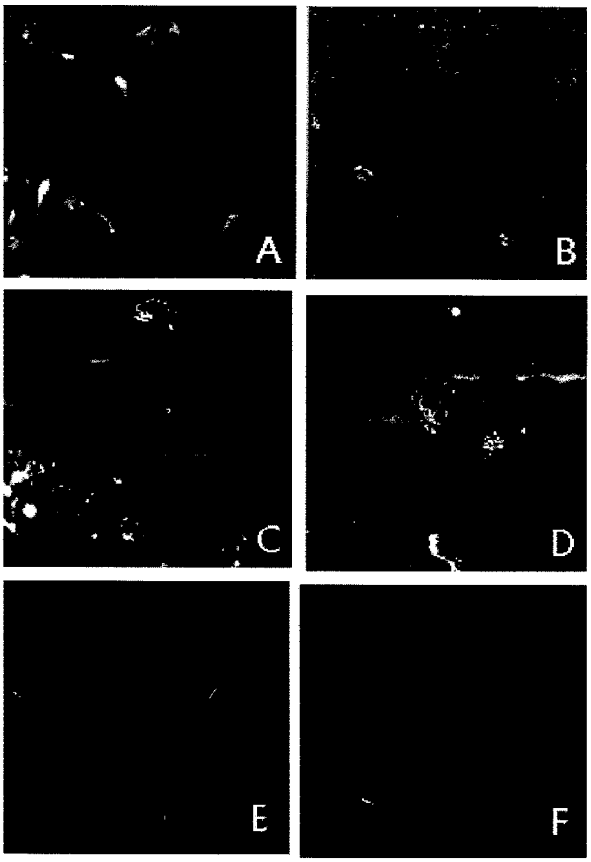


Figure 1. **Fluorescent micrographs of BAEC stained for f-actin in the presence of PKC inhibitor without FBS at 24, 48 and 72 hr under stationary (A, C &E) and hypergravity (6G) conditions (B, D & F).**

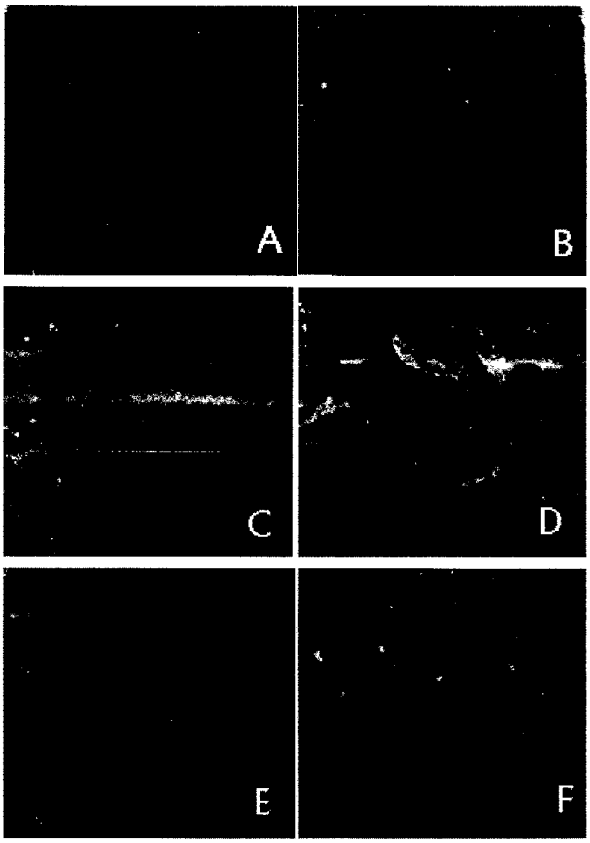


Figure 2. **Fluorescent micrographs of BAEC similarly to cells in Fig. 1, except carried out in the presence of 10% FBS. Stationary (A, C & E) and hypergravity (6G) treated cultures (B, D &F) at 24, 48 and 72 hr.**

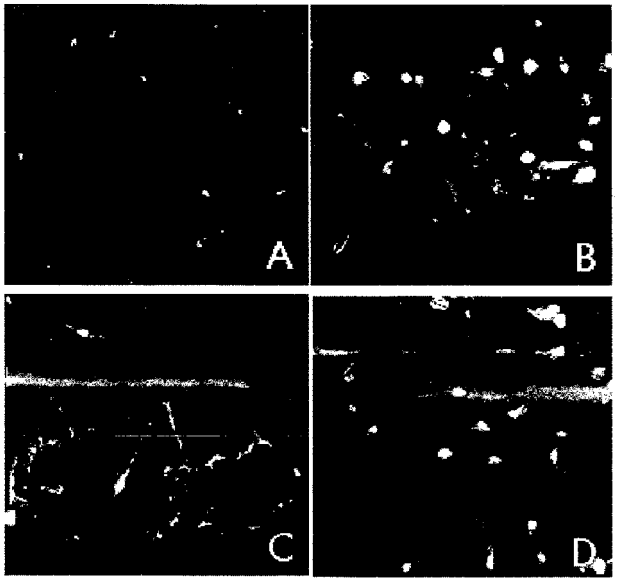


Figure 3. **Fluorescent micrographs of BAEC stained for f-actin in the presence of PKA inhibitor at 24 and 48 hr under stationary (A & C) and hypergravity (6G) conditions (B& D).**

Following fixation, the cells were stained with FITC-phalloidin and examined by fluorescence microscopy.

RESULTS

We found that cells treated with PKI's and maintained under stationary or hypergravity conditions did not show any morphological differences by phase contrast. Subconfluent BAECs maintained in DMEM (without FBS) and subjected to hypergravity in the presence of the PKI KT5720 was found to have decreased f-actin after 24 hr compared to controls (Fig. 1). This decrease was not seen at 48 or 72 hr; hypergravity treated cells showed the same fluorescence as controls for these times. In contrast, cells maintained in DMEM with 10% FBS, showed no differences in f-actin staining between hypergravity treated and stationary controls (Fig.2). These studies indicate that inhibiting PKA activity did not alter f-actin expression by BAEC in the presence of high serum and only transiently affected f-actin expression in the absence of serum. This suggests that PKA may not be involved in the signaling pathways for gravity effects on f-actin cytoskeletal structures.

Figure 3 shows the results of cherylerythrine treatment on f-actin expression after 24 and 48 hr of hypergravity or under stationary conditions. At both times studied there was increased f-actin fluorescence in hypergravity treated cells compared to stationary controls. This finding suggests that PKC is involved in the signaling pathway mediating the cellular effects of hypergravity.

ACKNOWLEDGMENTS: This study was supported by grants NASA NAG9-644 and NCCW-008.

REFERENCES

1. Hughes-Fulford M (1991). Altered cell function in microgravity. Exp Gerontol 26:247-256.
2. Rijken PJ, Boonstra J, and de Laat SW (1994). Effects of gravity on the cellular response to epidermal growth factor. Adv Space Biol Med 4:159-188.
3. Gruener R and Hoeger G (1990). Vector-free gravity disrupts synapse formation in cell cultures. Am. J Physiol 258: C489-C494.
4. Tschopp A and Cogoli A, Hypergravity promotes cell proliferation. Expermentia 39, S75-S76, 1988.
5. Philpott DE, Popova I, Kato K, et al. (1990). Morphological and biochemical examination of Cosmos 1887 rat heart tissue: Part I--Ultrastructure. FASEB J 4:16.
6. DeGroot KP, et al., Nuclear responses to protein kinase C signal transduction are sensitive to gravity changes. Exp. Cell Res. 197:87-90, 1991.

Biodegradation of Fresh vs. Oven-Dried Inedible Crop Residue in a Continuously Stirred Tank Reactor

Kamau Crawford[1], and Richard Strayer[2]
Center for Food & Environmental Systems for Human Exploration of Space, Tuskegee University, Tuskegee, AL 36088, U.S.A.

ABSTRACT

The degradation of soluble organics and mineral recovery from fresh and oven-dried biomass were compared in an Intermediate-Scale Aerobic Bioreactor (8 L working volume) to determine if drying crop residue improves performance in a continuously stirred tank reactor (CSTR). The study was conducted in an Intermediate-Scale Aerobic Bioreactor (ISAB) CSTR with dimensions of 390mm height x 204 mm diameter. The pH in the bioreactor was controlled at 6.0, temperature at 30°C, and aeration at 7.0 $L_{air}min^{-1}$. Gases monitored were CO_2 evolution and dissolved oxygen. Homogeneously mixed wheat cultures, used either fresh or oven-dried biomass and were leached, then placed in the I-SAB for a 4-day degradation period. Studies found that mineral recovery was greater for leached oven-dried crop residue. However, after activity by the mixed microbial communities in the I-SAB CSTR, there were little notable differences in the measured mineral recovery and degradation of soluble organic compounds. Degradation of soluble organic compounds was also shown to improve for leached oven-dried crop residue, but after mixing in the CSTR, the degradation of the fresh biomass seemed to be slightly greater. Time for the biomass to turn in the CSTR appeared to be one factor for the experimental differences between the fresh and oven-dried biomass. Other factors, although not as defined, were the differing physical structures in the cell walls and varying microbial components of the fresh and oven-dried treatments due to changes in chemical composition after drying of the biomass.

INTRODUCTION

In a closed environment (such as a space station), optimal use of all resources becomes a necessity. The utilization of waste materials (crew and plant derived) becomes an important topic for research by scientists involved in Advanced Life Support (ALS). It has been shown that edible crops can effectively be grown for space using a hydroponic system (l). However, because such a large percentage of crops are inedible, crop residue becomes a major resource for use in a closed environment. Bioregeneration of plant nutrients from inedible crop residue has a key role in enhancing life support on long-term space missions (2). This method basically involves the utilization of physiochemical and biological processes to recover minerals from crop biomass (3). The limiting factors in missions, such as cost, energy, and labor, require that bioregenerative techniques be selected with strict criteria.

Although inedible crop residue is an excellent nutrient source, it contains a large amount of soluble organic matter that becomes problematic when incorporated directly into a nutrient system (2,4). A surplus of this organic matter can 1) form microbial biofilms on both the plant and nutrient delivery system, 2) cause increased root respiration, and 3) lead to increased denitrification at the rhizosphere microbial community (0,6). It, therefore, becomes necessary to explore methods of degrading these soluble organic compounds before introduction into the nutrient delivery system (7). At KSC, two methods, Continuously Stirred Tank Reactors (CSTR) and composers, were found to adequately reduce the soluble organics while simultaneously recovering the inorganic minerals (8). To date, at KSC, composter studies generally utilize fresh biomass while CSTR studies use dry (oven-dried and frozen) biomass.

Oven-drying crops requires additional heat energy not needed for using fresh biomass. However, results from previous KSC comporting, leaching and CSTR studies indicated that mineral recovery and the total organic carbon (TOC) values were lower for fresh crop residue than for oven-dried crop residue. This raised an interesting question--was oven-drying really a better method for recovering minerals, and if so, was the cost of oven-drying worth the benefits of releasing more minerals and soluble organic materials?

To address the first part of the question, a study was conducted comparing fresh and oven-dried crop residue under the same conditions. The objective of this experiment was to determine the differences between fresh and oven-dried inedible crop residue with regards to degradation of soluble organic compounds and recovery of inorganic minerals.

MATERIALS AND METHODS

Environmental Parameters: A 20-day study was conducted with composite wheat (*Triticum aestivum L.)* residue consisting of four cultures -- Veery, Superdwarf, Apogee, and Yecora roja. The study was conducted in duplicate Intermediate Scale Aerobic Bioreactor (I-SAB) Continuously Stirred Tank Reactors (CSTR). The reactors had dimensions of 390 mm x 204 mm (height x dimension) and an 8L working volumes. Three parameters--temperature, aeration, and pH-- defined treatment conditions. Temperature was controlled at 30°C by an external heating jacket. The air flow rate and stirring were adjusted to keep the dissolved oxygen above 2.0 mg O_2 L-' and to control excessive foaming. This desired rate was attained at a aeration rate of 7.0 L_{air} rein-'. The pH level was maintained at 6.0 with one normal nitric acid (IN HNO3). Dissolved oxygen and CO_2 production were also monitored throughout the experiment. The I-SAB system was described previously by Finger and Strayer (0).

Wheat pre-treatment: Substrate consisted of wheat stems and leaves from a biomass production chamber (BPC). The stems and leaves were cut into 1 cm lengths with a floral stem cutter and manually mixed to homogeneity in a plastic bin. The mixed chopped wheat biomass was separated into two equal portions. One portion was oven-dried (70°C to constant weight) and the other portion of fresh biomass was refrigerated at 4°C.

Leaching Procedure: Prior to the first bioreactor run, fresh wheat residue was leached for 2 hours in 600 mL beakers at room temperature. Three replicate samples, each containing 00 g flesh wheat biomass in 200 mL distilled water per beaker, were leached. The process was performed at room temperature being disturbed only by stirring every 30 minutes. The stirring was manual with a stirring rod. The wheat biomass was vacuum filtered (Coming) using 0.2 mm pore size filters (Coming 20982-PF). The filtered contents were used for analysis.

Experimental Procedure: The dry biomass percent moisture was determined prior to the first run using fresh and dry biomass. This value of 86.7% was used to calculate the amount of dry biomass to be used in the oven-dried run. Each I-SAB ran concurrently first with fresh biomass and then with oven-dried biomass. The substrate was fed in batch mode at the start of each run.

The bioreactors were filled with 8L of distilled water. They were then inoculated with four frozen cubes consisting of contents from a previous I-MB amendment study. The inoculum cubes contained active microbial communities.

Of the two treatments, the fresh wheat was fed to the I-SAB CSTRs first to ensure its fresh status. I-MB #1 was fed 422g fresh weight (equivalent to 26.09g dry weight), and I-SAB #2 was fed 222g fresh weight (equivalent to 03. 17g dry weight) of the wheat crop residue. The ISAB CSTRs ran until steady state was reached, which was defined by similar CO_2 production curves and total daily CO_2 values. The resulting time course was 4 days for completion of one run. During these 4 days, ~022 mL of the biomass (solids and liquids) was harvested 1 day after feeding for data analysis. All water loss (including evaporated water) was accounted for by the addition of distilled water back into the system. This was done by manually pouring the water

into the bioreactor through a constructed opening in the lid. After 4 days, -4222 ml of the biomass (solids and liquids) was harvested for a data analysis. Due to difficulties with the CO_2 analyzer at the start of the flesh run, we fed another 422g and 222g to I-SAB #1 and I-SAB#2, respectively, to collect CO_2 production data, This biomass was fed, on day O, adding to the remnants of wheat biomass already left over from the previous run. Samples were not taken from the second batch of wheat biomass.

Following the fresh run, the oven-dried run was conducted under the same experimental conditions, with the exception of adjustments made for moisture content. I-SAB #1 and I-SAB #2 were fed 26.6g and 03 .2g dry wheat crop residue, respectively. The dry wheat had been refrigerated at 4°C for 24 days before being fed to the I-SABs.

Sampling Techniques: Day 2 samples were taken from the leached biomass prior to the first run. As mentioned earlier, they were filtered with vacuum filters at 2.2 mm pores size filters. Day 1 and 4 samples were bailed from each I-SAB with a 4L beaker. Samples were taken from the bioreactors while the biomass was stirring in order to obtain homogeneous mixtures, They were then filtered through 4 metal screen sifters (Fisher Scientific Company) ranging in pore sizes of 2 mm to 202 mm, and separated into solid and liquid phases. The liquid phase was defined by those particles that passed through the sifters. The solids were collected together from the sifter membranes and used for analysis. Some of the liquid (soluble residue) was used for dry/ash weights. The rest was allowed to settle for approximately 24 hours, after which, the remaining liquid layer was separated and run through a MinitanTM Ultrafiltration System(Millipore) at 2.1 micron pore size. This filtered effluent was used for more sample analyses.

Response Variables: Results were compared using five main response variables -- three were used for organic degradation and two for mineral recovery. Organic degradation was estimated using CO_2 production, dry weight/ash weights, and total organic content (TOC) values. CO_2 production monitored microbial activity by way of a carbon dioxide analyzer (Model LI-6201, LI COR). Dry and ash weights of the crop residue, taken on day 2 and day 4, evaluated percent volatile solids lost and estimated percent degradation of organic matter. TOC values were determined from filtered samples using an Inductively Coupled Argon Plasma Spectrometer (ICAP). Consequently, TOC values were the primary assessment for soluble organic degradation. Mineral recovery was assayed with the ICAP and with Visible Light Spectrophotometry by Segmented Flow Analyzers (Technicon). The ICAP was used for elemental analysis (K+, Ca^{+}, Mg^{2+}...) while Visible Light Spectrophotometry was used more for anionic concentration detection. Both methods for mineral recovery utilized filtered samples.

RESULTS AND DISCUSSION

Organic Degradation: Two runs, one with fresh and one with oven-dried wheat residue, were used to determine the differences in selected response variables. Of these variables, only CO_2 production was continuously measured throughout each run. The other response variables were measured using collected samples. The I-SABs were fed inedible wheat residue at the beginning of each run, and were allowed to continue running until CO_2 levels were at a steady state. Steady state was generally reached within 48 hours of the feeding. The usual trend was peak CO_2 production around the 12th hour of feeding which was accompanied by a rapid decrease soon thereafter. Total CO_2 production in I-SAB #1 was 11.8 and 9.32g CO_2-C, for fresh and oven-dried biomass, respectively, and in I-SAB #2 0.06 and 0.26g CO_2-C for fresh and oven-dried biomass, respectively. Figure 1 displays CO_2 production curves for the fresh and oven-dried biomass in I-SAB #1.

Dry weights and ash weights were used to estimate the percentage of degraded organic matter, both soluble and insoluble, for each experimental run. The focus in this paper is on the soluble organic matter of carbon degradation. Ash weight data indicated a 1.0% reduction of

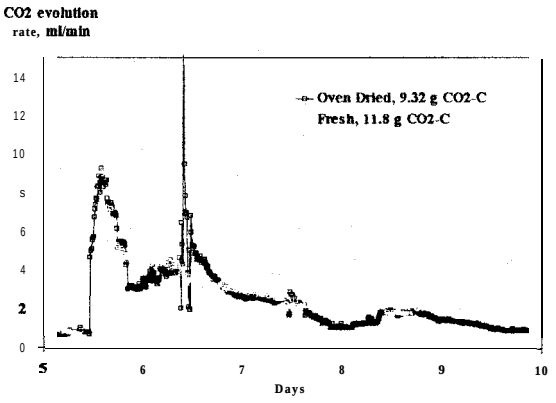


Figure 1. CO2 evolution in I-SAB #1 of fresh Vs. oven-dried wheat residue in mL of CO_2/min, taken from Day 5 to Day 12. Total CO2-C in g CO2-C.

volatile solids during the oven-drying process. After four days, studies showed conflicting data for percent degradation between the fresh and oven-dried biomass. I-SAB #1 displayed 36% degradation of soluble organic C for the fresh biomass compared to 49% degradation of the oven-dried biomass, a fairly large difference. I-SAB #2, on the other hand, exhibited an essentially equal degree of degradation between the two treatments, with 32% v. 34% degradation for the fresh and oven-dried biomass, respectively. Notice the lack of correlation between the dry/ash weights and CO_2 data. While the product formation, represented by respiration of CO_2, was higher for fresh wheat, substrate disappearance was higher for the oven-dried wheat residue. There was some question as to the validity of this data. The discrepancies could have resulted from inconsistent sampling. Taking homogeneous samples for analysis was quite difficult due to the large and differing sizes of the substrate particles. Inconsistencies in sampling could have influenced solid: liquid distribution ratios, subsequently causing errors in collected data. Therefore, dry/ash weights were not used in the final interpretation of the data. Carbon dioxide production was a continuous function, therefore we accepted those data.

TOC values correlated with the CO_2 data, further supporting the claim that CO_2 data was indeed trustworthy. As shown in figure 2, more soluble carbon was released from the dry leached biomass, but not much more came out after that. However, fresh had a large release on day 1 and fairly equal to the dry biomass by day 4. This could mean that it just took a longer time for the microbes to begin degrading the fresh biomass perhaps because of the varying chemical compositions of the fresh and oven-dried crop residue.

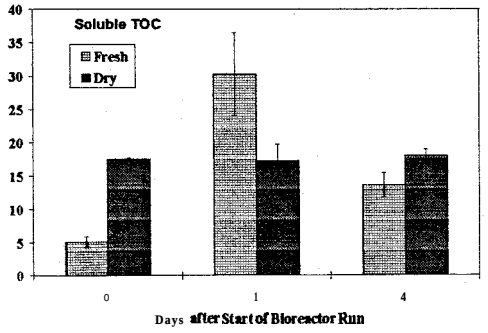


Figure 2. Total Organic Content of fresh vs. oven-dried wheat residue in mg C/g crop residue added taken from day 2 to day 4.

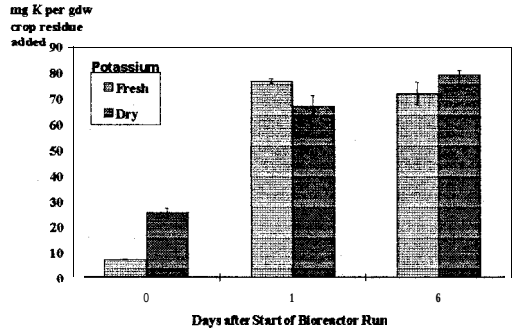


Figure 3. Potassium content in effluent of fresh vs. oven-dried wheat residue in mg K/g crop residue added.

From the TOC values and CO_2 data, we concluded that degradation of the leached wheat residue was improved by oven-drying. This was already expected from the previous studies. However, it was also determined that although degradation from day 2 to day 1 was very similar, it was slightly better in the fresh biomass.

Mineral Recovery: Mineral analysis was performed using a ICAP and Visible Light Spectrophotometry. Figure 3 shows some potassium results from the mineral analyses. Like potassium, most of the macronutrients were recovered better in the leached oven-dried biomass (day 2) with calcium being the only exception. However, for most minerals, recovery on days 1 and 4 were fairly similar between the fresh and dry. As for the soluble organic degradation, this could mean that fresh biomass releases minerals a little slower than dry biomass.

Table 1. Comparison of mineral content of fresh and oven-dried crop residue. Data are from the average of I-SAB #1 and 2 after conversion from mg/L to mg/ gdw crop residue added

	Component (mg/ gdw)									
	PO4-P	Potassium	Calcium	Magnesium	Iron	Boron	Copper	Manganese	Zinc	Chloride
Fresh (day)										
0	0.5049	6.9575	0.8784	0.3516	0.0031	0.0065	0.0056	0.0188	0.0059	0.4766
1	5.3085	76.7644	5.1189	2.1476	0.0158	0.0292	0.0186	0.1023	0.0557	7.7336
4	4.8651	71.8845	4.9689	2.0405	0.0243	0.0135	0.0137	0.0907	0.0310	11.5829
Oven-dried (day)										
0	1.6044	25.2661	0.4641	0.4859	0.0056	0.0031	0.0028	0.0105	0.0065	3.7151
1	4.1591	71.7424	3.8416	1.9402	0.0205	0.0166	0.0158	0.0649	0.0491	8.0724
4	3.5341	79.0761	5.9800	2.4572	0.0274	0.0135	0.0155	0.1033	0.0520	7.7830

Although time was a factor for the leached biomass, it did not seem to improve the recovery of most of the minerals after day 1. These trends were also seen in the macronutrients, specifically phosphorus, calcium, magnesium, and potassium. On the other hand, many of the micronutrients exhibited different trends. For instance, iron showed a steady increase from day 2 to day 4, perhaps indicating that with time more of that mineral could have been recovered. Boron recovery peaked at day 1, with a subsequent decrease from day 1 to day 4. All of the micronutrients analyzed, specifically iron, sodium, boron, and copper, released out more in the fresh leached biomass. The opposite was true for the macronutrients. Data for these nutrients were difficult to define because they were at such low detection levels.

SUMMARY AND CONCLUSION

More in-depth studies need to be performed in order to determine if the results were in fact consistent and reliable. Successive studies could provide definitive relationships between the fresh and oven-dried crop residue. This experiment could serve as a preliminary for these future studies, providing the limits at which the experiment could be optimally conducted. In this

particular experiment, other factors could have led to the increased mineral recovery and carbon release in the leached oven-dried biomass. Because we were more familiar with the leaching of oven-dried biomass, we based our leaching times on past oven-dried studies. However, fresh biomass may require more time to complete the leaching process. A time course for the leaching of the fresh biomass has been suggested for future study. Structural differences may have also affected the CO_2 production curves. Some of the measured microbial respiration in the fresh biomass may not have been due to soluble organic degradation, but rather to microbial respiration and/or consumption of other microbes, Also, since some of the microbial populations may have been decimated during the drying process, CO_2 production may not be an accurate representation of differences between the soluble organic degradation of the two treatments.

As mentioned, homogeneous sampling also posed a problem due to the large and varying sizes of the substrate particles. In order to resolve this, it has been suggested that the entire contents be harvested for sampling.

ACKNOWLEDGEMENT

The authors are grateful for support of this work by the following-the United States Aeronautics & Space Administration, Grant No. NCC 9-51, G.W. Carver Agricultural Experiment Station, Tuskegee University, United States Department of Agriculture/CSREES, Mike Alazaraki of Dynamac Corporation and Drs. Audrey Trotman, and Phillip Loretan of Tuskegee University.

REFERENCES

(1) AIBS-KSC Biomass Processing Technical Panel. John F. Kennedy Space Center. Biomedical Operations and Research Office. Conversion of Inedible Wheat Biomass to Edible Products for Space Missions: Final Report.

(2) Garland, J. L., C.L. Mackowiak, J. C. Sager. 1993. Hydroponic Crop Production Using Recycled Nutrients From Inedible Crop Residue. SAE Technical Paper, Series 932173,

(3) Barta, D.J. and D.C. Heninger. 1994. Regenerative Life Support Systems-Why Do We Need Them? Adv. Space Res. 14(1 1):423-412.

(4) Garland, J.L. 1992. Coupling Plant Growth and Waste Recycling Systems in a Controlled Ecological Life Support System (CELSS). NASA Tech. Memorandum 127447.

(5) Finger, B.W. and R.F. Strayer. Development of an Intermediate-Scale Aerobic Bioreactor to Regenerate Nutrients from Inedible Crop Residues. SAE Technical Paper Series 941421, Presented at the 24th International Conference on Environmental Systems (ICES) and 4th European Symposium in Space Environmental Control Systems, Friedrich Shafen, Germany, 1994.

(6) Garland, J.L. 1992. Characterization of the Water Soluble Components of Inedible Residue from Candidate CELSS Crops. NASA Tech. Memorandum 127447.

(7) Strayer, R.F. and K. Cook. 1995. Recycling Plant Nutrients at NASA's KSC-CELSS Breadboard Project Biological Performance of the Breadboard Scale Aerobic Bioreactor During Two Runs. SAE Tech. Paper 951708.

(8) Garland, J. L., and C.L. Mackowiak. 1990. Utilization of the Water Soluble Fraction of Wheat Straw as a Plant Nutrient Source. NASA Tech. Memorandum 103497.

THE DEVELOPMENT OF AN EDIBLE PEANUT PROTEIN FILM

N. Patrick, G. Jones, H. Aglan* and J. Lu

Center For Food and Environmental Systems For Human Exploration of Space
Tuskegee University, Tuskegee AL 36088
Campbell Hall

ABSTRACT

The peanut is one of the crops chosen for use in NASA's Advanced Life Support Program (ALS). The peanut is a source of both oil and protein. After oil is extracted from the peanut, a protein rich flour remains. An edible peanut protein film is one use for this flour. Two types of film are developed for this study, one set of film contains 10% fat while the other set contains no fat. For film without fat, the defatting of the peanut by the Soxhlet method is the first step in the manufacturing process of the film. Secondly, the protein is precipitated at its isoelectric point (pH 4.5) and centrifuged to separate the protein from the non-protein. After freeze-drying the protein, it is milled in a ceramic ball mill to decrease particle size and sifted through a series of sieves to determine particle size distribution. Those particles retained on the 100 mesh sieves are utilized for film formation. Larger particles are re-ground and sifted. Five grams of protein is mixed with 50 mL of distilled water, 70 mL of 80% ethanol, 15 mL of 6N ammonium hydroxide and a plasticizer. This mixture is heated for 30 minutes until the temperature reaches 70' C. The mixture is then poured onto a level Teflon coated glass surface. After allowing the film to form overnight under a ventilation hood, it is manually removed from the plate. The processes and methods adopted have created flexible films of uniform thickness that are free of air bubbles. Thickness of films made from fully defatted peanut protein and partially defatted peanut protein were 0.10 μm and 0.13 μm, respectively. Films with natural peanut fat are approximately three times as flexible and almost four times as strong as the films made without fat. Further research will be performed to evaluate its mechanical properties.

This paper will greatly contribute to food preservation and waste management. Potential applications of this film are edible/biodegradable containers, wrapping for food preservation (against water, oxygen and oil), storage packets for seasonings or other ingredients and encapsulation for pharmaceutical use.

Keywords Peanut, edible film, protein, and mechanical properties,

---

*Corresponding Author
Tel: (334) 727-8973
Fax: (334) 727-8090
E-mail: Aglanh@acd.tusk.edu

Introduction

Edible films and coatings are typically derived from lipids, proteins, carbohydrates, or combinations of the three (Gennadios and Weller, 1991). They have become of great interest for their film forming ability and the extra protection they provide against desiccation, oxidation, rancidity and other means of food deterioration (Brandenburg et al., 1993). Regardless of whether films are made from lipid, carbohydrate or protein, they are adaptive to a great number of applications. As barriers from oil, water or oxygen, they preserve the freshness of foods. In addition, they can control oil, water or oxygen activity, preventing either loss or uptake. Edible films work as inhibitors of moisture and solute migration among layers in such foods as pies, cakes or pizzas (fresh or frozen) (Gennadios and Weller, 199 1; Krochta and De Mulder-Johnston, 1997). An edible coating system, developed and patented by Kerry Ingredients, (Beloit, WI.) and Hercules, (Wilmington, DE.), known as Fry Shield, reduces the amount of fat uptake during frying by 20-40% and in addition, reduces moisture loss (Krochta and De Mulder-Johnston, 1997). They serve as packaging materials for small food items and products such as candies, dried fruits, nuts, cereals and vegetables, as flavor packets for the food industry or as encapsulation for pharmaceutical use. Because these film are made of natural ingredients they can be consumed with the packaged food (Gennadios and Weller, 1991). As binding, they may be applied to the surfaces of foods to serve as a foundation or adhesion medium for seasonings. As glazes they enhance the appearance of baked goods with shiny, glossy finishes or colorful soft or hard shell-like coatings in addition to offering flavor.

The protein content of the peanut seed is between 22 and 30 % (Ahmed and Young, 1982). When films are formed from this peanut protein, there are several factors that must be considered in order to produce a successful film with desired specifications. In relation to film yield and formation, Aboagye and Stanley (1985) found that temperature was the most significant factor. The closer the temperature to the boiling point of water, the greater the film yield and formation rate. It was also reported that alkaline pH conditions and higher protein concentration improved the film forming process. In contrast, elevated pH levels decreased film strength. For optimal film strength, a pH of 7 was recommended (Aboagye and Stanley, 1985). Peanuts and soybeans are classified as legumes and are very similar in composition and functionality. The production, characteristics and mechanical properties of soy protein films are also comparable to that of peanuts. There is limited literature on films developed from peanut protein; therefore, literature from soy protein films will also be reviewed.

Soybeans have customarily been utilized in the Far East for the production of edible films (Gennadios and Weller, 1991). In the same manner, peanut protein based films have also been formed. This procedure involved the heating of peanut milk in large open pans and the removal of films formed on the surface (Wu and Bates, 1973). The method in which soy films could be prepared from isolated soy protein was developed by Jaynes and Chou (1975). The procedure involved the spreading of an isolated soy protein slurry (pH 6.6) on Teflon-coated baking pans and drying at 100° C for 1 hr. This procedure has been proven to work for the production of other protein-based films,

Protein-based films generally exhibit poor water vapor permeability as a result of their hydrophilic nature; however, they arc very effective oxygen barriers (Gennadios and Weller, 1991; Herald et al., 1995). The addition of plasticizing agents to these films contributes to their flexibility, extensibility and workability, thereby avoiding cracking and chipping during handling and storage (Wu and Bates, 1973). There arc many available plasticizing agents, however, mono-, di- or oligo-polysaccharides, lipids and their derivatives are the most commonly used (Gontard et al., 1993). The usefulness of protein-based films exists in various categories. Balassa (1970) patented a method for encapsulation and stabilization of aromas and flavors. The greatest attribute of edible films is their natural renewability in a continuous life cycle. Recyclable or biodegradable materials eliminates the accumulation of synthetic materials thereby reducing waste. Packaging represents approximately 30% by weight of municipal solid waste, occupying approximately two-thirds of trash can volume (Hunt et al, 1990). Understandably, food packaging has become a major concern of waste reduction efforts. In situations where the accumulation of waste products may be hazardous to human health, as in space travel, the development of an edible peanut protein film for food preservation and/or packaging from the waste or byproduct of peanuts is ideal. Therefore, the objective of this research was to develop a film from peanut residue, Currently, protein films from peanuts are being developed for use in space applications where waste reduction is a critical factor. The peanut is a source for both oil and protein. After the oil is extracted for use in cooking, a protein rich residue remains. From the peanut residue, protein can be extracted and used for the production of an edible peanut protein film.

Materials and Methods

Peanut Flour: Unroasted, partially defatted peanut flour was used for this experiment (Seabrook Enterprises, Atlanta, GA). Two sets of film, (1) made with partially defatted (10% fat) peanut flour and (2) made with fully defatted peanut flour were developed for testing. Each set of film contained sorbitol as the plasticizing agent. The effect of the inclusion and the exclusion of peanut fat in addition to a plasticizer were investigated.

Defatting: Peanut flour was defatted using the Soxhlet method for solvent extraction with petroleum ether (AOAC, 1990).

Protein Extraction: One liter of distilled water was added to 100 g of defatted or partially defatted peanut flour. The pH was adjusted to 9.0 with 6N sodium hydroxide, stirred for 1 hour, then centrifuged at 3,300 rpm for 30 minutes to remove any non-protein material. The supernatant, which contained the soluble protein, was collected. To precipitate the protein, 6N hydrochloric acid was added to the supernatant until it reached its isoelectric point (pH 4.5). The solution was allowed to precipitate overnight. The precipitated protein was centrifuged at 6,000 rpm for 30 minutes. The protein pellet was washed with distilled water and centrifuged again at 6,000 rpm for 30 minutes. Finally, the precipitated protein was freeze-dried (Schnepf, 1996). The protein was milled for 2.5 hours, then sifted through a series of 40, 60 and 100 mesh sieves. The protein retained from the 100 mesh sieve was used for film formation. Protein remaining on the 40 and 60 mesh sieves were ground and sifted again until they were retained or passed through to the 100 mesh sieve.

Film Formation: In preparation for film formation, a level stand with a Teflon treated surface was created. In a beaker, 5 grams of protein, a plasticizer (2-4 g in amount), 50 mL of distilled water, 70 mL of 80% ethanol and 15 mL of ammonium hydroxide were combined. This solution was stirred on a hot plate until 70°C was reached. To remove air bubbles, the solution (protein gel) was poured onto (203.2mm x 279.4mm) Teflon treated surfaces under vacuum. After 5 minutes, the vacuum supply was disconnected and the liquid was allowed to form a film overnight under a ventilation fan.

Film Thickness: The thickness of films was measured to the nearest 0.1 μm at 4 random positions around the film, using a micrometer (B.C. Ames Co., Waltham, MA). They were cut into 0.8 -0.9 mm wide by 13.9 -14.0 mm long samples.

Film Surface Morphology: Selected film samples were subjected to a high intensity microscope. Samples were mounted on glass slides with a mounting medium and then covered with a thin glass slide sheet. Viewing of the films and photography was made possible by use of a Bright Field Ultra Microscope and camera (C. Squared Corp., Tamarac, FL).

Film Tensile Strength and Percent Elongation: Tensile strength and strain were evaluated with an Instron Universal Testing Machine (Model 48085 Instron Engineering Corporation Canton, MA) according to ASTM Standard Method E 882-88 (ASTM, 1988). The Instron was fitted with a 100-lb load cell and initial grip separation was 0.022 m.

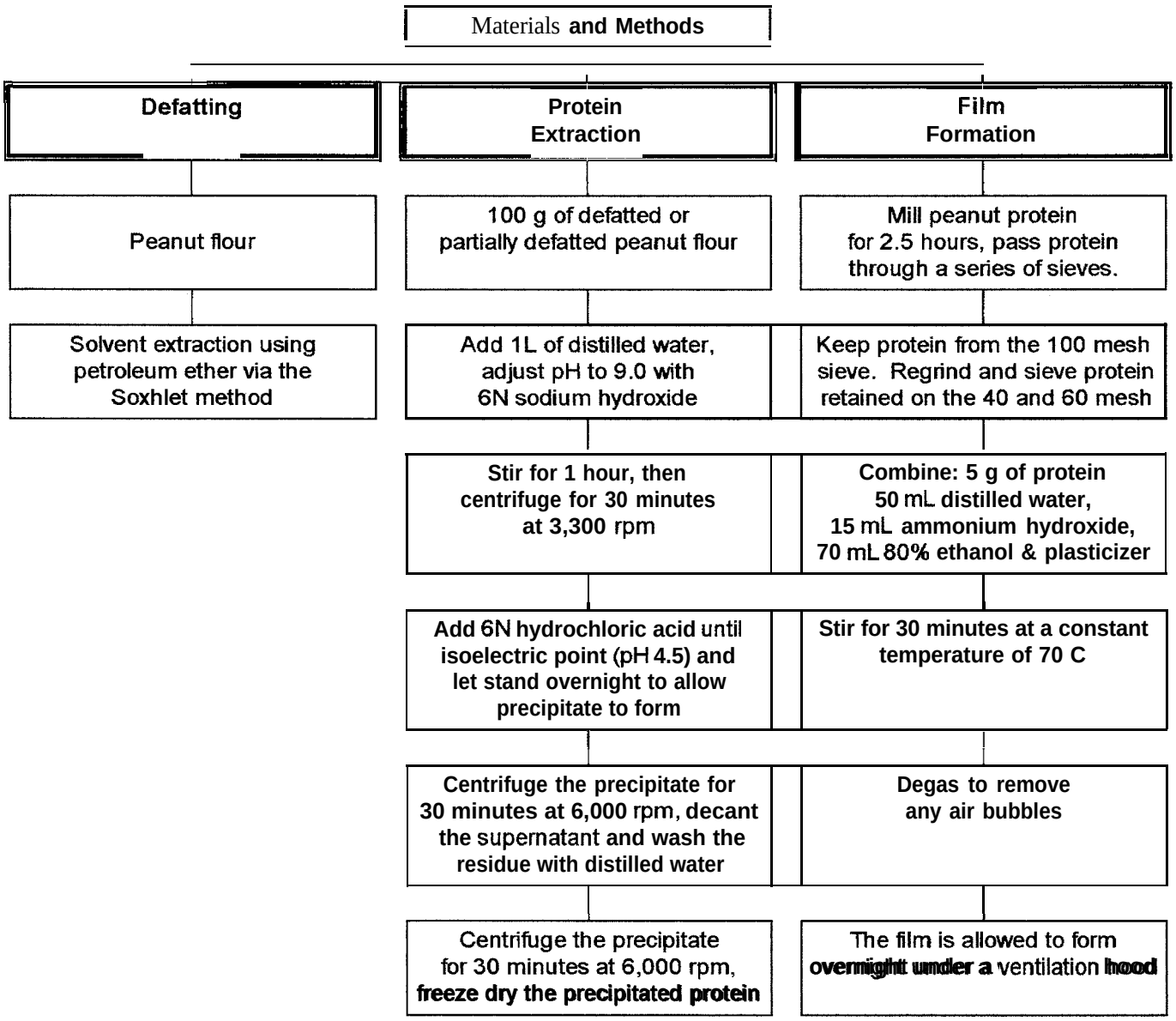


Fig. 1.- Flow chart of processing procedures for the development of an edible peanut protein film.

Results and Discussion

Appearance & Microstructure

The appearance of film made from fully defatted peanut protein were generally clear, glossy and transparent, however, more voids were present in these films than in films made from partially defatted peanut protein, which were smoother and almost free of voids. The presence of fat in these films imparts an opaque appearance, but they are also transparent. The tendency of the film forming solution to foam during solidification was less with films made with fat. An increase in lipid concentration resulted in a more compact structure, filling in the areas of possible voids (fig. 3). According to Brandenburg et al., (1993), pH affects the appearance of films. It is stated that films at pH 6 are opaque and have insoluble particles, while films with pH in the range of 8-12 are clear. The films made from peanut protein were made at pH 9 and appeared to be fairly clear.

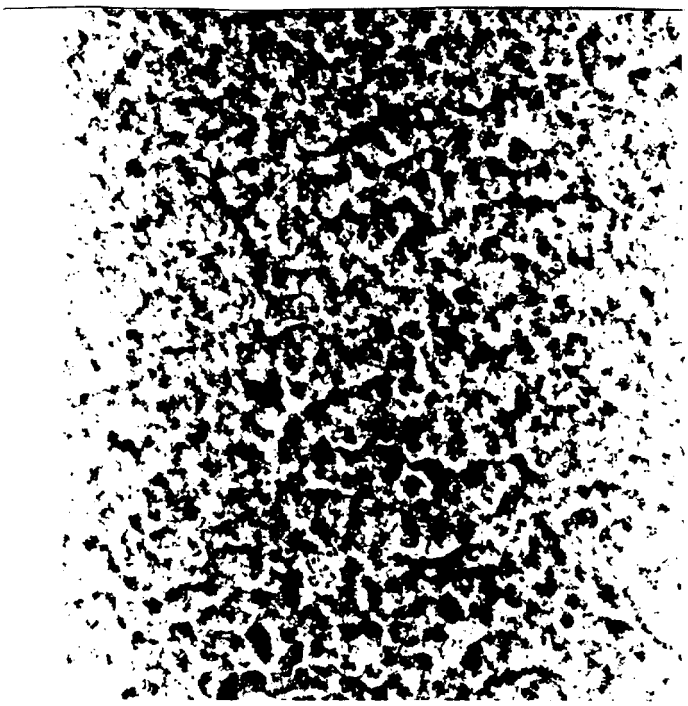


Fig. 2.- Isolated peanut protein film with fat, viewed at 355 magnification.

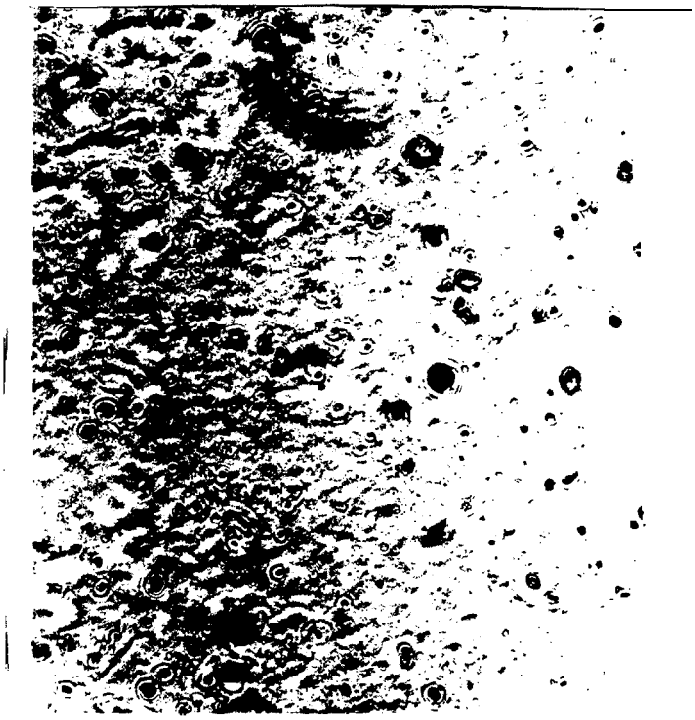


Fig. 3.- Isolated peanut protein without fat, viewed at 355 magnification.

Mechanical Properties

The thickness of films made from defatted peanut protein and partially defatted peanut protein were 0.10 μm and 0.13 μm, respectively. The stress-strain relationship for the two films is shown in Figure 4. The ultimate tensile strength of films made from partially defatted peanut protein was 7.5 MPa, while the ultimate tensile strength of films made from defatted peanut protein was 2.1 Mpa (Fig. 4). The strain values for films with fat and without fat were 6.2 % and 2.2 %, respectively (Fig. **4). The** film with fat is approximately three times as flexible and almost four times as strong as that made without fat. The modulus of elasticity for films with fat was approximately 1.2 MPa and without fat was approximately 0.8 MPa. On this basis, the film made with fat is much stronger and durable as indicated by the area under the stress-strain curve. Thus, it appears that the addition of fat enhances the mechanical behavior of peanut protein films.

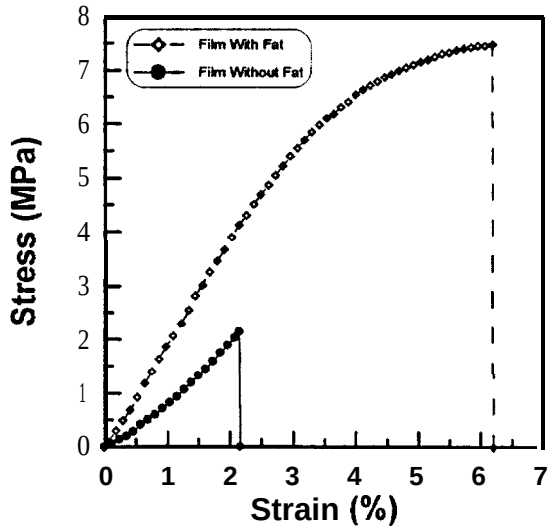


Fig. 4.- Stress/Strain relationship for edible peanut protein films with and without fat.

Conclusion

The addition of plasticizing agents to the films lends flexibility, extensibility and workability, which deters cracking and chipping during handling and storage. In addition to the plasticizing agents, the incorporation of the peanuts natural fat also contributes to these factors and increases film strength and durability. Mechanically, the tensile strength of these films is enhanced They arc approximately three times as flexible and almost four times as strong as the films made without fat (Fig. 4.). Another attribute of the addition of fat is the compactness that it provides to the microstructure of the film. Filling of the areas of otherwise possible voids, creates a more stable system. Although the appearance of the films with fat is more opaque than films without fat, they are still transparent.

Acknowledgement: This research was supported by the National Aeronautics and Space Administration (Cooperative Agreement NCC-9-5 1).

References

Aboage, Y. and Stanley, D. W. 1985. Texturization of peanut proteins by surface film formation. 1. Influence of process parameters on film forming properties. Can. Inst. Food Sci. Technol. J., 18:12-20.

Ahmed, E.M. and Young, C.T. 1982. Composition, quality and flavor of peanuts, in Peanut Science and Technology, Pattee, H.E. and Young, C. T.. eds., Yoakum, TX: American Peanut Research and Education Society. Inc.. pp. 655-688.

AOAC. 1990. Official Methods of Analysis, 15th ed. Association of official Analytical Chemists, Washington DC.).

Balassa, L.L. February 17, 1970. U.S. patent 3,495,988.

Basha, S.M.M. and Cherry, J.P. 1976. Composition, volubility and gel electrophoretic properties of proteins isolated from florunner (*Arachis hypogaea L)* peanut seeds. J. Agric. Food Chem. 24: 359-365.

Brandenburg, AH., Weller, C.L., and Testin, R.F. 1993. Edible films and coatings from soy protein. J. Food Sci. 58(5): 1086-1089.

Gennadios, A., and Weller, C.L. 1991. Edible films and coatings from soymilk and soy protein. J. Food Sci. 36(12): 1004-1009.

Gontard, N, Guilbert, S, and Cuq, J.L 1993. Water and glycerol as plasticizers affect mechanical and water vapor barrier properties of an edible wheat gluten film. J. Food Sci. 59(1): 20(+211.

Herald, T. J., Gnanasambandam, R., McGuire, B. H., and Hachmeister, K.A. 1995. Degradable wheat gluten films: preparation, properties and application.
J. Food Sci. 60(5): 1147-1150, 1156.

Hunt, R. G., Sellers, V.R., Franklin, W.E., Nelson, J.M., Rathje, W.L., Hughes, W. W., and Wilson, D.C. 1990. Estimates of the volume of MSW and selected components in trash cans and lanfills. Franklin Associates, Ltd., Prairie Village, Kansas and The Garbage Project, Tucson, AZ.

Jaynes, H. O. and Chou, W. N. 1975. New method to produce soy protein-lipid films. Food Prod Dev. 9(4): 86.

Krochta, J. M., and De Mulder-Johnston, C. 1997. Edible and biodegradable polymer films: challenges and opportunities. Food Technol. 51(2): 60-74.

Schnepf, 1996. Personal Communication. Dept. of Nutritional Science and Dietetics, Univ. of Nebraska, Lincoln, NE.

Wu, L. C., and Bates, R.P. 1973. Influence of ingredients upon edible protein-lipid film characteristics. J. Food Sci. 38: 783-787.

ABSTRACT

Development of Paper Products from Dried Sweetpotato Stems and Peanut Shells

R. McConnell, R. Smith, G. Jones' and J. Y. Lu

Center for Food and Environmental Systems for Human Exploration of Space
Tuskegee University, Tuskegee, AL 36088
Campbell Hall

One of the goals of NASA's Advanced Life Support Program (ALS) for sustaining human life in space is to achieve a closed system in plant production and usage. That means all inedible plant parts should be recycled or used in some way. A Tuskegee University team researching sweetpotato and peanut for ALS has developed paper products from dried sweetpotato stems and peanut shells. In this study, the sweetpotato stems and peanut shells were soaked separately in water for 48 hours. After 48 hours, researchers manually separated the pulp and the unusable parts. To form the paper, 160 g of pulp and water mixture was poured through a 15.1 cm (diameter) filtration funnel and the pulp was trapped on 15 cm (diameter) falter paper. The falter paper and pulp were dried in an air oven, and the filter paper was removed, An examination under a scanning electron microscope showed that the sweetpotato paper was composed of "fibers", whereas the peanut shell paper was composed of "blocks". Results of physical testing showed that the sweetpotato stem paper was stronger than the peanut shell paper. It is anticipated that there may be other uses of these products such as writing paper, bags and packaging material. Because of its biodegradability, it can be incorporated into the resource recycling system at the end of its use.

Keywords: Sweetpotatoes, peanuts and paper

* Corresponding Author
Tel: (334) 727-8955
Fax: (334) 724-4451
E-mail: G_Jones@acd.tusk.edu

Development of Paper Products from Dried Sweetpotato Stems and Peanut Shells

Introduction

Sweetpotatoes and peanuts are two crops that NASA has proposed to grow in space. Sweetpotato stems and peanut shells are part of the inedible portion of the plant. One of the goals of NASA's Advanced Life Support (ALS) Program is to utilize the entire plant. Thereby, have zero waste. In an effort to accomplish this goal, a Tuskegee University research team has developed paper products from dried sweetpotato stems and peanut shells. It is proposed that the paper produced will be used as a medium for seed germination. Because the paper is biodegradable, after its use, it can be incorporated into the resource recycling system.

Paper is defined as a felted sheet formed on a fine screen from a water suspension of fibers (Smook, 1992). As a papermaking raw material, wood is a relative newcomer. For nine-tenths of its history, paper was made almost exclusively from non-wood fibers. Although wood now supplies 93% of the worldwide papermaking raw material, some developing countries do not have an adequate supply of wood. However, many of these countries do have large quantities of agricultural residues and other non-wood plant fibers which can be used for paper making (Atchison and McGovern, 1993).

The highest ratios on non-wood plant pulping capacity to total papermaking pulp are found in the developing market economies of Asia, Africa and Latin America. Some of these regions have greater non-wood pulping capacity than wood-pulping capacity. Furthermore, total non-wood plant pulping capacity worldwide is increasing faster than wood pulping capacity (Atchison, 1993a).

Non-wood plant fibers will play an increasing role in the world's pulp and paper industry. It appears that two billion tons of non-wood material could be available for paper-making. It has been shown that by selecting the appropriate mixture of non-wood

plant fibers and appropriate pulping process, any grade of paper and paperboard can be produced without the addition of any wood pulp (Atchison, 1993b). This is especially good news for developing countries and countries that do not have adequate domestic supply of wood fibers.

The objective of this research project was to develop two types of paper products from inedible plant materials. One was made from sweetpotato stems and the other was made from peanut shells. Both products were made in the same manner.

Materials and "Methods

Paper Making

Sweetpotato stems and peanut shells were soaked separately in tapwater for 48 hrs (Fig. 1). The pulp was removed with a hand mixer and separated from the unusable parts. To form the paper 160 g of a pulp and water mixture was poured through a 15.1 cm (diameter) filtration funnel. The pulp was trapped on 15 cm (diameter) Whatman's #4 filter paper. The filter paper and pulp were dried in an air oven, and the filter paper was removed.

Evaluation of Physical Properties

The structure of the paper was examined under a Hitachi S-2150 Scanning Electron Microscope (SEM) (Hitachi, LTD., Tokyo, Japan). The tensile strength was determined by using the Instron Universal Machine (Model 48085; Instron Engineering Corporation, Canton, MA).

Results and Discussion

Results of the SEM showed that the sweetpotato paper is composed of long fibrous materials (Fig. 2). The peanut shell paper was composed of "blocks". Results of the

Instron Universal Testing Machine showed that the sweetpotato paper was much stronger than the peanut shell paper.

Currently this paper is being evaluated as a medium for seed germination. If this product works as a germination medium, the remaining material can be degraded in a bioreactor. This would aid in utilizing the entire plant.

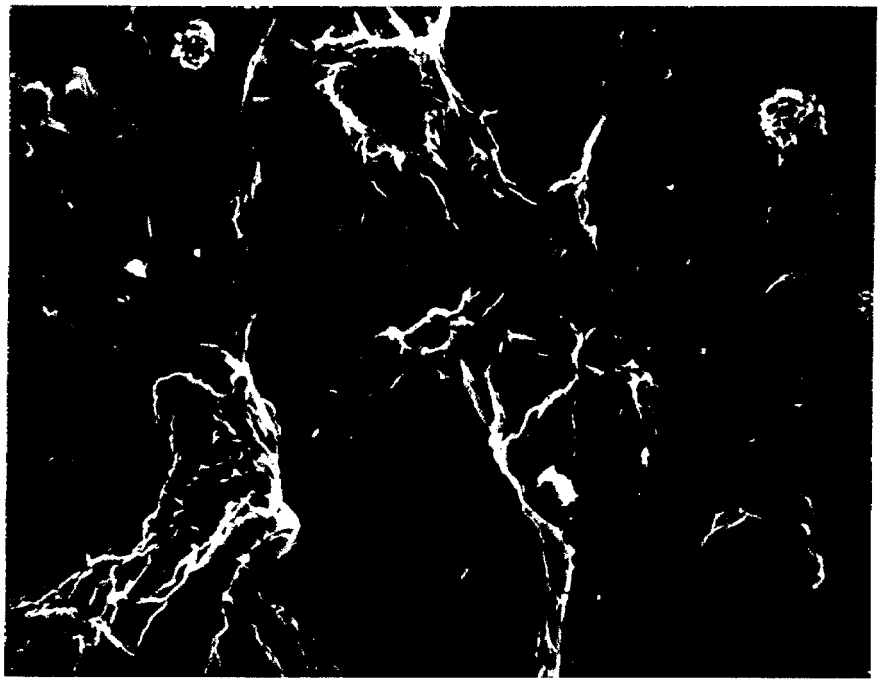


Fig **2.** Microstructure of sweetpotato paper at 500 magnification.

Conclusion

Paper products have been developed from sweetpotato and peanut in support of ALS objectives for lunar/Mars missions. Development and testing of this product will continue. However preliminary testing indicates that they provide viable materials for use in space while utilizing inedible plant parts. As we continue to develop this paper product, other uses may be identified. However, we are currently concentrating on physical properties and the use of this paper as a seed germination medium.

References

Atchison, J.E. and McGovern, J.N. 1993. History of paper and the importance of non-wood plant fibers. Ch. 1 in *Pulp and Paper Manufacture: Secondary Fibers and Non Wood Pulping.* 3rd ed. TAPPI, Atlanta, GA.

Atchison, J.E. 1993a. Data on non-wood plant fibers. Ch. 2 in *Pulp and Paper Manufacture: Secondary Fibers and Non-Wood Pulping.* 3rd ed. TAPPI, Atlanta, GA.

Atchison, J.E. 1993b. The future of non-wood plant fibers in pulp and papermaking. Ch. 3 in *Pulp and Paper Manufacture: Secondary Fibers and Non- Wood Pulping.* 3rd ed. TAPPI, Atlanta, GA.

Smook, G.A. 1999. *Handbook for Pulp & Paper Technologists.* 9th ed. Angus Wilde Publications. Bellingham, WA.

Acknowledgments

This research was supported by funds from the National Aeronautics and Space Administration (Grant No. NCC 9-51) and USDA/CSREES/Evans-Allen (Grant No. ALX-FS-2).

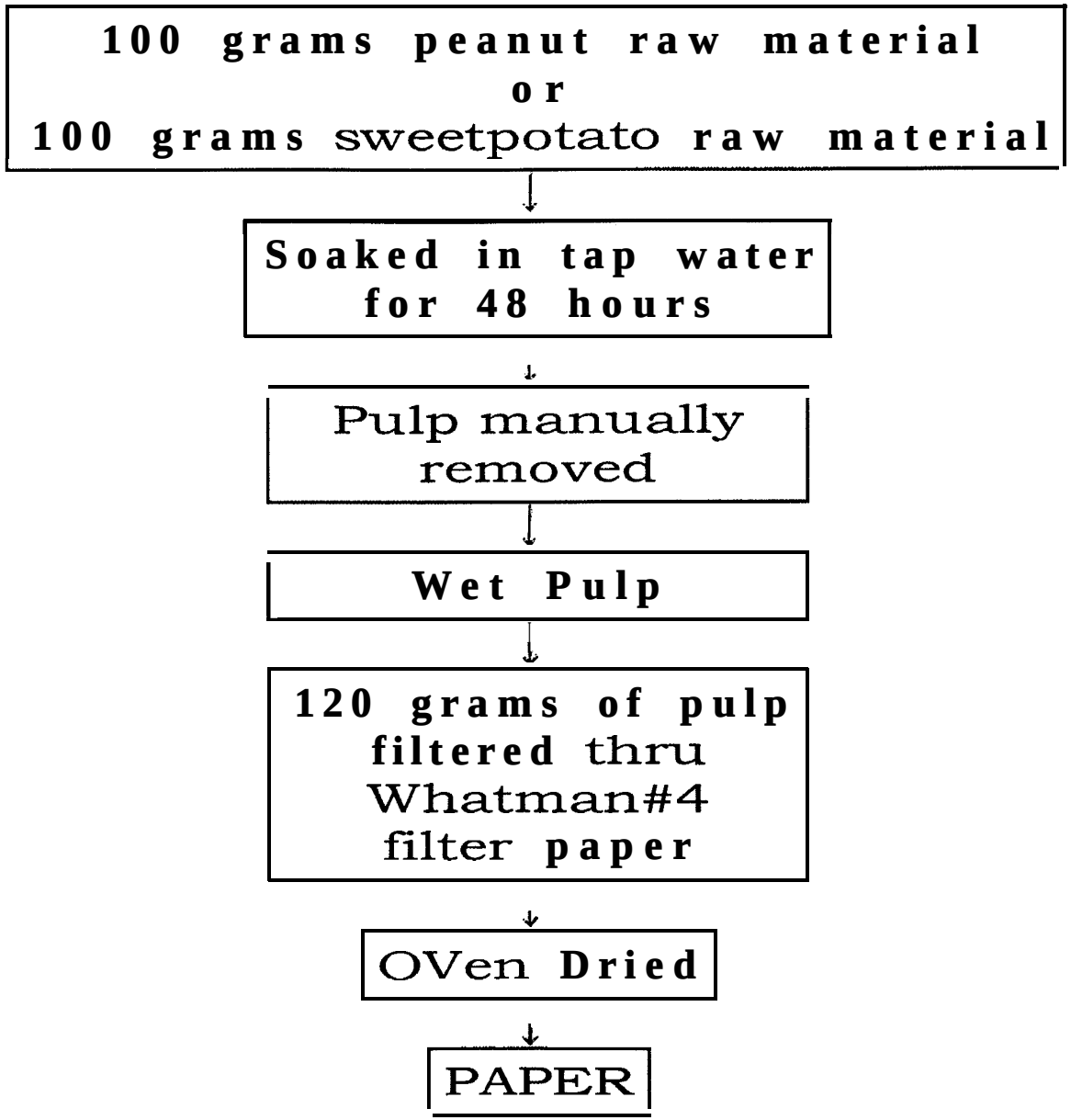


Figure 1. Peanut/Sweetpotato Paper Production

SYSTEMS CHARACTERIZATION OF TEMPERATURE, PH AND ELECTRICAL CONDUCTIVITY IN AEROBIC BIODEGRADATION OF WHEAT BIOMASS AT DIFFERING MIXING RATES

M. Calhoun, A. Trotman*, and H. Aglan
Center for Food and Environmental Systems for Human Exploration of Space (CFESH)
Tuskegee University, Tuskegee, AL 36088

ABSTRACT

The purpose of this preliminary study is to observe and relate the rate of mixing to pH and electrical conductivity in an aerobic, continuously stirred bioreactor. The objective is to use data collected from successive experiments as a means of a system characterization. Tests were conducted to obtain these data using a continuously stirred 20 L Cytostir glass reaction vessel as a bioreactor operated without built-in temperature or pH controls. The tests were conducted on the lab bench at ambient temperatures. The substrate in the bioreactor was ground wheat biomass obtained from the Biomass Production Chamber at NASA Kennedy Space Center. In this study, the data reflect characteristics of the native (uninoculated) systems as well as inoculated systems. In the native systems, it was found that pi-l levels became stable after approximately 2 to 3 days. The electrical conductivity levels for the native systems tended to decrease over time. In contrast, ion activity was increased after the introduction of bacteria into the system. This could be correlated with the release of nutrients, due to the activity of the bacteria. Also, there were slight increases in pH in the inoculated system, a result which is expected for a system with no active pH controls. The data will be used to test a mathematical model in an automated system.

Keywords: aerobic, biomass, biodegradation, bioreactor, electrical conductivity, pH, resource recovery

INTRODUCTION

A system to recycle inedible components of crops is vital to the effort for long-term space travel. The design of such a system would require that the operating parameters are reliant on baseline data indicative of the behavior of the system. It has been demonstrated that biological recovery processes are a viable alternative to physical/chemical methods. First, they require little pre-processing and need no significant consumables. Also, they use less power and yield greater mass savings than the alternatives (Supra et. al., 1997). For these reasons, a major thrust of the waste management and resource recovery working group of the NASA Center for Food and Environmental Systems for Human Exploration of Space (CFESH) at Tuskegee University bioregenerative involves methods for solid waste degradation,

* Corresponding author. Fax: (334) 724-4451, Trotmana@Acd.Tusk.Edu

Stirring is the primary source of energy addition to the system. The speed of stirring promotes a homogeneous substrate mixture, and it has been shown that continuous stirring increases the rate of degradation (Illmer et. al., 1997). It would seen then, that higher speeds are most desirable; but at high speeds, frothing tends to occur. Inaccurate measurements could occur in a system with controls as a result of frothing, because any addition of acids, bases or other substances may become suspended in the froth. Conversely, high loads at low speeds cause drag and possibly premature fatigue of motor parts. The rate of mixing under such conditions could cause inconsistent speeds or complete motor failure. For this study, a speed range was chosen to ensure that neither frothing nor engine drag would occur.

in addition, the use of continuously stirred bioreactors are efficient and reliable methods for the biodegradation of solid waste materials (Strayer et. al., 1997). They are efficient due to the fact that the increased rate of biodegradation facilitates an increase in the rate of release of nutrients ideal for growing crops hydroponically. This type of system is reliable because it tends toward steady state conditions even if there are fluctuations in temperature, pH or other parameters.

Aerobic biodegradation of solid waste in solution utilizes microorganisms as the decomposing agent to metabolize gross particles for the ultimate release of nutrients. The microbial activity that occurs during the process may result in the depletion of oxygen in the system. In closed systems such as these, finite amounts of resources are present. Microbes use oxygen in the biodegradation process, therefore the overall levels of oxygen in the system decrease. Thus, the addition of oxygen is necessary in the biodegradation process (VanderGheynst et. al., 1997; Supra et. al., 1997). An aerobic process involves the addition of oxygen. This is sometimes achieved by the use of an air pump with an attached tube emitting oxygen into the system. However, a system which allows the exchange of even small amounts of gases could be called aerobic in ambient conditions, provided that at least 50% of the gas mixture in the system is air. This is the type of aerobic biodegradation system that was used for these experiment.

The objective of this study was to gather data that could be used to describe a system which uses aerobic biodegradation, and *a* predetermined range for continuous rate of mixing (speed). The data were compared with findings of previous studies, in order to establish the validity of the results. All data will be used in the future to test a computer model with an automated system.

METHODOLOGY

A 20 L Kontes Cyto-stir bioreactor with two arms and a mechanical impeller was used as the vessel for all experiments. A 15 L quantity of a basal salts medium (BSM) was prepared in a separate vessel, and the pH was adjusted to within a range of 6.8-7.0. This is a range that will favor microbial degradation. The liquid to biomass ratio for all experiments was 10:1 by volume, so 1500 mL of milled wheat biomass was prepared as the substrate in the bioreactor for each. The wheat biomass was obtained from the Biomass Production Chamber at NASA Kennedy Space Center in May, 1997. The wheat particles were too large for the experiment, so the biomass was milled a second time, using a #20 mesh screen. The twice-milled biomass was added to the bioreactor and autoclave prior to experimental set up.

Two 5-day experiments were conducted to characterize the native or baseline system. The speed setting was 100 rpm with 445.5 g dry weight of wheat biomass for the first experiment, and the speed was set at 156 rpm with 453.3 g dry weight of wheat biomass for the second experimental run. The baseline system gives information about the characteristics of the native system, therefore there was no addition of bacteria. The other tests were conducted for 14 days

each and were inoculated with 1 mL of a bacterial culture designated as WDST 3A (*Serratia marcesens*), in triptic soy broth (TSB). The inoculated culture was incubated for 40 hours prior to inoculation of the bioreactor. The bacterial culture was added to the system on Day O and Day 8. The bacterial culture WDSt 3A was isolated from the sweet potato fields on the Tuskegee University Carver research farm. For the two 14- day inoculated tests, the speed was set at 100 rpm with 479.2 g dry weight of wheat biomass for the first, and the speed was set at 156 rpm with 475.8 g dry weight of wheat biomass for the second of these experiments.

All of the materials for the experiments were autoclave (at 121 °C, 15 psi, for 30 min.) prior to set up. That included all parts and accessories: bioreactor with biomass, BSM, rubber tubing and rubber plugs. After cooling, all items were placed under a Laminar hood until set up. The set up entailed pumping the BSM into the bioreactor using a peristaltic pump and sterile tubing. After addition of the medium, the leads for the pH, electrical conductivity and thermocouple probes were inserted into a slotted, sterile #6 rubber plug that was secured into one arm of the bioreactor. The unwashed bacterial cells (> 107 cfu/mL) in nutrient broth were then added for all experiment runs requiring inoculation. The mechanical impeller was then set to the appropriate experimental speed setting and the system was left to incubate for 24 hours. Readings were taken every 24 hours thereafter.

RESULTS AND DISCUSSION

Data were collected from four experiment runs. From the data gathered, comparisons were made between the baseline and inoculated systems, as well as between the two speed settings. Figures 1 and 2 show pH vs. time for the baseline and inoculated systems respectively. The pH levels in the baseline systems stabilize after an adjustment period, becoming nearly constant after 3 days in each case. The inoculated systems showed a slight upward trend after the introduction of bacteria, which is expected for aerobic biodegradation in the absence of pH controls (Strayer et. al., 1997). Yet pH does become more stable over time. As the literature suggests, these types of systems tend to return to steady state conditions over time, even when there are fluctuations in pH due to external reasons (Strayer et. al., 1997). The data collected support that claim.

Figure 3 shows electrical conductivity (EC) vs. time for the baseline system. The graph shows a downward trend, indicating that EC decreases over time in both experimental runs for the baseline system. It should be noted though, that both curves would flatten before they reached 0μS, because there will always be some measurable amount of ion activity in these systems. Conversely, for the inoculated systems, the graph of EC vs. time (Figure 4) generally shows an upward trend. In both experimental runs, after an adjustment period of about 48 hours, EC levels increased overall. This is due to increased ion activity due to enhances degradation by the release of enzymes from the bacterial inoculum. As the enzyme supply is depleted, EC levels begin to stabilize, and move toward steady state conditions. Hence, the reintroduction of inoculum is necessary to maintain a desired level of ion activity.

Figures 5 and 6 show ambient temperature vs. time for baseline and inoculated systems respectively. A temperature increase in the systems would be expected, particularly if one or both of the following hold: 1) the mixing speed is increased; 2) inoculum is present. However, in the absence of temperature controls, any slight increase in temperature is dissipated through the system. So, it should be emphasized that these are ambient temperatures recorded inside the bioreactor, and that temperature fluctuations that occurred inside the bioreactor came about as a result of temperature spikes in the laboratory. During both of the baseline experiment runs as well as the 156 rpm inoculated experiment run, there were such spikes recorded, while during the 100

rpm inoculated experiment run, there were more constant temperatures recorded. In spite of this irregularity though, all of the data supports previous findings.

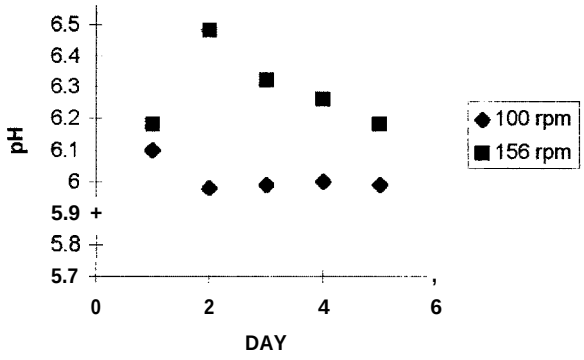


Figure L Graph of pH vs. time for baseline system.

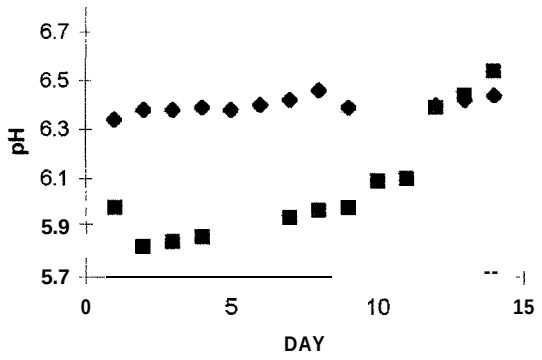


Figure 2. Graph of pH vs. time for inoculated system.

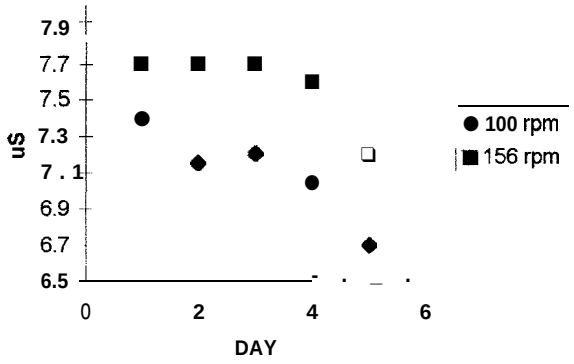


Figure 3. Graph of EC vs. time for baseline system.

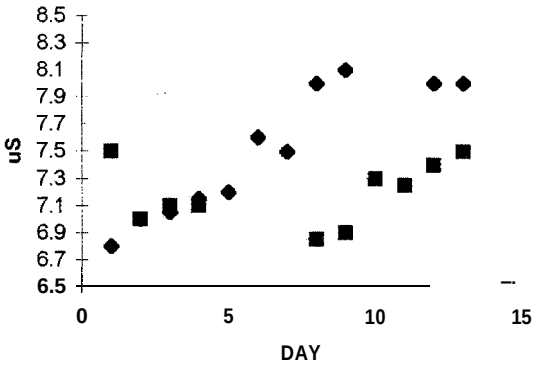


Figure 4. Graph of EC vs. time for inoculated system.

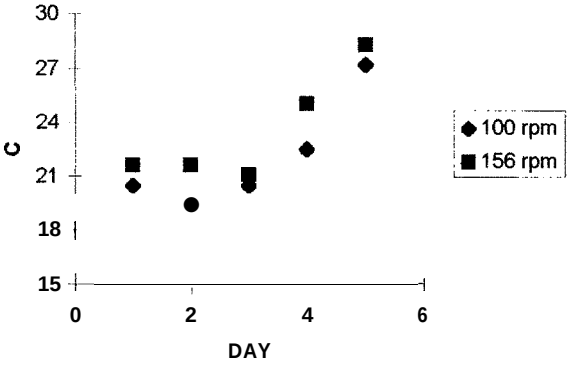


Figure 5. Graph of Ambient temperature vs. time for baseline system.

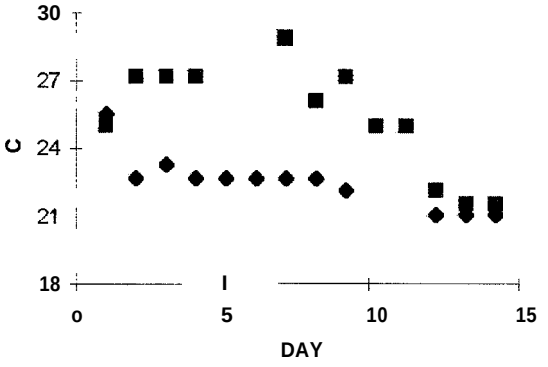


Figure 6. Graph of Ambient temperature vs. time for inoculated system.

CONCLUSION

This study was conducted in order to characterize an aerobic biodegradation system. All data collected were compared with the results of previous studies. The hypothesis was that is a system were run at relatively high speeds for a long period of time, and experienced unplanned fluctuations in pH and temperature, yet maintained steady state conditions, then a system which was run at lower speeds for a shorter period of time should maintain steady state conditions even if unplanned fluctuations in pH and temperature occur. In this study, unplanned fluctuations in pH, electrical conductivity and ambient temperature all resulted in a return to steady state conditions after a period of adjustment. It should be noted that the experiments conducted for this study were both very short-term and used only a moderate speed range. Thus, the time and amount of fluctuation were restricted by speed and time constraints. [n a previous study conducted at the Kennedy Space Center breadboard facility, it was also determined that unplanned fluctuations in pH, dissolved oxygen and temperature resulted in a return to steady state conditions (Strayer et. al., 1997). The time period and speed range in this control study were both greater than that of the study conducted at the CFESH facility at Tuskegee University, so the range of fluctuation was not as constrained by time and speed. In both studies, most fluctuations fell within the known tolerance range of bacterial processes, so the return to steady state conditions was expected. Therefore, it was determined that the results obtained from experiments conducted for this study were supported by data previously recorded.

These preliminary test results give an indication as to how bioreactors should behave under the same or similar conditions. The data gathered to date have provided the framework. for this goal. Current and future research will supply additional information about biodegradation over longer periods of time. All data gathered will be used to establish an acceptable range of pH, electrical conductivity, temperature; and later, mineral content, nitrogen levels and dissolved oxygen levels in the production of a filtrate suitable for delivery to a hydroponic growth system.

ACKNOWLEDGEMENTS

This research was supported by funds from the National Aeronautics and Space Administration (grant # NCC9-51) and the USDA/CSREES (grant # ALX-FS-2).

REFERENCES

1) Supra, L.N., M. Reddig, M.A. Edeen, K.D. Pickering, and N.J. Packham. 1997. Regenerative water recovery system testing and model correlation, SAE Paper 972550, *27th International Conference on Environmental Systems*, Lake Tahoe, NV, July 14-17, 1997.

2) Illmer, P. and F. Schinner. 1997. Compost turning- a central factor for a rapid and high-quality degradation in household comporting, *Bioresource Technology, 59:157-162.*

3) Strayer, R., B.W. Finger, and M. Alazraki. 1997. Stability and reliability of biological reactors, SAE Paper 972549, *27th International Conference on Environmental Systems,* Lake Tahoe, NV, July 14-17, 1997.

4) VanderGheynst, J. S., L.P. Walker, and J. Parlange. 1997. Energy transport in a high-solids aerobic degradation process: mathematical modeling and analysis, *Biotechnol. Prog., 13:23%-248.*

Applied Statistical Methods to Predict the Northward Migration of Africanized Honey Bees

Gabriel Vargas, Margarita Muniz
Pan American Center for Earth and Environmental Studies
University of Texas at El Paso
El Paso, Texas 79968
e-mail: gvargas@utep.edu

Africanized Honey Bees have been slowly migrating north from South America since the 1950's. Miles and miles make up their migration path that seems to ignore boundaries, crossing the Texas Mexico border in 1990. Many counties in the lower valley of Texas became infested zones and serious attacks, some including one thousand plus stings on humans were reported in some areas. Our research team is committed to find out what will stop the northward migration pattern. We believe that there exists natural phenomena which we consider are and will be variables that alter their migration pattern.

The State of Texas has been broken into ten ecological areas to search for characteristics that are exclusive to each. Variables such as annual average temperature, precipitation, water bodies, vegetation, and the number of successive days between frosts have been careful] y selected to start our research. The northern portion of Texas contains a unique feature that the AHB's have not encountered in their long journey. This feature is colder temperatures than those found in the South. We are working on a model that will consist of a series of experiments that will conclude with a high percent of certainty that temperature is of high impact on the AHB's migration pattern.

To aid in the tracking of the AHB's ground cover, the Spans Explorer, a Geographical Information System (GIS) tool is being utilized. The Spans Explorer software transforms our PC into a library of data, which is collected from various resources, This database is enriched with information for each of the 255 counties in the state of Texas that can be analyzed and correlated.

Our research is based on information which, can be extracted from geological maps. Information on temperature, precipitation, water bodies, vegetation, and the number of successive days between frosts are patiently taken for each county. Our database is setup for user input of all county information. The various variables are compared and correlation between them is determined.

The variables mentioned for our research were gathered for all areas whether they are mountains and basins, plains, coastal areas, prairies, or piney woods. A total often separate areas divide the entire state and these ten areas are used as the classification scheme for the collected data. These ten areas are; 1. Piney woods, 2. Gulf prairies and Marshes, 3. Post oak Savannah, 4. Backland prairies, 5. Coast Timbers and Prairies, 6. South Texas Plains, 7. Edwards Plateau, 8. Rolling Plains 9. High plains, and 10. Trans Pecos and Mountains. All information pertaining to annual precipitation, temperature, the number of frost-free days, and the amount of surface lake acres have been entered to the Spans software for each and every county in each of the ten ecological areas.

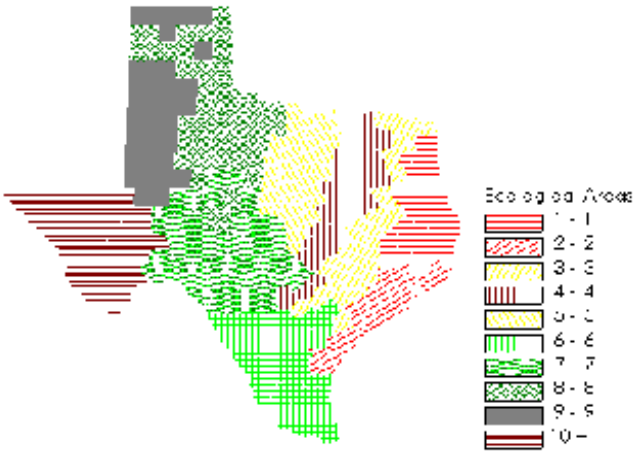


Figure 1: Ten Ecological Areas

This software provides desktop mapping which makes Figure 1 possible. Desktop mapping allows the information to be displayed on the computer screen with a geographical dimension. In a state with two hundred and fifty five counties it is extremely difficult to keep track and remember where exactly each county is. With the aid of geographic representation of the data, the quantities in the database can be pin pointed in an exact location on the map. With tenacity the state slowly enlightens with information. Bunches of classified data can be seen instantly to determine its proportion among the different ecological areas.

The software provides a Query option to enable searches of specific quantities. In the database the searched information is highlighted to distinguish it from the information not selected. In another window the results of query can be seen in a geographical representation. The map display provides a better understanding of the magnitude of the information than a simple tabulation of numbers. A query example can be a search for those counties that have had bee reports. The desktop mapping feature will display the result as shown in Figure 2 after the software manipulates the database and graphically provides those locations of Texas that have had AHB's sightings.

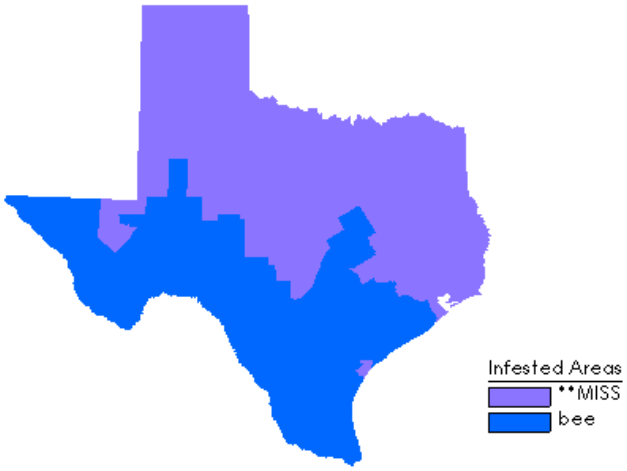


Figure 2: Results of the Africanized Bee Query

The numeric data in the database is used to perform statistical analysis. The Students T Test was chosen as the first of a series of experiments implemented on the variables and those places where the bees have been reported. The T -Test results yield the percentage of an event happening or not happening. The event consisted of one degree of change between the variable and the location where it was determined if the variable could be accepted as a migration factor or not. To implement this test the mean, standard deviation, and the number of counties with bees and without bees in the same ecological area are considered.

For each of the ten ecological areas the number of counties were noted for the presence of AHB's. Two separate categories were established which included those counties with bee reports and those without. Each of the variables mentioned above, temp, rainfall, etc., was a test for the two cases. Matlab (a high performance mathematical computational program) was used in the crunching of numbers required in the T-Test. The results from the T-Test are in the form of reject and do not reject. Reject results reinforce the variable as an established factor contributing to the northward migration pattern and do not reject results discard the possibility of any correlation to the pattern. From the T test we can safely conclude the following.

	Avg. Annual Temperature	**Avg. Annual Precipitation**	**Frost Free days**	**Surface water acres**
Ecological area 1	D N R	D N R	D N R	D N R
Ecological area 2	D N R	D N R	D N R	REJECT
Ecological area 3	REJECT	REJECT	REJECT	D N R
Ecological area 4	REJECT	D N R	REJECT	
Ecological area 5	D N R	D N R		
Ecological area 6	REJECT	D N R		D N R
Ecological area 7	D N R	REJECT	REJECT	D N R
Ecological area 8	D N R	D N R	REJECT	D N R
Ecological area 9	REJECT	D N R	REJECT	REJECT
Ecological area 10	D N R	D N R	D N R	

Table 1: Student’s T Test Result

The T-Test results show no correlation between the amount of water contained in an ecological area and the bee migration. As expected the number of days between frosts is correlated to where the bees are found. Referring to the temperature results of table 1 the temperature variable seems to have failed having more do not rejects present than rejects. This is not to conclude that the temperature variable failed. AHB's are only present in the south, which is not the majority of the state. The counties in the ecological areas that cover the south tested as rejects which translates to correlation of temperature and bee migration.

The average temperature for bee infested and non infested counties are shown for ecological areas 2,3,4,7

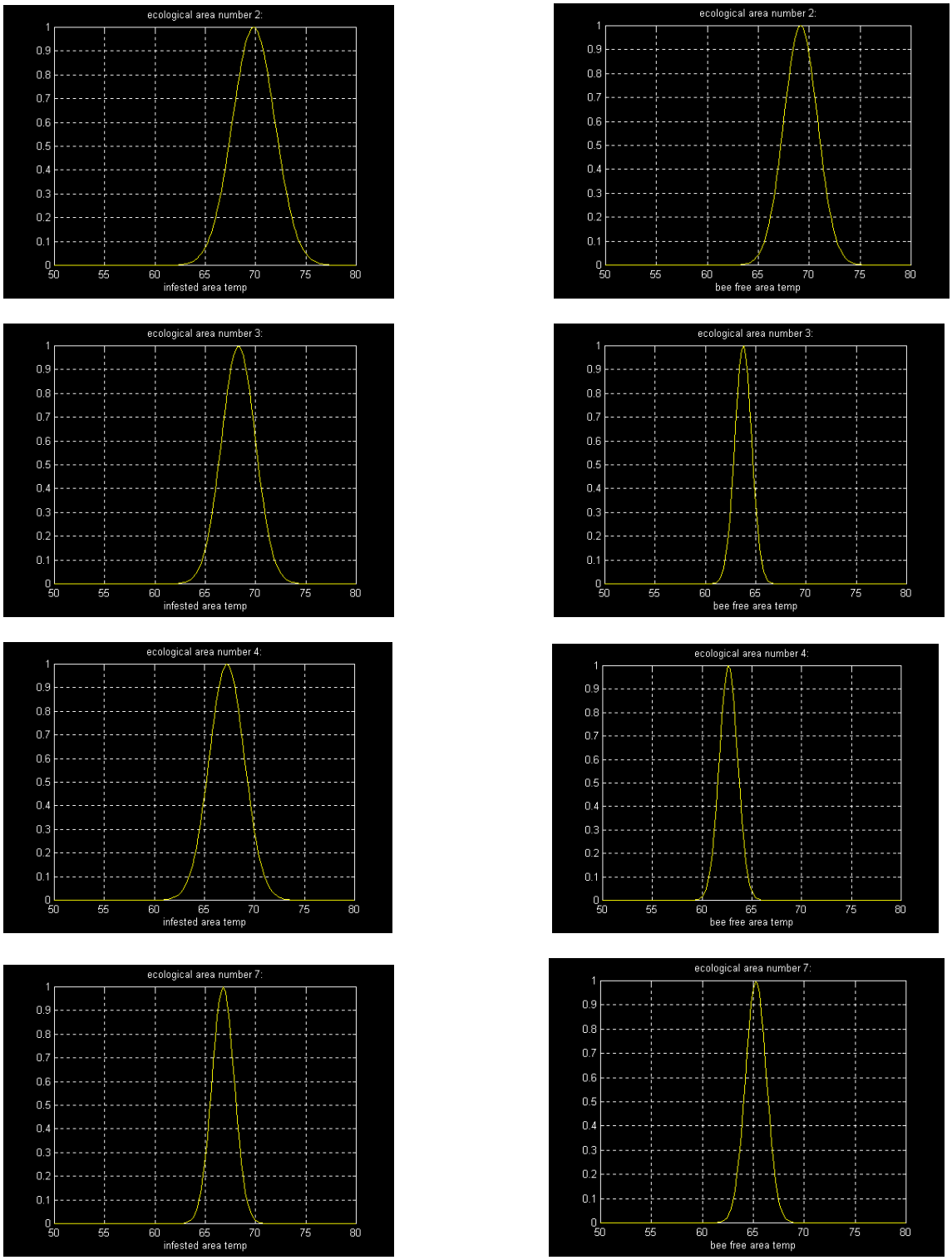


Figure 3: Matlab Results

The MATLAB results in Figure 3. show that for the same ecological area, bees were reported in counties that averaged a higher annual average temperature. The average temperature of the bee free areas is always lower than those of the infested areas. Ecological areas 8, 9, 10 also have this characteristic. Ecological areas 1, and 5 did not have any bee reports and their average temperature value were 65.5°F and 64.17°F. Ecological area 6 counties all had reports with a temperature of 69.65 °F. This encourages us to proceed with our intuition believing that temperature is a strong factor. The following table summarizes the temperature vales vs. ecological areas.

Ecological Areas	Total Counties	Reports	Mean Temp	No Report	Mean Temp
1	24	0	0	24	65.56
2	15	9	69.81	6	69.16
3	28	8	68.3	20	63.76
4	26	9	67.24	17	62.65
5	24	0	0	24	64.17
6	24	24	69.65	0	0
7	33	19	66.77	14	65.26
8	40	1	65	39	58.33
9	29	4	60	25	57.3
10	12	9	62.97	3	60

Table 2.

Future plans include implementation of alternate statistical analysis methods on the same variables to reiterate the results of Student's T-Test. More variables are being searched for to determine if other factors also contribute to the northward migration pattern of the AHB's.

ACKNOWLEDGEMENTS

We acknowledge our director Dr. Scott Starks and two members of the Pan American Center for Earth and Environmental Studies, Raul Reyes and Richard Duran, for their contribution to our study.

REFRENCES

[1] Kennamer, L., and Reese, J.V., Texas-Land of Contrast its History and Geography, Austin, Texas, W. S. Benson 1972
[2] University of Texas, Texas Looks Ahead, The University of Texas Board of Regents, Austin, 1994.
[3] State Department of Highways and Public Transportation, Cities and Towns of Texas Transportation Planning Division Austin Texas in Cooperation with the U. S. Department of Transportation Federal Highway Administration 1986.

Lori Alvarado, Dr. Ann Q. Gates, Bob Gray* and Raul Reyes
Department of Computer Science
*Center for Environmental Resource Management
The University of Texas at El Paso
El Paso, TX 79968
{ alvarado, agates, rreyes } @cs.utep.edu bob@utep.edu

Abstract

Tilting the Balance: Climate Variability and Water Resource Management in the Southwest, a regional conference *hosted by the Pan American Center for Environmental Studies, will be held at The University of Texas at El Paso on March 2-4, 1998. The conference is supported through the US Global Change Research Program (USGCRP) established by the President in 1989, and codified by Congress in the Global Change Research Act of 1990. The NASA Mission to Planet Earth program is one of the workshops sponsors. The purpose of the regional workshops is to improve understanding of the consequences of global change. This workshop will be focused on issues along the border and the Rio Grande River and thus will bring together stakeholders from Mexico, California, Texas, New Mexico, Arizona and Colorado representing federal, state, and local governments; universities and laboratories; industry, agricultural and natural resource managers; and non-governmental organizations. This paper discusses the efforts of the NASA PACES center create a GIS-based decision-support system that can be used to facilitate discussion of the complex issues of resource management within the targeted international region.*

1. Introduction

A regional workshop, *Tilting the Balance: Climate Variability and Water Resource Management in the Southwest* will be hosted by the Pan American Center for Earth and Environmental Studies (PACES) at the University of Texas at El Paso on March 2-4, 1998. The workshop is supported through the US Global Change Research Program (USGCRP) established by the President in 1989, and codified by Congress in the Global Change Research Act of 1990. NASA PACES, National Oceanic and Atmospheric Administration (NOAA), Environmental Protection Agency (EPA), US Department of Agriculture (USDA), and US Geological Survey (USGS) are sponsoring the workshop. The focus of the workshop is on the effect of short-term variations in climate, caused by droughts and El Niño, on arid and semiarid basins in the Southwest region of the United States, and long-term changes due to global climate change. The limited water supply and the health-related water issues that are pertinent to the fast-growing Southwest region makes the workshop and the need to discuss plans for the future highly relevant. White papers and presentations on the effect of past, present, and projected climate changes will be given. In addition, there will be breakout sessions that cover the following topics: agriculture, tools for understanding change, public health, drought preparation and management, and environmental education [1].

The goal of this regional workshop is to improve the understanding of the consequences of global change to the Southwest region. The workshop will bring together stakeholders from Mexico, California, Texas, New Mexico, Arizona and Colorado representing:

- federal, state, and local governments,
- universities and laboratories,
- industry, agricultural and natural resource managers, and
- non-governmental organizations.

This paper discusses the efforts of NASA PACES to collect and integrate data using satellite images, ground studies, gravity data, and digital line graphs (DLGs) to build a prototype decision-support system that can be demonstrated at the workshop. This activity serves several purposes: 1) to demonstrate the capabilities of a decision-support system with a Geographic Information System (GIS) as its base, 2) to showcase the range of data available through PACES, and 3) to illustrate the impact of integrating data from various sources and its role in facilitating discussion.

Section 2 of the paper discusses the goals and roles of NASA PACES. Section 3 summarizes the requirements of a decision-support system and the effort of PACES in developing such a prototype. The paper ends with a summary.

2. **Overview: Pan-American Center for Earth and Environmental Studies**

The objective of NASA PACES at the University of Texas at El Paso is to develop and maintain a repository of information acquired from NASA and other sources through remote sensing, ground studies or paper-based maps in an effort to engage in and advance the research areas that are relevant to NASA's Mission to Planet Earth program. The repository being established by PACES contains geographical, geological, and environmental information that covers various regions of Mexico and the southwest region of the US The role of the center is to provide access of this information to other government entities in the US and Mexico, and research groups from universities, national laboratories and industry [6] .

Because the workshop will be bringing together numerous policy-makers, stakeholders, and research groups, NASA PACES will have the opportunity to raise awareness of their mission as well as demonstrate the power of integrating and visualizing data collected from ground studies and satellites. Customized use of GIS provides the platform for integrating, visualizing, and analyzing data in a format that will facilitate discussion within selected groups of the workshop.

3. **Development of a Decision-Support System**

A GIS is a computer-based system that is capable of assembling, storing, integrating, manipulating, and displaying geographically-referenced information. This type of system provides the capability to measure, map, monitor, and model information that is stored. Measurement is concerned with the storage of data with various environmental parameters. Mapping refers to the spatial display of different types of data. Monitoring allows the user to observe changes in data over a period of time, and modeling allows the user to create models based on relations [6].

The ability of GIS to manipulate data from both temporal and spatial data dimensions provides the foundation to build a decision-support system. Traditional GIS software can be enhanced with statistical analysis, image processing, presentation graphics, and other applications in order to present the data in a format that can support decision making. Data that has been accurately and soundly collected, managed, analyzed and presented though the decision-support system can be used to facilitate decisions that is of interest to government officials, business people, community leaders, planners and teachers, e.g., planning new water transport schemes, business placement, subdivisions, or emergency evacuation routes [3][6]. A decision-support system provides two approaches to the decision maker:

- The technical staff can use the unique combination of tools to present information to the decision maker.
- The decision maker is provided with interactive tools that promotes a user-friendly atmosphere for manipulating, analyzing, and displaying information.

The prototype GIS that is being developed at NASA PACES will provide an opportunity for the attendees to become aware of the data that is available through PACES. The demo also will highlight the capabilities of a decision-support system especially from the view of integrating and visualizing data from multiple sources. The steps in creating a decision-support system includes: define the problem, identify the targeted audience, select the test area, identify the time frames and data of interest, gather the data, and integrate applications with GIS. The subsections that follow discuss each step from the perspective of building the prototype system.

3.1 Defining the Problem

The first step in developing a decision-support system is to define the problem to be addressed. The prototype focuses on three issues that are critical to the El Paso and Ciudad Juarez areas: rapid population growth, finite water supply, and quality of water resources, The rationale is to develop a platform that will assist in identifying and prioritizing problems in the region, defining and analyzing alternatives, and developing and supporting appropriate plans of actions.

A brief **summary** of these issues are discussed next,

- ***Population growth in the El Paso and Juarez areas [5].*** The border of Texas with Mexico showed the highest percentage growth in the state--27 percent during the 1980-1990 period compared to 20 percent for the entire state. El Paso County growth over the past seventeen years has averaged a steady 2.2 percent growth per year; above both the state and national averages. The population and predicted figures for El Paso, Texas and Ciudad Juarez, Mexico can be summarized as follows:

Area	1990	1996	2010	Change
City of El Paso	515,342	583,421	**750,000**	2.3'%
El Paso County	591,610	667,532	**935,000**	2.9%
Ciudad Juarez Mexico	798,499	914,855	1,292,243	3.1%

- ***Sources of water [4].*** El Paso County obtains water from two different sources: groundwater and surface water. Surface water is obtained from the Rio Grande, while groundwater is acquired from two aquifers, the Mesilla Bolson at the northwest area of the county and Hueco Bolson at the eastern side of the city. About 35 percent of the total water comes from the two aquifers, mainly for municipal use. Meanwhile, most of the surface water is used for irrigation. The following table can outline the use and sources of water in El Paso County.

Use:	% Ground Source	% Surface Source
Manufacturing	88%	12%
Municipal	79%	21%
Irrigation	2%	98%
Power	100%	0%
Livestock	95%	5%
Total	35%	65'%

- ***Quality of water [4][5].*** In El Paso County, the Rio Grande is divided into three segments for classification and monitoring purposes. All three segments exhibit pollution problems. The El Paso-Anthony segment is affected by toxic chemicals from wastewater discharges from both maquilas in Ciudad Juarez and wastewater discharges from the city of El Paso. The second segment of the Rio Grande, which runs through both cities, manifest the same type of pollutants; part of the discharges are received from El Paso's wastewater treatment plants and from the maquilas. Among substances of concern are phosphates, orthophosphates, arsenic, lead, copper, mercury, zinc, and fecal

coliform. The last segment is from El Paso County through Hudspeth and Presidio Counties, and this was ranked in 1994 by the *Texas Natural Resources Conservation Commission* as the sixth most contaminated river segment in the state. The quality of groundwater is potentially threatened because of multiple factors including: leaking underground of both septic tanks and petroleum storage tanks, illegal dumping of hazardous substances, overuse of pesticides, contamination from hazardous waste spills at industrial facilities, and overpumping of water in the Mesilla and Hueco Bolsons, which lead to higher mineral and saline conditions.

3.2 Identifying the Data of Interest

Based on the problem definition, the test area and time parameters can be determined. A general list of questions that should be asked to narrow the scope of interest includes:

. What is the objective of the decision-support system?
. Who is the audience?
. What questions do they want answered?
. What are the available or attainable sources and formats of data?

In the prototype, the test area is defined as the El Paso/Ciudad Juarez areas, and the time period of interest is 1975-2000. From the issues, it was determined that the following data will be useful: ground water, aquifers (shallow and deep), salinity contours, well, and gravity data.

3.3 Gathering the Data

The next step involves gathering data that will aid in characterizing the problem. This requires identifying the types and sources of relevant information. In this project, the base map was defined by the USGS digital line graphs (DLGs), digital elevation models (DEMs), and corresponding satellite imagery. This is being supplemented principally by other appropriate information gathered from Mexico's National Institute of Statistics, Geography, and Informatics (INEGI) data files, gravity models, governmental agencies (such as Texas Water Development Board), and PACES .

3.4 Developing the Decision-Support System

No single source of information exists that combines these data in a format to assist researchers and decision makers in identifying problems and their sources. GIS allows data be input into a common structure and displayed. Typical GIS software requires extensive training and experience to operate. In addition, GIS applications are used to answer specific questions or meet specific application requirements. Consequently, expanding GIS to include decision support requires careful customization to make it both easy to use and meaningful to decision makers.

The decision-support system is being designed to provide an interactive tool set that can be used by decision makers must consider the following:

1. The system must display complex data sets in common formats that assist the decision maker in identifying critical issues such as rapid population growth, combined with limited water supplies and exacerbated by decreasing water quality.

2. Enough information about the issues must be available to define clearly both spatial and temporal characteristics.

3. Integrated views of data are necessary to understand relationships between key factors such as residential water demand and its long term impacts on limited capacity aquifers or concentrations of heavy metals in river water, near or downstream of industrial concentrations.

4. The decision-support system should allow decision makers to relate problems with potential sources or causes, spawning the development of a variety of alternative solutions.

5. The interface and information should allow intuitive evaluation of various alternatives such that they can be prioritized based upon their potential to solve problems or provide benefit.

In the prototype being developed, we will be identifying the characteristics of population growth, use of water resources, the resulting decrease in water supply, and the deterioration of water quality that is currently a problem in the Rio Grande region. It is believed that this information will help the decision makers become increasingly aware of the problems, for example, related to depletion and the increasing salinity in static water supplies. The next step will be for the decision makers to define alternatives to this problem, e.g., putting a dam at a location that is depleting the resource, or recharge the aquifer from a distant reservoir to the affected area. Through the decision-support system, it will be possible to make a first-pass estimate of which alternative would be the most desirable solution. For example, a location for a new reservoir may not be a reasonable option due to surrounding land use or other limits in the terrain. Decision makers can now develop a strategic plan of action for performing further studies and/or developing additional research requirements.

4. **Summary**

A regional workshop, *Tilting the Balance: Climate Variability and Water Resource Management in the Southwest* will be hosted by PACES to discuss the effects of short- and long-term global climate changes. Numerous policy-makers and research groups will be brought together which will provide NASA PACES an opportunity to raise awareness of their mission and to demonstrate a prototype decision-support system that enhances a traditional GIS with applications that include statistical analysis, image processing, and presentation graphics.

The prototype system being developed for the workshop will focus on three issues that are critical to the El Paso and Ciudad Juarez areas: the population growth in the region, “finite water supply, and the quality of the water supply. The rationale for creating the decision-support system is to provide a platform that assists stakeholders in identifying and prioritizing problems in the region, defining and analyzing alternatives, and developing and supporting plans that may be put into affect. By providing enough base information to support discussion on the issues, solutions, alternative solutions, and/or a clearer understanding to the problems may result. The goal of this project is to use GIS as the foundation to build the decision-support system and to demonstrate it application in visualizing, analyzing, and modeling information about the region that is of concern to this workshop.

Acknowledgments. This research is being funded through NASA Contract NCCW-0089.

References

[1] http://edcwww.cr.usgs.gov/nsdi/gendlg.htm

[2] Acosta, J. Alvarado, L., “A Study of Data Issues in the GIS Environment”, Proceedings of the National Conference on Undergraduate Research-97 (NCUR), Austin, TX, april 1997.

[3] Mitchell, A., "Zeroing In", Environmental Systems Research Institute, Inc., 1997.

[4] Mryamoto, S., Fenn, L. B., and D. Swietlik, "Flow, Salts, and Trace Elements in the Rio Grande: A Review," Texas Water Resources Institute and Texas Agricultural Experiment Station, The Texas A & M University System, College Station, Texas, 1995.

[5] Reed, C., "The El Paso County Environmental Assessment," Texas Center for Policy Studies, Austin, Texas, 1997.

[6] Worboys, M., "GIS: A Computing Perspective", Bristol: Taylor and Francis, 1995.

[7] University of Texas at El Paso, "Proposal for the Pan American Center for Earth and Environmental Studies," submitted to NASA Headquarters, March 1995.

An Image Retrieval and Processing Expert System for the World Wide Web

Ricardo Rodriguez, Angélica Rondón, María I. Bruno, and Ramón Vásquez
Laboratory of Applied Remote Sensing and Image Processing
University of Puerto Rico - Mayagüez Campus
Dept. of Electrical and Computer Engineering
P.O. Box 9042
Mayagüez, P. R. 00681-9042

ABSTRACT

This paper presents a system that is being developed in the Laboratory of Applied Remote Sensing and Image Processing at the University of P.R. at Mayagüez. It describes the components that constitute its architecture. The main elements are: a Data Warehouse, an Image Processing Engine and an Expert System. Together, they provide a complete solution to researchers from different fields that make use of images in their investigations. Also, since it is available to the World Wide Web, it provides remote access and processing of images.

1.0 INTRODUCTION

Images are commonly used in numerous areas such as agriculture, geology, forestry, hydrology, marine science, and GIS. Researchers from those areas use different criteria to retrieve the images they are seeking. It can be difficult and time consuming for the users to browse and search within large amounts of data without having the ability to filter it based on the criteria that is meaningful to them.

Data is generally hidden in independent GIS and image processing environments. It is common that not all available data is visible to the users unless a search through all available systems is performed. This may even require previous experience on each system in order to get the desired data unless direct assistance is given. Moreover, the information may only be available to local researchers, impeding the exposure of these resources to external users.

Frequently, the researcher needs to apply processing techniques to the retrieved data in order to determine if it satisfies the desired requirements. This requires extra work and time. The processing requires candidate images to be already downloaded. Normally, it involves the use of an image processing application. The result of this processing determines the validity of the image for the research. This entire process should be performed for each image.

Having a centralized repository that stores images and spatial data makes easier the analysis, retrieval and sharing of information. An image processing engine that transparently utilizes different image processing systems facilitates the analysis and processing of the images. An expert system helps the researcher, no matter his or her expertise, to use the system. And, by making the system available to the World Wide Web, it is possible to provide remote access and processing of images.

2.0 AN INTEGRATED SOLUTION

An integrated solution for on-line image retrieval and processing must address the presented issues. Figure 1 shows the architecture of the system. The Data Warehouse, Image Processing Engine, Expert System and Web Server are the major elements of the system.

The end-user application runs as an applet in a remote machine using a Web browser that supports the JDK 1.1. The inference engine is included as part of the applet code. It interacts with a Data Gateway client process to get the required information from the knowledge base. Also, the applet interacts with this process to get the attribute and spatial data from the Data Warehouse. The major elements of the system are discussed in the following sections.

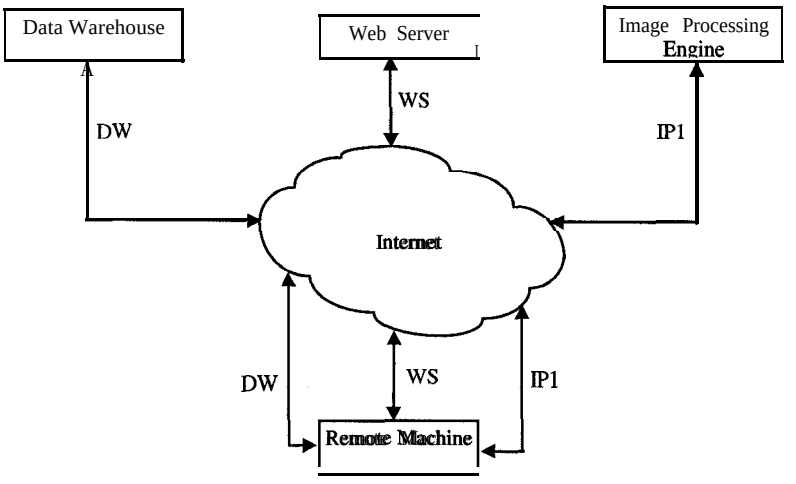


Figure 1. Architecture of the Image Retrieval and Processing Expert System

2.1 DATA WAREHOUSE

The Data Warehouse is designed in a way that addresses the needs of the various users for context-specific analysis of images and spatial data. Available data is extracted from different sources like existing GIS and image processing tools, digitized maps, aerial, satellite and radar image data. The collected data is organized based on the specifications and then placed in the data warehouse.

As opposed to the traditional relational model used on the design of most image and spatial databases, a multidimensional model is being used to match the specific requirements of spatial and image data analysis. Location and Temporal dimensions support the analysis of spatial data as well as other attribute data. Spatial data mining case studies, such as in [Koperski96] are being used to help identify how users want to analyze the data to map the appropriate dimensions. The flexibility provided by the multidimensional model, makes easier the On-Line Analytical Processing (OLAP).

The resulting data warehouse provides a strong platform for mining scientific data. It forms the basis for the knowledge representation of the expert system that defines the set of rules for data search and retrieval. Queries can be constructed against spatial and attribute data using values, ranges of values, location and time.

Figure 2 presents the actual mechanism used to access the Data Warehouse. The remote machine interacts with the Data Warehouse through a server process. This process is implemented by the Data gateway Server. It provides the mechanisms to access both local and remote databases. In the case of our system it is used together with a Database Engine to transparently access an Oracle RDBMS, where the Data Warehouse and the Knowledge Base are implemented.

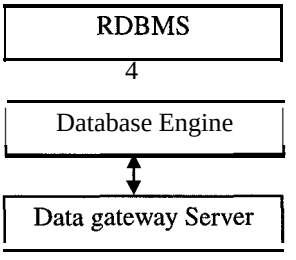


Figure 2. Accessing the Data Warehouse

2.2 IMAGE PROCESSING

The image processing engine is actually a collection of different image processing systems. Figure 3 presents a block diagram describing the elements that constitute the engine and their relationships.

An image processing application programming interface (IPAPI) is being designed. It provides a common interface for the various APIs from different image processing systems. Therefore, different processing algorithms from various systems can be applied to candidate images retrieved from the system. The image processing capability needed for a particular task depends on the kind of researcher as well as the image currently selected. Therefore, an intrinsic relation must exist between the Expert System and the Image Processing Engine. In order to add a new image processing system, it is necessary to create a driver that conforms to the IPAPI specification.

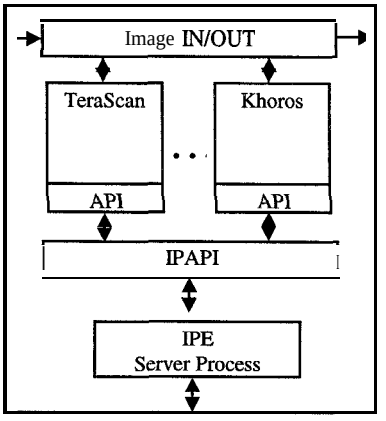


Figure 3. Block Diagram of the Image Processing Engine

The access to the image processing engine is done through a server process. Remote Method Invocation techniques (RMI) and the Common Object Request Broker Architecture (CORBA) are being analyzed and tested for the client-server interaction between the engine and the end-user application. The client process in the remote machine establishes a TCP/IP connection with the server process. In addition to specify the images and the processes to be applied, the communication protocol provides mechanisms to retrieve actual information from the Image Processing Engine. For example, available image processing systems, their procedures and status can be requested.

2.3 EXPERT SYSTEM

By using an expert system it is possible to perform effective search, analysis and processing of images based on different criteria, Therefore, users with little or no knowledge and experience in image processing can determine the appropriate processing techniques to extract specific features from the images that are more meaningful. This is done by following intelligent suggestions.

At this time a framework based on an expert system is being developed to help the end-users in the area of image analysis. This system has the expertise to select specific images in the database and to suggest techniques that can be applied to them.

The original contribution of this expert system is that it offers to the inexperienced user a way to retrieve images from a database and the necessary mechanisms to apply different processing algorithms to those images.

2.4 WEB SERVER

This element of the architecture makes possible the deployment of the system as an on-line service. The end-user application is sent as an applet to the browser in the remote machine. After receiving the applet, two additional connections are established, one with the Data Gateway Server, in order access the data from the Data Warehouse and the Knowledge Base, and the other with the server process of the Image Processing Engine. If the applet is considered a trusted applet, the connections are not limited to the II? address of the Web Server.

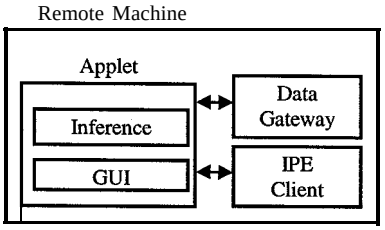


Figure 4. Diagram of the Software Architecture in the Remote Machine

Figure 4 presents the software components needed in the remote machine in order to use the system. The end-user Graphical User Interface and the Inference Engine are coded as part of the applet. The former interacts with the researcher, while the latter receives the inputs from the user and get the actions from the Knowledge Base. The client process of the Data Gateway is necessary to establish the connection with the Data Warehouse and the Knowledge Base. Also, the Image Processing Engine Client Process (IPE) is required to access the Image Processing Engine.

3.0 EXPERIMENTAL RESULTS

A horizontal prototype is being constructed and used to gather additional requirements and specifications.

At this moment, the fully functional part is the one that deals with the access to the Data Warehouse. It includes sample Graphical User Interfaces coded in the applet, the use of the Data Gateway Server and Client processes as well as the Database Engine and the RDBMS. Borland's JBuilder, DataGateway and Database Engine products are being used together with Oracle RDBMS 7.3 with the Spatial Data and Parallel Query Options. The Java Web Server from Sun is being used to deploy the system as an on-line application.

Figure 5 shows a sample applet for viewing images. This viewer searches the Data Warehouse to get information of available images. The user is able to select and view an image. This simple applet exposed the most obvious drawback of this and any other system that uses the World Wide Web, the long time required to send images through the Internet. To solve this problem, a wavelet decompression algorithm is being coded as part of the applet. Wavelet-compressed images are smaller than their JPEG and GIF counterparts, hence their transmission time is reduced.

Translation problems for data intimation, monitoring changes in data source, view maintenance and multiple view optimization are some of the problems associated to data warehousing.

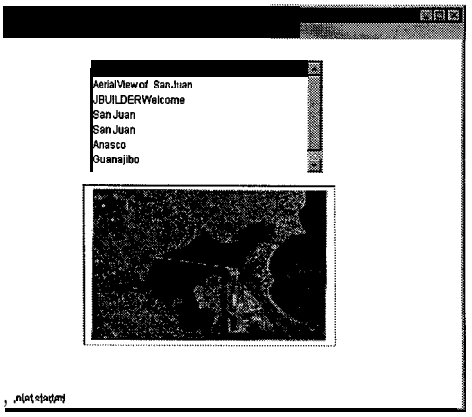


Figure 5. **Image Viewer Applet**

The Data Warehouse access using the Data Gateway server and client components proved to be a great way to get remote database access through the Internet. Also, by using data-aware components and the visual design interface of JBuilder, rapid prototyping and coding is possible.

4.0 CONCLUSION

The integration of data warehouse and an image processing engine into one system asseverates the retrieval and processing of images. Adding an expert system makes it easier to use and its employment as an online application aloud researchers from around the world to retrieve and process images remodeling.

5.0 ACKNOWLEDGEMENTS

This work was supported by NCCS-252 and the Department of Electrical and Computer Engineering, University of Puerto Rico al Mayagüez Campus.

6.0 REFERENCES

[Corey97] Corey Michael J., Abbey Michael. Oracle Data Warehousing. McGraw-Hill. 1997.

[Fayyad96] Fayyad, Usama, David Haussler, and Paul Stolorz. Mining Scientific Data. Communications of the ACM, November 1996.

[Graves96] Graves, Sara J., Thomas H. Hinke and Shalini Kansal. Metadata: The Golden Nuggets of Mining Data. First IEEE Metadata Conference, Silver Springs, Maryland, April 16-18, 1996.

[Gonzalez96] Gonzalez Rafael, Richard. Woods, *Digital Image Processing,* Addison-Wesley, New York. Hayes-Roth Federick, Donald A. Waterman, Douglas B. Lenat, *Building Expert Systems,* Adisson-Wesley, New York, 1996.

[Hamilton97] Hamilton Graham, Cattell Rick, Fisher Maydene. JDBC Database Access with Java. Addison Wesley .1997.

[Kapteyn96] Kapteyn Astronomical Institute, "System for Astronomical Image Processing", Kapteyn Astronomical .Institute, 1996.

[Koperski96] Koperski, K., J.Adhikary, and J.Han. Knowledge Discovery in Spatial Databases Progress and Challenges. In Proceedings of Workshop on Research Issues on Data Mining and Knowledge Discovery, Montreal, QB, June 1996.

[Lu93] Lu, W., J.Han and B. C. Ooi. Discovery of General Knowledge in Large Spatial Databases. Proc. of 1993 Far East Workshop on Geographic Information Systems (FEGIS'93), Singapore, June 1993, pp. 275-289.

[Springmeyer96] Springmeyer, Rebecca, Nancy Werner, and Jeffery Long. Mining Scientific Data Archives through Metadata Generation. 1st IEEE Metadata Conference, April 1996.

[Waterman86] Waterman Donald A., *A guide to expert system,* Addison-Wesley, New York.,1986.

[Weber97] Weber Joseph. Special Edition using Java. Que Corporation. Third Edition.1997.

Preserving Technical Knowledge by Using Web-Like Tools

D. Moreman (l), J. Dyer (l), J. Coffey (2)
(1) Southern University
(2) University of West Florica

A national treasure in technology is disappearing like melting snow. The senior engineers from the early days, who designed the systems still launching our missions into space, are leaving NASA. In many technical areas, no comparable people replace them. Areas exist in which documentation of creations and discoveries is sparse or has not been gathered into safe archive -- and certainly not into archive accessible by modern methods such as CD-ROMs and Web-like Intranets. We will now report on tools and methods we have developed to address this problem and this opportunity. We see our tools as part of a revolution transforming the great social net of sharable wisdom.

In 1993, Ken Davidian of NASA LeRC (Lewis Research Center in Cleveland) decided to attack the problem of knowledge-loss, focusing on cryogenic propulsion. Through Don Noga of NASA Lewis, Davidian had funds for encouraging research at minority institutions. I, Douglas Moreman, was the lucky person at Southern University who happened "to answer the phone" when Davidian called. We developed a plan based upon the knowledge of the day and followed it for two years. A revolutionary new plan has emerged from ideas that came to us as we were introduced to the World Wide Web, the great mother of new knowledge revolutions.

Our original idea was to tape record interviews with retired rocket engineers still living near NASA Lewis. Knowledge gleaned from the tapes would then be made accessible via something like an "expert system. " We, at Southern, happened to have personal contacts with two widely respected experts in expert systems who agreed to help us launch the project. Through these contacts we met and funded John Coffey of the University of West Florida (UWF). Coffey had just finished his master's thesis project (NUCS) under Ken Ford. NUCS is a most impressive expert system for interpreting visual output from one kind of cardiac sensor. John Coffey would teach the latest knowledge-technology to John Dyer and Douglas Moreman, team members from Southern University.

Our team was warmly accepted by our targeted "domain expert" Bill Tabata of NASA Lewis Research Center. Bill Tabata, expert in the Centaur cryogenic rocket, is widely respected in NASA and hugely respected by our team. For two years we used the old tools of "knowledge acquisition" and "knowledge representation. " I call them "old" with a wry smile, as these tools were state-of-the-art until two years ago. Unfortunately they were insufficient for the problem. Wisdom was bleeding away from NASA faster than any practical scale-up of our efforts could address. We all four admitted this to each other the day Tabata introduced us to Ed Prokasky. Knowledge spouted from Prokasky like water from a fire hydrant. After two years, we had not done justice to Tabata. Were we going to ignore Prokasky and the other senior engineers Tabata could introduce? A senior engineer in Denver added to our urgency. His work was critical to the success of the upper stage satellite-launching vehicle, Centaur. We asked "How many people in the world understand this process you have described running from design through manufacturing?" "Two, " he answered. Shocked, we determined to find new methods involving

much faster knowledge acquisition and much less time massaging the knowledge acquired. "Expert systems," were out. We would be obliged to find a new tool. Within a year, we had invented the concept of a Library of Netbooks.

A Netbook can be created in the following way. Experts, singly or in groups, are put into situations conducive to their remembering and divulging deep, often personal, wisdom that is in danger of being lost. In our case, this social part is easy because we are, literally, working with rocket scientists. We use video cameras and may someday use the resulting tapes, but, the main function of cameras has been psychological: to maintain task-focus amongst the team involved in the recording process. The audio recording is the gold produced. From the audio, a transcription is made. The transcription is transformed into a hypertext document with internal links. We also create links to images and to supplementary text. Our own automated tools are used to produce a hyperlinked index. The digitized audio and the text and images are recorded onto a disk (in 1997 this is a CD). The contents of the disk constitute a "netbook."

A netbook can be used via a CD drive in one PC, via a CD jukebox, or can be copied to a hard drive. A netbook can be used in stand-alone mode on one PC, or in an intranet, or can be added to the World Wide Web. The user can just read text, just listen to audio, or listen while watching synchronized text and pictures.

We predict that netbooks will be the basis of a knowledge revolution. The revolution will come, with or without us, because the Web exists and some concept of netbook with be obvious to everyone. The revolution will sweep through all areas of knowledge. For example, consider the tradition called "Oral History." Two years ago I did not even know there was such a tradition. Now, with the help of Roger Launius, Chief Historian of NASA, I see how that tradition will be transformed. Oral Histories will, via netbooks, be so accessible and interesting that someday everyone will have seen one (and heard it too, which is not today in the tradition!). The newly formed Oral History Program of Johnson Space Center will, I expect, be the first to produce netbooks for oral history. The forward thinkers of the X33 Project have already budgeted 100 oral histories. Perhaps they will put them on netbooks? What about medical schools? What about family histories?

Seeing that netbooks will spring into existence everywhere brings us to concerns on a higher level. Let's return to our work at Southern for NASA Lewis Research Center and take a skeptical, higher level view. One netbook makes an impressive demo. But is it worth its cost of production? The potential audience for a technical book on some aspect of rocket design may be fewer than 100 per year. And that audience is scattered and cannot be found. We plan to attach our books to the NASA LeRC rocket engine simulation program called RENS. An engineer using RENS might click on the appropriate icon and run a search of our library of books. Given a Library of too limited scope, the probability of him being disappointed is high. He will learn to NOT use our system in the future, as it wastes his time. We can address this fact of life in a way that impacts the bigger, social picture.

We propose a Library of Netbooks. Such a library must have its global "Library Index" linking to the contents of the component netbooks. When a user requests information on "boost pumps," he is not required to specify one book at" a time. If the Library's size and scope

surpasses some critical mass, the probability of a rewarding search is so high that the user is encouraged to return in the future.

The contents of a Library Index can be updated automatically by software that operates on the indexes of the component netbooks. But, this requires standardization of those component indexes. And this requires a standards committee.

Links from one netbook to another may seem easy - until one realizes that today's version of a targeted netbook may be replaced with tomorrow's. So, another task for the Standards Committee for Netbooks will be to specify a linkage system that can be maintained. The Committee might also publish a guidebook for producing netbooks. We expect to print the first guidebook at Southern University.

You see, that to make fruitful our efforts at Southern, we are forced to reach out around the country, finding and bringing together interested parties. Thus, we have met heads of projects mentioned above. We have met NASA-affiliated historians and the executive producer of the National Historical Landmarks Program of the American Chemical Society. We encourage everyone interested to contact us. Though over-loaded and under-funded ourselves, we can still make introductions among people who can make things happen.

Let's return our focus to a single netbook, seeking suggestions for would-be producers. Our guiding principle has become: minimize labor. Notice that using a web browser reduces the time needed to produce a user-interface. Unfortunately, there seems to be no way to automatically create satisfactory hyperlinks. Furthermore, these links require intelligent work from someone who can understand the material. The Indexes, so crucial to ultimate success, can be created, in first draft, by software. Our colleague John Dyer has demonstrated this. But, he points out a need for human labor in finishing the job.

When you produce a netbook, you might omit the audio. But we encourage you to follow us to the opposite extreme. Driven by hard experience, we plan to purchase radio lapel-microphones and a multitrack digital recorder - spending over $2000 for a system that can be carried in a briefcase. The most fundamental mistake you can make is to underfund the production of audio recording. Think of the audio recording as pure gold. It is the foundation of the book.

We do not see much virtue in long videos of talking engineers. Short clips maybe more useful than distracting, particularly if dense with technical information. Such video can, today, be seen in a netbook by means of the same software that already runs our audio (Real Audio 5.0). There are two problems with video: 1) it eats up memory and 2) it can easily be more distracting than focusing.

Hire a good transcription service then check their work. A good audio recording will save money here. Have your experts check the transcription and plan to hyperlink any additions they think of in the process. Encourage them to think of pictures as well as words. If they have physical objects, make superfluously many photographs of them while you have the chance. Photograph your experts. Collect any diagrams and old photographs the experts offer. Scanners are cheap and their output, via the right software, easy to massage for a webpage. The normal bandwidth-restrictions on images are greatly relaxed for stand-alone CDs and for intranets.

Imagine your product being important 20 years from now. Keep it simple and easily transformable to suit the computer-based tools of that future time.

While developing strikingly new tools for knowledge preservation, we continue to advance the tools we started with. John Coffey proposes that concept maps be used as a navigation tool for a Library, standing on top of the Library Index. The user will see a graphical display of a domain of knowledge, with nodal concepts being joined by links having epistemological meaning. His tools may allow for future application of methods of expert systems for intelligent searching of a knowledge base founded on a Netbook Library.

An Object Model for a Rocket Engine Numerical Simulator

D. Mitra, P.N. Bhalla, V. Pratap and P. Reddy
Jackson State University
Department of Computer Science
P.O. Box 18839, Jackson MS 39217

Abstract

Rocket Engine Numerical Simulator (RENS) is a packet of software which numerically simulates the behavior of a rocket engine. Different parameters of the components of an engine is the input to these programs. Depending on these given parameters the programs output the behaviors of those components. These behavioral values are then used to guide the design of or to diagnose a model of a rocket engine "built" by a composition of these programs simulating different components of the engine system.

In order to use this software package effectively one needs to have a flexible model of a rocket engine. These programs simulating different components then should be plugged into this modular representation.

Our project is to develop an object based model of such an engine system. We are following an iterative and incremental approach in developing the model, as is the standard practice in the area of object oriented design and analysis of softwares. This process involves three stages: object modeling to represent the components and sub-components of a rocket engine, dynamic modeling to capture the temporal and behavioral aspects of the system, and functional modeling to represent the transformational aspects.

This article reports on the first phase of our activity under a grant (RENS) from the NASA Lewis Research center. We have utilized Rambaugh's object modeling technique and the tool UML for this purpose. The classes of a rocket engine propulsion system are developed and some of them are presented in this report. The next step, developing a dynamic model for RENS, is also touched upon here.

In this paper we will also discuss the advantages of using object-based modeling for developing this type of an integrated simulator over other tools like an expert systems shell or a procedural language, e.g., FORTRAN. Attempts have been made in the past to use such techniques.

Introduction

The process of engineering design often involves iterative cycles of design and verification of design parameters for their correctness and effectiveness. The latter part of this cycle requires numerical simulation of the idealized behavior of the designed system. Such idealized behaviors are codified with mathematical formulas and equations. A complex system comprises of many sub-components and processes each being simulated by a formula. The simulation of such systems often involves studying a part of the system by using appropriate set (or a chain) of formulas for the relevant part. For this reason a simulation software for a complex system is a package of sub-routines and functions for representing the mathematical formulas. Typically, during a simulation job, an engineer-simulator would assemble an appropriate subset of those sub-routines in a proper order, possibly for simulating a sub-component of the whole system, and compile them to create the required run-time environment.

Rocket engine is a complex system whose sub-components are being simulated with a set of well defined formulas. Over the last few years traditional rocket engines with liquid propulsion has not seen many major changes in their design. This made it possible to carefully study and understand the behavior of its different components and processes, and thus, to create standard subroutine packages for the purpose of simulation. For example, "Rocets" is one such simulation package developed by Pratt and Whitney. Although such packages are developed with a modular approach with the objective of better software maintenance, little thought have been given to organize the software from a point of view of the ease of their use or the compatibility of the software organization with the real architecture of a rocket engine. An engineer does not have any help whatsoever from such a software in his or her choice of appropriate sub-routines and organizing them in consonance with his or her simulation goal. In other words, the software has no in-built knowledge about the organization of a rocket engine. As far as a user is concerned it is a flat collection of sub-routines, about which the user is supposed to have a thorough knowledge.

The purpose of our current project is to develop an architecture corresponding to a liquid propulsion-rocket engine, for the purpose of organizing a simulation software of the engine. The architecture is being developed by using object oriented design and analysis technique. Such an architecture, ideally should act as an interface (not to be confused with a graphical user interface) between the software and a rocket engineer. By looking at this architecture the engineer would create a sub-module of a rocket engine, which is to be simulated, and the software will combine necessary subroutines in order to achieve the required simulation. A more informed engineer should be allowed to choose a particular sub-routine from a set of such sub-routines doing the same calculation. If a particular component of the system is to be replaced with another one the software would know which subroutines are to be replaced in order to create a new simulation. Thus, the architecture acts as a knowledge-base about the system - to be taken advantage of by a human simulator (Zhang and Zeigler, 1989). We call this architectural component of the rocket engine simulation software an Expert System Architecture of Rocket Engine

Numerical Simulator (ESA-RENS). In this article we have reported on some of our activities in designing the ESA-RENS.

Object Oriented Design: Background

Object oriented design of a complex system involves three phases of modeling: object modeling, dynamic modeling and fictional modeling. Modeling is utilized as a tool for the design (Rambaugh et al, 1991) activity. The target system is analyzed critically in order to recognize its different components and their interaction processes. This analytical approach leads to the design of a software architecture simulating the physical system.

The first phase, the object modeling, involves recognizing the physical or conceptual objects in the system, and organizing them into a hierarchy of classes. Looking at the noun phrases of the description of the system often helps one in identifying the classes. Higher levels in a hierarchical scheme could be the abstract classes while the leaves are the real ones corresponding to the objects of the physical world. Each class packages the data structures and functions required to manipulate those data structures. A lower level class inherits the designated data structure and functions of its ancestor(s). Software maintenance becomes much easier by organizing data and their manipulating-functions into conceptually cohesive classes. However, an explicit objective of our project is to utilize such an organization scheme to help a user in manipulating a dynamically reconfigurable software, like a simulation package.

In the dynamic modeling phase of the design, the temporal relations between the processes in the physical system are studied. Thus, classes of sequential events are organized into time lines, and the temporal orderings between events across different time lines are identified. Multiple possible temporal scenarios of the events are studied in this phase before such a dynamic model of the system is developed. On the other hand, the fictional modeling phase identifies the data flow between different functions or processes in the system. This is opposed to the study of control flow in the dynamic modeling stage. The dynamic model and the functional model of a system capture the behavior of the object-classes identified in the object modeling phase. This is an integration issue over the three models.

The Unified Modeling Language (UML) is a tool for the purpose of object oriented design of systems. It employs specific syntax for the three modeling techniques. The “Rational Rose” from Rational Software Incorporation is a software which implements UML (Rational, 1996). Skeleton of the program (comments and function headers) can be automatically generated by the Rational Rose from a design module. We are using this software for our project.

Object Model of the Rocket Engine

The topmost level of the rocket system which is being modeled in our work is the Propulsion System. It has two components: the "propellant reservoir" and the "engine system. " The Engine System, in turn is comprised of the "feed system" and the "thrust chamber. " A feed system has "turbines," "pumps," "regenerative cooling channels," and "valves." A thrust chamber has also four components: "combustion chamber, " "throat section," "injector," and "nozzles."

The components could be of different types and could be designed with different parameters. For example, a pump could be centrifugal pump, multi-stage pump or single-stage pump. Similarly, a throat section of a thrust chamber has ***upstream radius of curvature, upstream tangent angle, downstream radius of curvature, downstream tangent angle, throat radius,*** and ***type of material*** as its design parameters. Our objective in this phase was to develop an object model for the propulsion system.

The object model of the propulsion system is designed by searching for the noun phrases in the description of the system (Huzel and Huang, 1992). About one hundred and twenty noun phrases are initially identified as feasible classes for the model. Some of the spurious classes are eliminated as being redundant, unrelated to the problem domain, attributes of classes in reality rather than classes themselves, or operations in reality rather than classes etc. The following twenty seven are the final list of classes: feed system, turbo pumps, pumps, regen_cooling channels, thrust chamber, propellant reservoir, combustion chamber, flow control orifices, adaptive cooling, throat section, nozzle, injector, valves, lines, cooling system, spherical_cc, cylindrical_cc, oblong_cc, ducts, propellant supply ducts, impulse turbine, reaction turbine, turbines, co-axial injector, propellants, impinging injector, and engine system.

The next phase was to define the class hierarchy and data dictionary for each class. Instance variables for the classes are being identified at this stage. After this stage associations between classes are being generated. There are three types of associations: inheritance (for the class hierarchy), aggregation (part-whole relationships), and use relationship (who uses what). The association structure also plays a significant role in developing a simulation-run time system of some sub-components of the system, This is because the assembly of the sub-components or the processes is affected by the relationships between the parts as coded in the association structure.

Future Directions

A natural next phase of our project is the development of dynamic and fictional models of the rocket engine. But at this stage a distinction has to be made in the design of a real system (e.g., for developing a rocket propulsion system) and the design of its simulation software, One has to keep in mind that such simulation packages already exist. The engineers would not like to rebuild or even modify their time-tested subroutines.

From that angle, our project involves re-engineering of an existing simulation software, or some of such softwares. A timing analysis should reflect appropriate sequences of sub-programs. However, the process is more than mere re-engineering of a software, in the sense that our umbrella program (the expert system architecture) should be flexible enough to accommodate any (or at least some) such software(s). Hence, most of the functions within the object-classes are not to be rewritten, rather those classes should be built in such a way that the relevant functions within them call the existing sub-routines from older packages. Even if the engineers decide to rewrite their sub-routines they should have full control in integrating their new functions (along with the old ones) under the umbrella of the expert system architecture. Thus, the expert system architecture would also allow a smooth evolution of the simulation software, rather than having to continue to use the old software stretched beyond its normal life-cycle, or having to overhaul it - a process distrusted by the users. However, this type of software architecture involves a cross-language program development issue, since the architecture and the actual simulation routines are likely to be written in different languages (e.g. C++ versus FORTRAN). The research in developing the architecture also addresses the issue of the standardization of function-interfaces, which will also benefit the target user community, in our case, the aerospace engineers.

With the above constraint in mind we are currently studying one of the rocket engine simulation softwares ("ROCETS" from Pratt and Whitney), which was written in FORTRAN. As an example, we will use the subroutines in this package as functions in our architecture. Eventually, our program should be independent enough to call any appropriate subroutine designated by the user, or automatically chosen by the expert system architecture for its effectiveness. This phase of the activity needs to be coordinated with our development of dynamic/functional models of the rocket engine.

Discussion

Developing an expert system architecture for a dynamically reconfigurable software, such as a simulation package, is an innovative idea. Our research is an experiment with a practical implementation of this idea. Our objective is to create a user-understandable architecture for the simulation software, and we are developing this architecture by studying the target physical system. This research employs object-oriented methodology for the purpose of knowledge acquisition and representation. This is a relatively new direction within the artificial intelligence (AI) area. Traditional knowledge representation languages in AI are logic-based, frame-based or rule-based (Gonzalez et al, 1993). Object-oriented design not only makes the resulting software better maintainable, but also more user friendly from a cognitive point of view. Adoption of the object-oriented paradigm within other disciplines like the software engineering and the databases make its use even more imperative within AI, and for projects like ours, for the sake of compatibility with the next generation softwares. Another additional benefit of such a project, as ours, is to create an automated knowledge-base of the target engineering

design. For many reasons a computerized archival of engineering knowledge is much superior to that in the form books or other media.

Acknowledgment

This research is done under a NASA grant (NCC3-437) from the Lewis Research Center. We are indebted to the RENS project's past technical coordinator Ken Davidian and the current technical coordinator Joseph Hemminger for their support in this work.

References

A. J. Gonzalez and D. D. Dankel, "The Engineering of Knowledge-based Systems: Theory and Practice, " Printice-Hall, Inc., New Jersey, 1993.

D. K. Huzel and D. H. Huang, "Modern Engineering for Design of Liquid-Propellant Rocket Engines," American Institute of Aeronautics and Astronautics, Washington D. C., 1992.

J. Rambaugh, M. Blaha, W. Premerlani, F. Eddy and W. Lorensen, "Object-oriented Modeling and Design," Printice-Hall, Inc., New Jersey, 1991.

Rational, User's manual: "Using Rational Rose 4.0," Rational Software Incorporation, Santa Clara, CA, 1996.

G. Zhang and B. P. Zeigler, "The System Entity Structure: Knowledge Representation for Simulation Modeling and Design," in L. E. Whidman, K. A. Loparo and N. R. Neilson (Eds.), "Artificial Intelligence, Simulation, and Modeling," John Wiley & Sons, New York, 1989.

Trace Pattern Matching Using Knuth-Morris-Pratt String Pattern Matching Algorithm

Yabo Wang
NASA Center for Autonomous Control Engineering
and
Department of Computer Science
North Carolina A & T State University
Greensboro, NC 27411

1 Introduction

The Trace Assertion Method is a formal method for' software module interface specification [7]. In this method, a software module is viewed as a finite state machine and the machine changes its state upon receiving an invocation of an access program of the module. A sequence of program invocations are called ***traces*** and a module's externally observable behavior is specified by a set of assertions, called trace assertions, about the traces of the module. A trace assertion is a parameterized state transition rule or an output role. For example, in the trace specification for a stack module, the effect of a POP program invocation on a stack of three elements may be specified by the trace assertion

$$[\text{PUSH}(x1).\text{PUSH}(x2).\text{PUSH}(x3)].\text{POP} =_e \text{PUSH}(x1).\text{PUSH}(x2)$$

Intuitively, this assertion states that if the stack is in the state denoted by a trace that can be matched to the trace pattern PUSH(x1).PUSH(x2).PUSH(x3), then after accepting the input symbol "POP", the stack changes to the new state denoted by the trace PUSH(x1).PUSH(x2) where xl and x2 bear the values bound by the pattern matching at the LHS of the assertion. This interpretation obviously resembles that of a rewriting rule.

A more general trace assertion about the effect of a POP program invocation on a non-empty stack may state

$$[T.\text{PUSH}(x)].\text{POP} =_e T$$

For any non-trivial software module, it is impossible to make assertions about the effect of every program invocation on every possible trace of the module by enumeration. Therefore, in the Trace Assertion Method, a complete specification is organized around a set of canonical traces and a set of reduction rules [7]. In the above two examples, the trace patterns in the square brackets are patterns denoting canonical traces. The most general form of a trace assertion takes the form

$$P(\chi(T.E)) :: T.E =_e T'$$

where T is a canonical trace pattern, E is an event pattern (a program identifier with its parameterized arguments), $\chi(T.E)$ is the set of variables contained in T.E, $P(\chi(T.E))$ is the condition of the assertion, and T' is a canonical trace pattern such that $\chi(T') \subseteq \chi(T.E)$.

To simulate the interface of a module **M, we** treat each trace assertion in **M's** trace specification as a (conditional) rewriting rule. We let the state of **M** be denoted by a set of canonical traces. Then, whenever a program of **M** is invoked, we try to find a trace assertion ***such that*** (1) the ***canonical trace denoting the current*** state plus the new program call match the LHS of the assertion; and (2) the match satisfies the condition of the assertion. The RHS of the assertion then becomes the canonical trace denoting the new module state.

In this rewriting process, one of the most important step is the matching of canonical traces (denoting the module states) with the LHS of trace assertions. Due to the fundamental difference between the syntactical structures of traces and terms, none of the term pattern matching algorithms can be applied to trace pattern matching. In this paper we describe how to apply the symbol table technique and the Knuth-Morris-Pratt string pattern matching algorithm (hereafter, KMP algorithm) [4] to accelerate the trace pattern matching process.

2 Preliminaries

The following terms are defined with regarding to a specific module **M.**

Definition 2.1 (event *expressions)* The ***event*** of ***an invocation*** of ***a program*** E of **M with** the ***arguments*** $a_1, \ldots a_k$ *is* denoted by the event expression

$$E(a_1, \ldots a_k)$$

where, E is a program name and a_i's are the actual values of program parameters. When k = O, we write E instead of E(). The set of all event expressions of **M is** denoted by **EM.**

Definition 2.2 (event patterns) The set of event patterns of **M, EP_M,** is defined as follows. If $E(a_1, \ldots a_k) \in \mathbf{EM}$, then $E(x_1, \ldots x_k) \in \mathbf{EP_M}$, where for each x_i $(1 \le i \le k)$, either $x_i = a_i$, or x_i is a variable, called ***argument variable.***

Intuitively, an event pattern is an event expression with zero or more of its actual parameters being replaced by argument variables. For example, PUSH(3) is an event expression, while PUSH(x) is an event pattern.

Definition 2.3 ***(traces)*** Let the set of traces of M be denoted by **TM.** Then,

(1) the empty trace, _ • TM,

(2) for every $E(a_1, \ldots a_k) \in$ EM, $E(a_1, \ldots a_k) \in$ TM, and

(3) if $T1, T2 \in \mathbf{T_M}$, then $T1.T2 \in \mathbf{T_M}$.

Definition 2.4 ***(subraces):*** Subtraces of a given trace $T = E_1.E_2 \ldots .E_n$ are the sequences of event expressions in:

$$T_s = \{E_j \ldots ..E_k \mid \text{where } 1 \le j \le k \le n\} \cup \{\_\}$$

The subtrace $E_j \ldots .E_k$ of T can be denoted as $T|_{j-1}^{k}$, and the empty trace _ as $T|_{0}^{0}$.

The subtrace notation $T|_{j-1}^{k}$ an be understood as follows: we view each non-empty trace T as having two invisible ". "s at the beginning and the end of the trace, respectively. We then number all "."s, including the two "invisible" ones, in T from left to right with non-negative integers. $T|_{j-1}^{k}$ is the subtrace of T between the two ". "S numbered j - 1 and k.

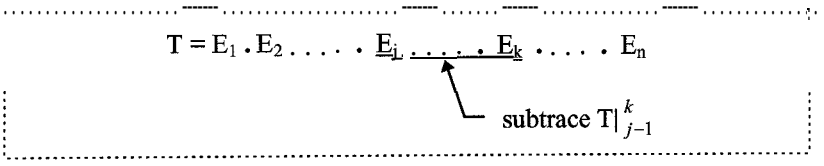


Figure 2.1 Subtrace Notation

Definition 2.5 ***(trace patterns)*** The set of trace patterns of **M, TP_M, is** defined as

(1) for every $E \in \mathbf{EP_M}$, $\mathbf{E} \in \mathbf{TP_M}$,

(2) if El, $E2 \in \mathbf{EP_M}$, then $E1.E2 \in \mathbf{TP_M}$,

(3) if $E \in \mathbf{TP_M}$, T is a trace variable denoting an element of $\mathbf{T_M}$, then, $E.T \in \mathbf{TP_M}$, and $T.E \in \mathbf{TP_M}$,

(4) if $T1, T2 \in \mathbf{TP_M}$, then $T1.T2 \in \mathbf{TP_M}$.

The subset of the trace patterns defined by (1) and (2) of Definition 2.5 is called ***S-patterns. An*** S-pattern contains no trace variables. For example, T. PUSH(x) is a trace pattern, but not an S-pattern (since T is a trace variable).

According to Definition 2.5, a general trace pattern has the format $S_1.T_1.S_2.T_2 \ldots .T_{n-1}.S_n$ where, each S_i is an S-pattern (possibly an empty trace) and each T_i is a trace variable.

3 Trace Pattern Matching

We formulate the trace pattern matching problem as follows: Given a trace T, and a trace pattern T_{Pat} as input, compute a substitution θ such that $\theta(T_{Pat}) = T$. If the computation succeeds, return θ and the flag ***true*** as the output; otherwise, return the flag ***false.*** For example, the trace pattern $T_{Pat} = T1.PUSH(x)$ matches trace PUSH(l) .PUSH(2).PUSH(3) by the substitution $\theta = \{T1 \mid\!\rightarrow PUSH(l) .PUSH(2), x \mid\!\rightarrow 3\}$.

To clearly distinguish a trace T from a trace pattern we hereafter call T the trace text. A substitution θ satisfies $\theta(T_{Pat}) = T$ is called a binding substitution of T_{Pat}.

3.1 Trace Pattern Matching Versus String Pattern Matching

We observe that, if each variable in a trace pattern is treated as a "wild card" that matches some trace elements, then a trace pattern resembles those in [5] called "patterns with wild cards". A wild card, "*", in a string pattern is a "don't care" symbol that matches several symbols in a string text. It is the constant symbols in the string pattern and the given string text that together decide what symbols a wild card symbol should match.

For example, a given pattern "a*ac" matches the text "abbaac". The substring "a" before "*" and the substring "at" after "*" in the pattern determine that "*" = "bba". The same pattern matches the text "aaaac" by matching "*" to "aa". We often say "a []ac" is the context in which this wild card occurs. In other words, the context of a wild card plus the given string text determine the symbols a wild card symbol should match.

Similar to the "*" symbol above, in a trace pattern, both argument variables and trace variables are wild cards in the sense that the former can match any arguments and the latter can match any subtraces. The actual values they are matched to are determined by (a) the context in which they occur, and (b) the given trace texts.

For example, the trace pattern PUSH(x).T1.TOP.T2 matches the trace text PUSH(4).TOP.PUSH(3).POP as shown

```
trace   =   PUSH(4) . TOP . PUSH(3) . POP
pattern =   PUSH(x).T1.TOP .       T2
```

Figure 3.1 A Trace Pattern Matching Example

In Figure 3.1, the positions of PUSH(x) and TOP, two S-patterns, determine that (1) T1 can only match the empty trace; (2) T2 matches the subtrace PUSH(3).POP; and (3) x, an argument variable, matches 4. The same trace pattern also matches another trace text PUSH(7).PUSH(2).TOP by matching T1 to PUSH(2), T2 to _, and x to 7.

In a trace pattern matching process, the S-patterns in the trace pattern is the context of the trace variables and hence should be processed first. For the general trace pattern $S_1.T_1.S_2.T_2 \ldots T_{n-1}.S_n$, each trace variable T_i has its context $S_i.[].S_{i+1}$. The match of a trace variable can only proceed after its context matching, i.e., the S-Pattern matching.

3.2 S-pattern Matching

In this section, we assume that the given trace pattern T_{Pat} is an S-pattern,

$$T_{Pat} = E_1(a_{11}, \ldots . a_{1i}).E_2(a_{21}, \ldots . a_{2j}) \ldots E_m(a_{m1}, \ldots . a_{mk}),$$

where each $E_i(a_{i1}, \ldots . a_{in})$ is an event pattern.

In the study of the general string pattern matching problems, there is no assumption on the profile of either the string text or the string pattern; they are arbitrary strings from an alphabet. In the famous KMP algorithm [4], the string pattern is pre-processed in order to compute the profile, i.e., the next sliding distances of the string pattern. In the trace pattern matching problem, however, the following facts about the profiles of the trace text and trace pattern are known:

(1) A trace text is a string with many repeated substrings. The substrings are program names defined in the syntax table of a trace specification. For example, any trace of the stack module contains many substrings of "PUSH", "POP", or "TOP".

(2) An event pattern is the logical unit to perform the matching operation. A match in string pattern matching operation, for example, when text = *EnterStaffAddr* and pattern = *StaffAddr*, is not a match in the context of trace pattern matching. After the initial matching fails, the sliding of the trace pattern to the fifth character of the trace text is a wasted operation.

Taking the advantages of the two facts, we can convert the costly string comparisons in the S-pattern matching to integer comparisons with the help of a symbol table. Our general scheme of the S-pattern matching is very similar to that of string pattern matching. The given S-pattern slides against the trace text in order to find a match with a subtrace of T. As indicated by (2) above, the minimum meaningful sliding distance should be the next "." position in T. The trace concatenator " ." in T serves as the next sliding position indicator as sown in Figure 3.2.

```
trace text:              PUSH(3).TOP.PUSH( 1).PUSH(4)
initial pattern position PUSH(x).PUSH(4)
first sliding:                 PUSH(x).PUSH(4)
second sliding:                       PUSH(x).PUSH(4)
```

Figure 3.2 Minimal Trace Pattern Sliding Distances

Let ***n*** be the number of event patterns in T_{Pat}. Suppose T_{Pat} is currently at the position ***i*** of the trace text T. Then, whether or not to further slide the S-pattern to the right of the current position is determined by the matching result of the current segment of T, i.e., the subtrace $T|_i^{i+n}$ and T_{Pat}. The segment matching involves (a) the matching of the events names in T_{Pat} and $T|_i^{i+n}$; and (b) the matching of the argument lists if(a) succeeds.

The name matching, as a string pattern matching, can be a costly operation. We therefore choose to operate on the integer representations of strings through a symbol table. As stated in the fact (1) above, as a substring, the event names have many repeated occurrences in a trace text and trace pattern. The symbol table for the events names can be constructed by just scanning the syntax table of the given trace specification. Then, the internal representation of event names in both the trace text and the S-pattern are based on their respective integer ID's assigned by the symbol table. Using this technique, a name matching now takes only one integer comparison.

The cost of symbol table operations will be analyzed in Section 4. We will see that this cost is well justified since the name matching will be repeatedly performed during the trace simulation process.

With the help of the symbol table, a more efficient S-pattern matching algorithm can be developed if we apply the hashing approach used in matching terms. The hashing approach, as described in [3], views all term patterns as forming a hashing table. Each term pattern in the table is a "bucket" with a signature bearing some characteristics of the pattern. This signature serves as a necessary condition that any matching term must satisfy. For a term to be matched, its signature is computed and used as the key to enter a bucket in the table where the further detailed checking is performed. The effectiveness of this approach depends on how simply the signature can be computed and how effective is the signature (so that most of the un-matchable terms can be eliminated).

For the S-pattern matching problem, there is an obvious signature that is both simple and effective: the list of events names in the S-pattern. For example, given the trace text

T = *AddTuple(3,x,9,y,a).AddTuple(4,5,z,6,7).AddSet(3,s).AddTuple(8,3,9,4,2)...*

and the trace pattern

T_{Pat} = *AddTuple(3,x,9,y,a).AddTuple(4,5,z,6,7).AddTuple(3,3,3,3,3)*

suppose that through the symbol table, the event names "*AddTuple*" and "*AddSet*" have the integer ID's 12 and 7, respectively. Then, the trace text has the internal representation

T'= *12(3,x,9,y,a).12(4,5,z,6,7), 7(3,s).* ***12(8,3,9,4,2)...***

and the trace pattern

T_{Pat}' = *12(3,x,9,y,a).12(4,5,z,6,* ***7).12(3,3,3,3,3).***

A necessary condition for T' to be matched to T_{Pat}' is that there are at least three consecutive 12's in T'.

In general, the list of event names (represented as integers) in T_{Pat}' should be used as the signature to check against the trace text T' before getting into the argument list matching. In the above example, the initial signature check fails at the third integer comparison. Whereas using the naive sequential comparison approach, it takes 12 comparisons (2 for event name matching, and 10 for arguments matching) before the detecting of the mismatching.

The biggest advantage of this signature checking approach is that, by storing internal representations of traces (and trace patterns) in appropriate data structure, some efficient string matching algorithms can be integrated into the approach for fast checking of the signatures.

To facilitate the signature matching, the signatures of an S-pattern should be stored in a contiguous data structure with random access, such as an array. We choose our internal S-pattern representation as an array of records. Each record is used to store an event expression. For the purpose of this paper, each record has an ID field representing an event name and a pointer to a list containing the argument list.

Treating each ID as a symbol, the name lists of T_{Pat}' and T', represented as integer lists, to be matched now become strings, and the signature checking becomes the string pattern matching. Thus, some fast string pattern matching algorithms, such as the KMP algorithm [4], can be applied to quickly find the substring in T' that matches the signature. In the above example, when the initial subtrace of T' fails to match T_{Pat}', according to the KMP algorithm, T_{Pat}' should slide to position 3 of T' as shown in Figure 3.3.

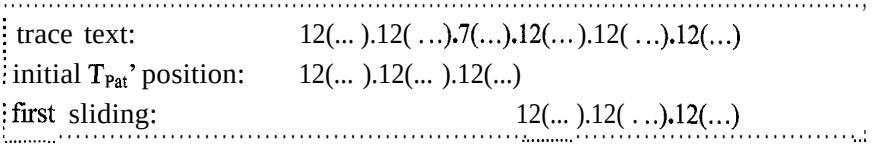


Figure 3.3 Applying KMP Algorithm to S-Matching

After the checking of the necessary condition, i.e., a successful signature checking, the next step of S-pattern matching is to check the sufficient condition for the checked subtrace of T' to match the S-pattern T_{Pat}' -- finding the substitutions of each argument in T_{Pat}' with the respective argument of the subtrace. This can only be done sequentially for the list of event pattern in T_{Pat}', we sequentially match the argument list of each event pattern; for

each argument list, we sequentially match each argument in the list. The argument substitution is performed by a straightforward algorithm **ListMatch** which will not be further discussed here.

The algorithm outlined above is named **S-Match.** Given a T_{Pat}', which is an S_i in a general trace pattern $T_{Pat} = S_1.T_1.S_2.T_2 \ldots T_{n-1}.S_n$, the **S-Match can only** operate at a certain segment of the trace text (for example, if S_{i-1} has been successfully matched to the trace text between (1, r), then, S_i can only start to match trace text starting from position at (r+1)). There are two special cases, i.e., i = 1 and i = n. For S_1, it has to match the prefix of trace text; while for S_n, it has to match the suffix of the trace text. For the other cases, we let the KMP algorithm to match S_i at the remaining middle segment of the trace text. For this reason, the original KMP algorithm is slightly modified so that we can pass the segment boundary positions to it.

3.3 Matching Trace Variables

With the above result, we can now match each variable in a general trace pattern $T_{Pat} = S_1.T_1.S_2.T_2 \ldots T_{n-1}.S_n$. For each S-pattern S_i, we use two positions, l_i and r_i, to record the subtrace of T it matches, i.e., $T|_{li}^{ri}$. We first match the prefix of T with the S-pattern S_1. If the matching succeeds, we record the matching subtrace of T by l_1 = O and r_1 = *length*(S_1); otherwise, the whole matching fails. Suppose the matching for S_1 succeeds, we then try to match S_2 to T starting from r_1. If it fails, then the whole matching fails; otherwise, the wild card T 1 is set to match with the subtrace of $T|_{r1}^{l2}$. as shown in Figure 3.4.

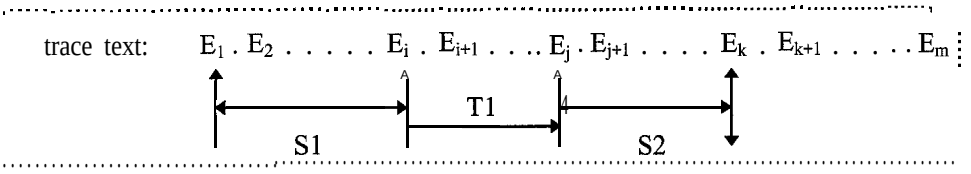


Figure 3.4 Matching Trace Variables

This process continues until either any of the S-pattern matching fails, which fails the whole pattern matching, or, S_n is matched (and thus T_{n-1} is matched), which successfully concludes the whole pattern matching. Note that S_n must match the suffix subtrace of the trace text, i.e., r_n = m.

The algorithm for matching trace variables is **FindSub**, This algorithm, employing the **S-Match** and matching trace variables in a trace pattern, searches for a binding substitution for the given trace pattern.

4 Discussion

For the readers familiar with term rewriting, they might expect that the structure of traces resembles the tree structure of algebraic terms. However, Definition 2.1 restricts traces to be flat strings since no function symbols is allowed in event expressions. We need to study the applicability of the KMP algorithm in trace pattern matching once the specification method and the rewriting system are extended to allow more general traces.

Note that, given a trace pattern and trace text, the **S-Match** algorithm presented in this paper finds the left-most substitution. In the rewriting process, the found substitution will be further processed to check whether the matched variables meet the conditions of a certain rewriting role. When the condition check is successful, rewriting will proceed; otherwise, the **S-Match** algorithm will be called again to match the next possible substitution of the same pattern and text. This backtracking process and the corresponding algorithm are beyond the scope of this paper and hence not presented here.

By using signatures of S-patterns as necessary conditions, the trace pattern matching process quickly eliminates useless and costly pattern matching on the event argument lists. By introducing a symbol table, the more expensive string comparisons in the signature matching is replaced by simple integer comparisons. The chosen data structure and the KMP algorithm then further accelerates the signature matching and hence the whole trace pattern matching process.

The overhead of the whole approach lies in the symbol table operations. The trace pattern matching algorithms discussed in this paper is specifically developed to support the simulation of module interfaces by the Stepwise Rewriting Strategy [6, 7]. Particularly, we desire to achieve a better run-time performance. In this strategy, we assume that a given trace specification can be preprocessed so that the "next" arrays of all S-patterns can be calculated as well as the symbol table (of access program names) can be constructed. When the simulation starts, events, represented by event expressions, occur sequentially. For each event expression, we can use the program name to look-up the symbol table in order to find its integer ID. For a module, let the number of programs provided

on the interface be n. In average, the symbol table look-up operation takes n/2 string comparison inside the symbol table. After the event name obtains its ID, the internal canonical trace representation is entirely built by using the integer ID's and the further trace pattern matching also operate on the ID's. Since, in a trace specification, there are potentially many more trace assertions (and hence trace patterns) to be matched than the number of programs, the time saved far exceeds the cost of a look-up operation; plus, the integer ID's enable the use of the KMP algorithm which results in farther saving of time.

We develop the trace pattern matching algorithms because the structure of trace is rather unique and hence, to our best knowledge, none known algorithm applies. Each event expression in a trace resembles a list (of arguments) of an S-expression. "But a trace is a sequence of such "list" without an operator at the "root" level and hence is not an S-expression. Because of the linear structure of traces, term pattern matching algorithms do not apply, either. It appears that the trace pattern matching can be treated as a string pattern matching and hence the trace rewriting as string rewriting [2]. However, unlike trace rewriting [7], in the string rewriting, (1) patterns are string constants; and (2) it only need to matching the substring in the string text with the given string pattern, not the whole string text.

As a component of a Trace Rewriting System, the detailed data and analysis on the trace pattern algorithms are not available until the whole system prototype is in operation. During the test runs of the algorithms, we noticed that in a trace specification, there are often some very short S-patterns (length of 1 or 2). The benefits of the KMP algorithm on those short trace patterns are not shown.

Acknowledgment

This work is partially supported by the NASA Center for Autonomous Control Engineering, under the grant ACE-48146.

Reference

[1] Aho, A. V. "Algorithms for Finding Patterns in Strings" Handbook of Theoretical Computer Science, van Leeuwen, J. (cd.), pp. 256-300, Elsevier Science Publishers B. V., 1990.

[2] Book, R. V., Otto, F. String-Rewriting Systems, Springer-Verlag, 1993.

[3] Cowan, R. M., and Griss, M. L. "Hashing -- The key to Rapid Pattern Matching" Symbolic and Algebraic Computations, Lecture Notes in Computer Science 72, pp. 266-278, 1980

[4] Knuth, De., Morris, J. H., and Pratt, V. R. "Fast Pattern Matching in Strings" SIAM Journal of Computer, Vol. 6, No. 2, pp. 323-350, 1977.

[5] Pinter, R. Y. "Efficient String Matching With Don't-care Patterns" Combinatorial Algorithms on Words, Apostolic, A., and Galil, Z. (eds.), Springer-Verlag, pp. 11-29, 1985.

[6] Wang, Y., and Parnas, D. L. "Trace Rewriting Systems" Proceedings of the Third International Workshop on Conditional Term Rewriting Systems, pp. 176-180, France, July 1992. Lecture Notes in Computer Science 656, pp. 343-356.

[7] Wang, Y., and Parnas, D. L. "Simulating the Behavior of Software Modules by Trace Rewriting" IEEE Transactions on Software Engineering, Vol. 20, No. 10, pp. 750-759.

HIGH PERFORMANCE DATABASE MANAGEMENT FOR EARTH SCIENCES*

Naphtali Rishe, David Barton, Frank Urban, Maxim Chekmasov, Maria Martinez, Elms Alvarez, Martha Gutierrez, Philippe Pardo

High Performance Database Research Center, School of Computer Science
Florida International University, ECS 243, University Park, Miami, FL 33199 U.S.A.
hpdrc@cs.fiu.edu, http://hpdrc.cs.fiu.edu

ABSTRACT

The High Performance Database Research Center at Florida International University is completing the development of a highly parallel database system based on the semantic/object-oriented approach. This system provides exceptional usability and flexibility. It allows shorter application design and programming cycles and gives the user control via an intuitive information structure. It empowers the end-user to pose complex *ad hoc* decision support queries. Superior efficiency is provided through a high level of optimization, which is transparent to the user. Manifold reduction in storage size is allowed for many applications. This system allows for operability via Internet browsers. The system will be used for the NASA Applications Center program to store remote sensing data, as well as for Earth Science applications.

INTRODUCTION

The NASA Regional Applications Center (RAC) at Florida International University (FIU) is being established as a storage facility and distribution center for all kinds of spatial and conventional data of local and regional interest. Access to this data is provided to a group of simultaneous local users via 150Mbps ATM communication lines and also to remote users via the Internet at much lower speeds. The volume of data is such that it can be represented in only one form in the database and for interactive access, must be compressed. Up to now access to such a large amount of satellite data has been by cumbersome FTP file transfer, sometimes taking a week or more to arrange for file availability. It is widely believed that this difficult access is hindering potentially valuable and widespread applications of remotely sensed satellite data. To make the compressed spatial data easily available to users we have implemented several database user views that enable user programs to interact with the data as if it were stored in many different formats. This technique greatly simplifies access to the spatial data and enables the rapid development of user programs.

The Everglades National Park is currently considered one of the most environmentally sensitive regions in the United States. The problem of monitoring this territory is increasingly important. Environmental surveys have been performed in the Park for more than half a century, and the results of these surveys need to be stored and managed in an efficient manner that is also easy to use. To achieve this goal a joint project between Everglades National Park (ENP) and the High Performance Database Research Center (HPDRC) at FIU has been established. The project is aimed at constructing a set of databases which contain the environmental data that has been collected in the park since 1911 to date. These data sets describe trends in vegetation, land and marine fauna as well as the physical parameters of the area. The design stage of the project is based on the semantic binary model database approach that is being developed at the HPDRC. The implementation stage consists of table creation, data loading, forms, user views and reports, as well as Internet access support using Oracle Relational Database Management System.

Along with the improved use and management of this environmental data, the ongoing research contributes towards better understanding and documentation of historical data. It will aid in future survey designs by allowing the opportunity to check the data sets and correct errors in them. Through the use of modern database technology, survey results can be made available to the public through the Internet. Thus, new environmental surveys will have a stronger research background and the data will be widely available for urban planing and educational purposes. Plans are currently underway for the construction of a semantic database, which will integrate the present textual data with satellite images of the territory. This will give a broader view of the development of South Florida.

* This research was supported in part by NASA (under grants NAGW-4080, NAG5-5095, and NRA-97-MTPE-05), NSF (CDA-971 1582, IRI-9409661, and HRD-9707076), ARO (DAAH04-96-1-0049 and DAAH04-96-1-0278), DoI (CA-5280-4-9044), NATO (HTECH.LG 931449), and State of Florida.

HPDRC is presently developing, under NASA sponsorship, a high performance semantic database system. Among its present applications is storage of spatial data collected from the OrbView-2 (formerly SeaStar) Satellite and other remote sensors. This system is used by the RAC. HPDRC's Semantic Database Management System (DBMS) is based on the Semantic Binary Model. In the Semantic Binary Model, information is represented by logical associations (relations) between pairs of objects and by the classification of objects into categories. The Semantic Binary Model is the most natural and convenient way of specifying the logical structure of information and for defining the concepts of an application's world. It is represented in the form of a semantic binary schema (Rishe 1992).

The Semantic Database models are potentially more efficient than the conventional models for two main reasons. The first is that all the physical aspects of the representation of information by data are invisible to the user and the second is that the system knows more about the meaning of the user's data and about the meaningful connections between such data. The first reason creates a potential for optimization by allowing more changes without affecting the user programs. The second allows this knowledge to be utilized to organize the data so that meaningful operations can be performed faster at the expense of less meaningful operations (Rishe 1992). Decomposing queries into atomic retrieval operations maximizes the efficiency of retrieval requests and each atomic retrieval request normally requires only one disk access.

Efficient updates are also a requirement of the semantic database. A transaction is composed of a set of facts to be deleted from the database, a set of facts to be inserted and additional information needed to verify that there is no interference between transactions of concurrent programs. A program fragment containing numerous update commands contained among other computations can generate a transaction. However, until the last command within a transaction is completed, the updates are not physically performed but instead are accumulated by the DBMS. Once the transaction is completed, the DBMS checks its integrity and performs the update. This insures the consistency of the database, with regard to applications and users. Until the transaction is completed, its effects are invisible (Rishe 1992). HPDRC's Semantic DBMS contains semantic facts and inverted semantic facts. This fact inversion scheme assures efficiency of queries including range queries and content access and also exhibits low entropy of data blocks, which facilitates compression. Even without special compression algorithms, the semantic DBMS is very storage efficient. However, contemporary compression techniques can help achieve significant savings (Rishe 1993).

The mathematical abstraction of the relational model has allowed the introduction of powerful and easy - to-use languages for retrieval and updates of databases. The semantic model however, offers a higher degree of abstraction, which results in more concise user programs, speedier processing (due to optimization), and a wealth of other features. Relational databases are good for general conventional database applications. However, in situations where the structure of information is complex, or where greater flexibility is required (objects with unknown identifiers, or objects moving from one category to another, etc.), or where non-conventional data is involved (spatial data, long text, images, etc.), semantic databases are successful.

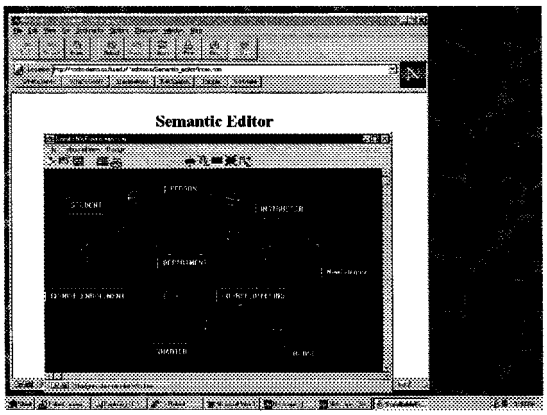
We are currently designing a parallel database system based on the semantic/object-oriented approach that will accommodate many servers, databases, and users. A server (S1) will gain access to other databases hosted by another server (S2) if there is a link between the two servers. S 1 will keep information about the accessed database through its cache and update the database on S2 upon commitment of transactions. The databases will be stored as binary files. A server will access blocks of a file through the file system interface. A file server may use many disks for storage and thus will be able to retrieve different blocks for different databases concurrently.

Our database system provides exceptional usability and flexibility. This strength is inherent to the object-oriented approach of our binary semantic model. A database developer will find designing in the semantic model intuitive since it mostly reflects the real world. The model captures the data along with its meanings (through relations). The model allows the production of the same database design for clients and developers. A client will have a greater understanding of a schema and will be in the position to offer more valuable input. This understanding comes at the price of more iteration in development. Once clients have a better understanding of their design, they usually find it easier to make more demands. A client reviewing the design can easily spot flaws in the understanding of the problem requirements.

A database for Earth Sciences data must address the issue of storage size reduction since the amount of data tends to be large. The relational model carries a lot of overhead in the space allocations for its tables and indices (keys). The semantic database reduces space allocation by eliminating this overhead. Although we also store inverse facts, the absence of additional tables and index files allows us to offer a better space management than relational model does. Usually database systems use various data representation structures where the semantic database system only uses a B-tree as its data structure that facilitates data packing. Since our data are in lexical

order, prefix packing is used in a block of data regarding an object; we only record the object id for the first entry. Access to the data is faster in semantic database since all facts about an object will be stored in contiguous disk blocks. One or a very few disk accesses are needed to satisfy some kinds of queries while in relational world, access to many tables maybe necessary thus causing many disk accesses per query.

A multi-user semantic database engine has been developed and is now in the testing phase. A user interface to this engine has been developed using C++ and is also being tested. A Java interface is nearing completion. We are also completing the development of a suite of tools for the semantic database engine: (1) text processing tools to manipulate our semantic data description language and to create databases and documentation and relational images of existing databases, now in alpha releases; (2) graphic tools to allow the creation and modification of semantic databases, which are now in the testing phase; (3) graphic tools for providing reporting functions on the database are partially functioning; (4) graphic tools to define and access a relational image of a semantic database are in the final testing phase.



Figure]: Semantic Editor Interface running on the WWW

One of Java based applications is a Semantic Editor. This is a tool, which works as an interface between the user and the semantic database, allowing the user to modify exiting semantic database schemas or create new ones. This is possible by creating new categories, and relations, modifying the attributes, and all the related information, and by being able to delete existing relations and categories. This is a portable and flexible application that can be run locally or remotely through the Internet. Figure 1 shows a picture of the system interface, after a semantic database schema for a university database was opened for modifications.

DATA ACCESSABILITY

At HPDRC, we have acquired from NASA Goddard Space Flight Center about 5G bytes of Landsat Thematic Mapper (TM) data observed by the Landsat 4 and 5 satellites. We have available 27 Quads of areas of the continental United States and 9 Scenes of areas of Brazil. For the storage of this TM data we used a Semantic Binary Database for the reasons mentioned in the above section. We have compressed the TM data prior to storage and so reduced the storage demands to about 2G bytes. We present here two Web based applications that demonstrate random rapid access to small parts of this compressed database via two distinct database views. The first view renders the compressed database as filtered rectangular false color maps of pixels. The second view renders the same database as a series of single sensor maps from which a movie can be generated

An HTML client program acts as the front-end application for the Landsat page. The user's database query is obtained by first showing a map of the region for which the database contains data. The user then selects an area (say Florida from a map of the U.S.). The client machine then displays an image from which the user selects a Landsat scene or quad. Finally from within the quad/scene the user selects the area to be displayed. Only a small portion of the total data set can be viewed in complete detail at one time owing to bandwidth and screen size limitations. Having selected the geographic area of interest, the user next selects the observation date from the dates presently available in the database for which this scene or quad was observed. The three sensors to apply as RGB colors for the false color composite image are next selected and, if required, an image-processing filter is chosen to modulate the desired image. The client program transmits the query data to a CGI script on the server, which makes a query to a view of the semantic database that yields a filtered false color image. This image is converted to JPG format to reduce the bandwidth demands on the network and returned to the client for display, this is shown in Figure 2.

In order to improve access to our Landsat database we have developed a program simulating a movie projector that allows the user to "fly" over the satellite images using the mouse or the keyboard to control the "flight path". The previous application provides access to satellite images, but because of the large size of the quad, the user is only able to see a small part of that image. The projector we have developed provides access to the entire satellite image, which is approximately 12M bytes, by gliding over the quad's surface in a continuous motion. The movie presently accesses only one sensor at a time for any given quad.

The movie system consists of a client/server combination and in this case the client side program must be downloaded from our site and installed as a browser helper. This helper program is written in the C++ programming language. It establishes a socket connection between the client and server through which the data is

transmitted. The user then chooses any Landsat quad in the database and about 2% of the central area of the quad is displayed in complete detail in a fixed size window. Then, using the mouse, the user can cause the window to move (or fly) over the remaining 980/. of the quad in any manner desired and at various speeds. Of course, it is possible to stop at any time and to "fly" backwards. Figure 2b below, shows the application along with the control panel. As the user navigates over the image the projector (client program) retrieves frames from the server. This server accesses the same compressed Landsat database that is used by the other applications described elsewhere in this paper. The server program makes database queries to a view of the Landsat database that generates movie sequences of the Landsat data that are computed in real time from the compressed database.

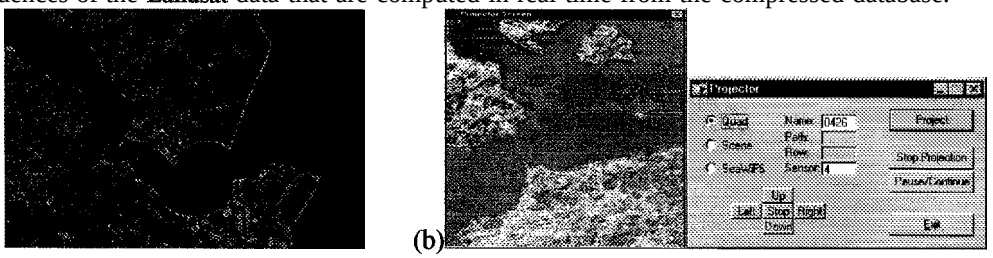


Figure 2(a)Landsat image (RGB=4, 3, 2) with Sobel edge detector (Jensen 1996).(b) Projector application with control panel.

A technique has been developed to reduce the bandwidth demands and requires less data transfer per frame. Transferring the border regions of the images that come into the new view instead of refreshing the entire image each time the frame changes does this. This technique reduced the bandwidth demands from a quadratic function of frame size to a linear function and will enable us to use a larger frame size or alternatively a faster refresh rate. In due course, Internet 2 should mitigate all the problems that arise from limited bandwidth.

The Ozone database at HPDRC contains all the data from the Total Ozone Mapping Spectrometer (TOMS) deployed on the Meteor-3 and Nimbus-7 satellites, and spanning a period of 15 years (1978-1994). The data is recorded for two different frequencies, monthly and daily. A total of about 0.5G bytes of Ozone Layer Thickness data are available from our on-line semantic database (Rishe 1992). We present here two Web based applications that demonstrate random rapid access to small parts of the database via a simple map view.

In order to manipulate, analyze, and display the information contained in this data set through different client platforms, an Ozone HTML/CGI based web page was developed. Through this page users can obtain detailed information about Ozone using queries to the spatial database and can retrieve textual information, as well as the Ozone maps from different dates, satellites, and frequencies. The query page is presented in Figure 4a. Once the user submits the projection parameters, the client passes the request to a CGI script that queries a database view of the Ozone database. The query returns a rectangular map of Ozone Layer thickness over the Earth's surface. This map is transformed on the fly to the requested projection and encoded as a GIF image for transfer to the client.

Through this page, users can view how the thickness of the Ozone layers change from season to season. Figure 4b shows the Ozone layers at the beginning of the year and as can be seen all values are greater than 230 Dobson Units (DU) (Asker 1993) (icair.iac.org.nz)(rtfm.mit.edu). Figure 4c depicts DU values less than 100 DU over the South Pole (Antarctic). These values were recorded by NASA's TOMS instrument flying on the Russia's Meteor-3 Satellite on October 6, 1993. On this day the ozone hole, the region where the total ozone column is less than 200 DU, covered 9 millions square miles, slightly smaller than the 1992 record ozone hole of 9.4 millions square miles.

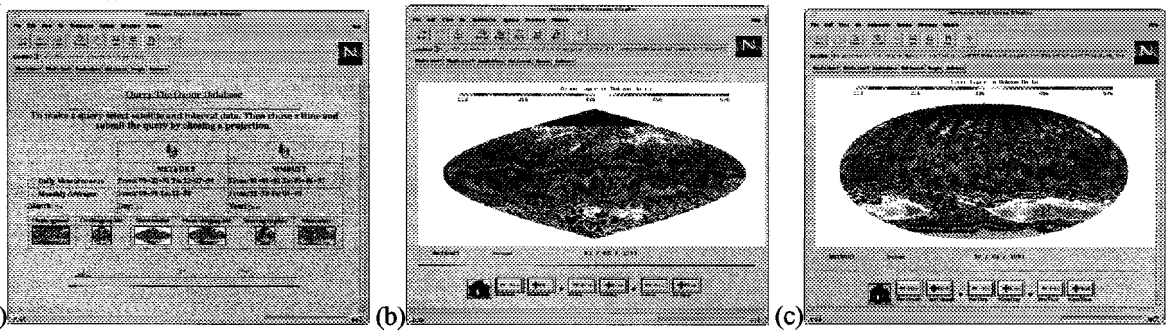


Figure 4.(a) The Query Page of the Ozone Web Page. The user enters a Satellite (Meteor-3 or Nimbus-7), Frequency (daily or monthly), Date, and then clicks on the projection desired to query the database (b) The result of a query from the Meteor-3 daily reading for January 1993 in the Sinusoidal projection. (c)The result of a query from the Meteor-3 daily reading for October 6, 1993 in the holographic projection. The light pink area over the South Pole represents the Ozone hole.

Virtual Reality Modeling Language (VRML) is a standard language describing interactive 3-D objects and worlds for delivery across the World Wide Web. The language is specified in a plain text file known as a VRML World. VRML is based on the SGI Open Inventor Format. VRML defines 3D graphics in terms of geometry, transformations, attributes, lighting, shading, and textures. Further VRML adds some language extensions to the Open Inventor Format to allow linking into the World Wide Web, and adds features and options to allow the user to choose the level of detail that the computer can handle efficiently (vag.vrml.erg) (hiwaay.net).

At HPDRC, VRML is used to produce a 3D viewing environment for scientific research applications using the Total Ozone Mapping Spectrometer (TOMS) for ozone layer thickness and ocean temperature data stored in a semantic database. These worldviews are created dynamically, exported through the World Wide Web, and then displayed locally on the client computer with a VRML viewer. The application provides a Web based form to allow input from the user. The application creates a worldview consisting of a set of two surfaces, the inner and the outer surface. The user chooses what spatial data to bind into each surface, that is to say what spatial data should be used to determine the colors rendered on that surface. The user also has the option to perturb each surface from the simple plane or sphere to an irregular "mountainous" terrain. Thus, in principal, the options would allow the simultaneous display of four spatial data sets; two to determine the color of the two surfaces and two to determine the "mountainous" perturbation of those surfaces.

Figure 5b shows a static picture of the VRML World View of two data sets, Ozone Layer Thickness and Ocean Temperature. The World contains two concentric spheres. The inner sphere is unperturbed and colored to represent Ocean Temperature with continents visible. The outer sphere is "colored" only by the Latitude and Longitude lines and otherwise is transparent gray. This outer sphere in greatly perturbed so that its radius represents the thickness of the Ozone Layer at the associated point on the surface of the Earth. The overall effect is that of the Earth with colored oceans representing temperature enclosed in a transparent gray plastic bag of Ozone on which are drawn latitude and longitude great circles. The huge dent in the bag over the South Pole clearly exhibits the Ozone hole.

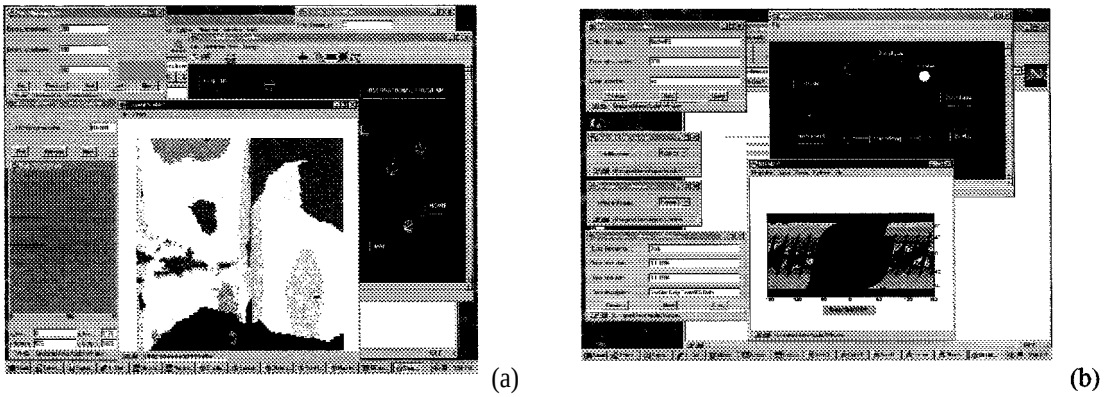
The VRML viewer (browser plug-in) on the client enables the World View, shown in Figure 5a, to be studied locally without further interaction with the server program that generated the view. The VRML viewer allows the user to rotate, zoom, spin in continuous motion, walk around, and indeed walk into the worldview, all without remote computation. However, the worldview itself represents a great deal of data. The worldview presented in Figure 5b, and available on our Web Site, is deliberately a low-resolution image (hence the blurred great circles) to ensure reasonable download time. However similar worldviews at higher resolution are possible on local clients and can be made more widely available over Internet 2. On local clients we have developed a similar application that uses the Inventor 3D Data format and presents the same 3D images that can then be viewed with 3D glasses and a suitable graphics workstation, this is shown on Figure 5c. The 3D-dataviewer system is implemented on the client using HTML and a VRML plug-in as shown on Figure 5c. On the server, a CGI script poses queries to a view of the database that presents spatial maps and then generates the VRML World View "on the fly" using the SGI Open Inventor package of tools.

Figure 5. (a) 3D Dataviewer HTML Form (b) VRML World: Ozone Perturbing the Outer Surface with Ocean Temperature in the Inner Surface(c) Local Version of 3D Dataviewer.

Another data set at HPDRC is simulated and real-time Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data. This data is derived from the broad area color imaging satellite, OrbView-2 satellite (formerly SeaStar) and SeaWiFS sensor. The SeaWiFS instrument is 1.1 Km multispectral sensor that acquires six visible and two near infrared channels for broad area applications such as fishing operations, harmful algae bloom monitoring, sediments monitoring, offshore oil and gas operations, agriculture and forestry management (Beck 1997). In order to efficiently store and manipulate this huge amount of data, we have created a SeaWiFS semantic database (Rishe, 1992) containing about 100 MB of simulated SeaWiFS data. We are in the process of building another spatial database to store the real-time OrbView-2 data. To view this data set, we presently have two Java based applications that portray random access to a small part of the data via a simple viewer interface.

In order to manipulate, analyze, and display the data contained in this data set through different platforms a SeaWiFS Java based web page was developed. Through this page, a user can retrieve textual and technical information about the SeaWiFS data including date/time, scan line, gain-tdi mirror mode, sensors, points-colors and many others, as well as the satellite coordinated SeaWiFS map. A user can select at random which area of the whole (1984x3560-pixel) image, they want to view and retrieve only this area from the database. Another feature about this system is that the user can request from the database a particular value of a point of the whole image, then the value (digital number) and the actual color are displayed on the window. Figure 6a shows a screen capture of the application running.

The other Java based application displays simulated SeaWiFS (SeaStar) data, Ocean Temperature data (supplied by the University of Miami Rosenstiel School of Marine and Atmospheric Science) and Ozone (TOMS) data. The main difference from this application and the previously describe is that it concentrates more in the image processing part and less in the technical information about the data. Using this cross-platform application a user can view the simulated SeaWiFS data geolocated and in different projections including Orthogonal, Orthographic, Sinusoidal, Holographic, Stenographic, and Macerator. Figure 6b portrays a screen capture of the system interface after retrieving some information.



(a) (b)

Figure 6: (a)Interactive System displaying some technical information and the map for the simulated Sea WiFS (SeaStar) data (b)Multi-dataset interactive system, after some information about the simulated SeaWiFS data that was retrieved from the spatial database.

SUMMARY

At the High Performance Database Research Center we are finalizing the development of a semantic database system to store remote sensed data as well as other Earth Science databases. Applications that use Internet technologies are used to overcome the access and display problems, which have limited the scientific community. These provide swift and efficient retrieval of huge amounts of data from the semantic database.

REFERENCES

Asker, J. R. "US/Russian satellite maps worst Antarctic ozone hole", Aviation Week & Space Technology, Vol. 139, No. 17, Ott 1993, p.72.

Rishe, N., Database Design: The Semantic Modeling Approach, McGraw-Hill, 1992.

Beck, J. M., Albasini, J., Hill, C., Techniques for Marketing SeaWiFS Ocean Color Satellite Imagery on the Internet, 1997.

Rishe, N., Accepted grant proposal to NASA. Grant NAGW-4080, 1993.

Jensen, J.R., Introductory Digital Image Processing-A Remote Sensing Perspective, 2nd cd., Prentice Hall, 1996.

http://vag.vrml.org/VRML_FAQ.html

http://hiwaay.net/~crispen/vrml/faq.html

http://icair.iac.org.nz/ozone/ozone.html

ftp://rtfm.mit.edu/pup/usennet/news.answers/ozone-depletion/intro

Investigation of Third Order Optical Nonlinearity and Reverse Saturable Absorption of Octa-alkoxy Metallophthalocyanines

Mohan Sanghadasa, In-Seek Shin, and Thomas A. Barr
Department of Physics, University of Alabama in Huntsville, Huntsville, AL 35899
Ronald D. Clark, Huaisong Guo, and Angela Martinez
School of Natural Sciences, New Mexico Highlands University, Las Vegas, NM 87701
Benjamin G. Penn
Space Sciences Laboratory, NASA Marshall Space Flight Center, Huntsville, AL 35812

ABSTRACT

In recent years, there has been a growing interest in the development of passive optical power limiters for the protection of the human eye and solid-state sensors from damage caused by energetic light pulses and also for other switching applications. One of the key issues involved is the search for appropriate materials that show effective reverse saturable absorption. Phthalocyanines seem to be good candidates for such applications because of their higher third order nonlinearity and the unique electronic absorption characteristics. A series of 1,4,8,11,15, 18,22,25-octa-alkoxy metallophthalocyanines containing various central metal atoms such as zinc, copper, palladium, cobalt and nickel were characterized for their third order nonlinearity and for their nonlinear absorptive properties to evaluate their suitability to function as reverse saturable absorbers.

1. INTRODUCTION

Optically nonlinear organic materials have attracted a lot of attention because of their fast and large nonlinear optical response due to the large hyperpolarizability of delocalized z-conjugated electrons. Furthermore, organic materials allow the tailoring of their chemical and physical properties by modifying the molecular structure. Charge transfer complexes, dyes, organometallic compounds and conjugated polymers are few groups of organic materials widely explored in this field, The magnitudes of the macroscopic nonlinear susceptibilities of these materials span a broad range depending on the resonance and non-resonance contributions.

This study is focused on one specific family of compounds known as phthalocyanines which is characterized by an extensive two-dimensional delocalized n-conjugated electron system. The tailoring of chemical and physical properties over a wide range of chemical modifications is possible in these materials by incorporating many different metal atoms into the ring and by substituting various fictional groups at peripheral sites. Phthalocyanines hold a considerable promise for the development of many nonlinear optical devices because of their large third order nonlinearity, unique electronic absorption characteristics, and high thermal and environmental stability. However, the major drawback in this family of materials is its poor volubility in organic solvents. In this study, a series of soluble octa-alkoxy metallophthalocyanines were synthesized and characterized for their optical nonlinearity including reverse saturable absorption (RSA)[1].

According to Beer's Law, the intensity of light exiting a material (I_o) is directly proportional to the intensity of light entering it (I_i). However, the absorption of light by some materials cannot be described by such a linear relation. In such materials, the ratio I_o / I_i may increase or decrease as the input intensity is increased. If this ratio increases, the material becomes more transparent with increasing Ii and such materials are known as saturable absorbers. On the other hand, some materials become stronger absorbers as Ii is increased, and therefore, such a property is referred to as RSA. Materials which exhibit RSA can limit the energy of short laser pulses and also shorten the pulses temporally by enhanced nonlinear absorption of the lagging edges of the pulses. In recent years, there is a growing interest in the development of passive optical power limiters. One of the key issues involved is the search for appropriate materials that show effective RSA. In an ideal optical limiter based on RSA, the light would propagate through the material with negligible absorption until the increasing incident fluence could no longer increase the output power.

In addition to the energy limiting and pulse shortening, reverse saturable absorber can also function as an optical element for power limiting and pulse smoothing because it can reduce the fluctuations in the laser amplitude and limit the power transmission through the material for long laser pulses. Reverse saturable absorbers which exhibit

the optical limiting effects show great promise for use in passive devices in the area of nonlinear optics and optoelectronics because of their special application potential. Since high-intensity visible light sources including lasers have become more prevalent in many fields, such devices for the protection of human eye and solid-state sensors from energetic light pulses and also for other switching applications are needed.

In this study, macroscopic third-order nonlinearity of a series of octa-alkoxy metallophthalocyanines was investigated. Samples in the form of thin films in polymethylmethacrylate (PMMA) matrix were used. They were further characterized by their nonlinear absorptive properties in liquid phase, hence to evaluate their suitability to function as reverse saturable absorbers.

2. SYNTHESIS

A series of 1,4,8,11,15, 18,22,25 -octa-alkoxy metallophthalocyanines (see Fig. 1) were synthesized in this study. Previous synthesis of octa-substituted metal phthalocyanines involved converting octa-substituted metal free phthalocyanines to octa-substituted metal phthalocyanines[2]. In this work, the appropriately substituted *phthalonitrile* was treated with lithium and 1-pentanol to form octa-alkoxy dilithium phthalocyanines, followed by refluxing a mixture of an octa-alkoxy dilithium phthalocyanines (dissolved in anhydrous ethanol or acetone) and the appropriate metal chloride (dissolved in anhydrous ethanol or acetone) to form the related octa-alkoxy metal phthalocyanines. Normally, this procedure is used for converting unsubstituted dilithium phthalocyanines to unsubstituted metal phthalocyanines[3]. In this work, it was found that this procedure was also suitable for converting substituted metal free phthalocyanines to substituted metal phthalocyanines.

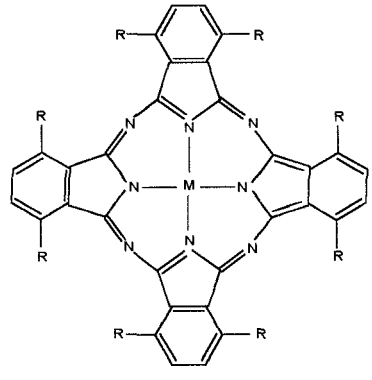


Figure 1 Molecular structure of 1,4,8,11,15, 18,22,25 -octa-alkoxy metallophthalocyanines
(R: butoxy/octyloxy/decyloxy/hexadecyloxy, M: Cu/Zn/Ni/Co/Pd)

Sublimation and dissolution in concentrated sulfuric acid followed by precipitation on ice are two classical methods used in the purification of phthalocyanines. They work well for unsubstituted phthalocyanines. However, for substituted phthalocyanines, these methods do not always work. In this research, successful purification methods were found to be column chromatography on silica gel using chloroform and ethyl acetate and extraction with chloroform-water and ethyl acetate-water. UV-VIS and ^{13}C-NMR spectra were used for final identification of the compounds.

All chemicals used in this study were purchased from Aldrich Chemical Company or Fisher Scientific unless otherwise indicated and were used without further purification. UV-VIS spectra were determined using a Perkin-Elmer model 553 Fast Scan UV-VIS Spectrophotometer. NMR spectra were obtained using a Bruker model AC - 300 MHz Spectrometer.

3. CHARACTERIZATION

The primary source used for optical characterization was a Nd:YAG laser (Spectra Physics DCR-11 with near Gaussian optics) operating in the Q-switch mode at a wavelength of 1064 mu at the pulse repetition rates between 3-10 Hz. The pulse duration of a single laser pulse was about 10 ns and the maximum energy was about 400 mJ. Absorption spectra of phthalocyanines typically include two absorption bands. One is due to the electronic transitions to the π-π^* states and is in the red region of the visible spectrum (Q band at 600-800 nm) and the other in the near *W* region (Soret band at 300-400 nm) as shown in Fig. 2.

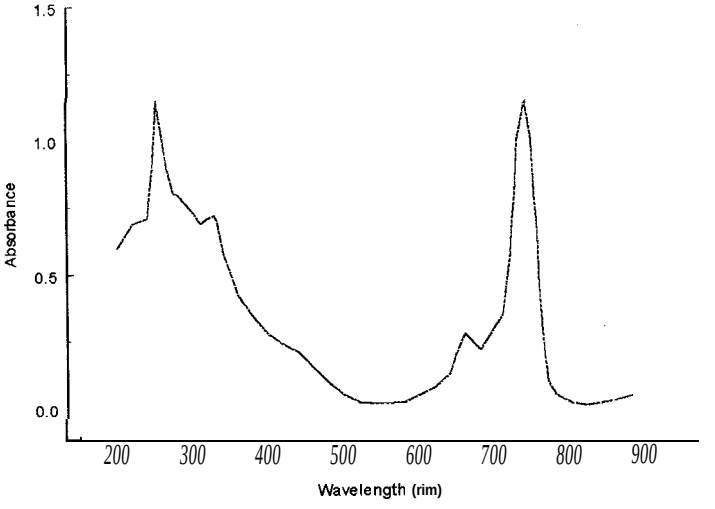


Figure 2. UV-VIS Spectrum of 1,4,8,11,15, 18,22,25 -octa-decyloxy palladium phthalocyanine in chloroform

Because of the unique absorption bands, the third order nonlinearity of these materials were investigated at the wavelength of 1907 nm to minimize the absorption at the third harmonic wavelength. RSA of these materials were investigated at the wavelength of 532 run. The linear absorption of these materials at this wavelength is not significant.

Detected signals either at 636 nm or 532 run wavelengths were amplified by fast pre-amplifiers (ORTEC, Model 9301) and then their average amplitudes were measured using a gated integrator and boxcar averager system (Stanford Research Systems, Model SR250) interfaced to a computer. The schematic diagram of the integrated experimental setups are shown in Fig. 3.

3.1 Third Order Nonlinearity

The third order nonlinearity of octa-alkoxy metal phthalocyanines was investigated using the third harmonic generation process. The 1064 nm laser wavelength was converted to 1907 nm wavelength by a 50 cm long Raman shifter filled with hydrogen gas at a pressure of about 7 MPa. The Raman antistokes at 1907 nm and any backscattering at 1064 nm due to Brillouin scattering were separated by using a dichroic mirror placed at 45° angle of incidence. Any residual 1064 nm radiation that transmitted through the mirror was blocked by using a Schott BG 18 broad band filter. A Corning 7-56 filter was used to block any visible light that might have originated from the Raman shifter.

Part of the beam was focused on a reference (vapor deposited polydiacetylene on a 1 mm thick quartz glass) which was used to cancel out laser fluctuations. The remainder of the beam was focused on the sample which was mounted on a stepper motor driven, computer controlled rotational stage (Ealing, Model 35-2500).

Samples used in this experiment were thin films of phthalocyanines in PMMA matrix . They were fabricated by spin coating mixtures of solutions of phthalocyanines and PMMA in chloroform onto a 0.25 mm thick 25x25 mm polished CFQ ribbon quartz plates (Heraeus-Amersil, Inc). Solutions of each compound in HPLC grade chloroform (Fisher Scientific)) were prepared with the concentrations in the range of 0.1 M to 1.0 M depending on the volubility and the availability of the material. Any undissolved particles were removed by filtering each solution using Acrodisc CR PTFE syringe filters with pore sizes of 0.45 μm and 0.2 pm. Each solution was mixed with 100 g/l solutions of PMMA (Kodak) with molecular weight of 350,000 in HPLC grade chloroform at 1:1 to 1:2.5 volume ratios. Spin coating was carried out using various rates of rotation and duration using a spin coating system (Headway Research Inc., Model 1-EC101D-R435).

The third harmonic light generated by both the sample and the reference was separated from the remaining fundamental beams by passing it through a Schott KG-1 broadband filter and an interference filter with a central wavelength of 636 nm. The signal was then detected by a photomultiplier tube (Hamamatsu, R4632) at each channel.

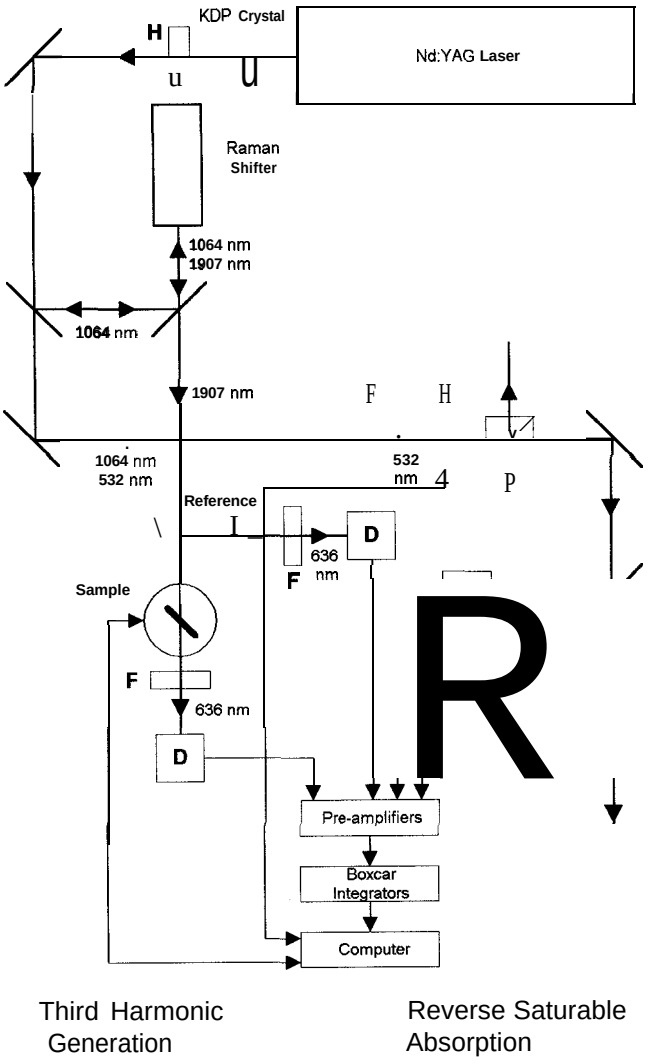


Figure 3. Schematic diagram of experimental setups for third harmonic generation and reverse saturable absorption. (D : detector, F : filters, H : half-wave plate, P: polarizing beamsplitter)

Third harmonic signal generated by the films was measured by varying the angle of incidence. At each run, about 2000 laser pulses were fired. The measurements were also repeated with various power levels of fundamental laser light. The third harmonic of this new wavelength was near the shoulder of the Q-band of some compounds and, therefore, appropriate corrections were made to the detected third harmonic intensity.

3.2 Reverse Saturable Absorption

The reverse saturable absorption properties of octa-alkoxy metal phthalocyanines were investigated by measuring the nanosecond nonlinear transmission in liquid phase. A linearly polarized 532 nm green light was obtained by generating the second harmonic light in a KDP crystal. The remaining 1064 nm light was removed by passing it through a solution of $CuSO_4$.

This second harmonic light passed through a half-wave plate and a polarizing beamsplitter before it was focussed into the solution with 1-cm pathlength. The half-wave plate was mounted on a computer controlled rotator so that the polarization of the green light could continuously be changed, hence the intensity of the light transmitted by the beamsplitter was varied continuously. The incident and transmitted pulses were reflected off beamsplitters before and after the sample and detected by high speed silicon photodiodes (Hamamatsu, S597 1). The maximum pulse

energy was chosen so that the intensity of the laser beam was below the damage threshold of the glass cell. The intensity of the laser beam was varied continuously and incident and transmitted intensities were measured simultaneously for each pulse. About 2000 laser pulses were fired at each run.

The measurements were made using solutions of octa-alkoxy metal phthalocyanines in chloroform with various concentrations ranging from zero to 6.8 x 10^{-4} M.

4. RESULTS AND DISCUSSION

By measuring the third harmonic intensity relative to fused silica, macroscopic third-order susceptibility $\chi^{(3)}(-3\omega;\omega,\omega,\omega)$ was determined Hence, the average value of microscopic nonlinear coefficient $\langle\gamma(-3\omega;\omega,\omega,\omega)\rangle$ was calculated. The values of $\langle\gamma(-3\omega;\omega,\omega,\omega)\rangle$ along with some other physical data such as the melting point and wavelength corresponding to the maximum absorption in the Q-band of each compound are listed in Table I.

Table 1. Average values of third order microscopic nonlinear coefficients, wavelengths (λ_{max}) corresponding to absorption peak in the Q-band and melting points of 1,4,8,11,15, 18,22,25-octa-alkoxy metallophthalocyanine

M	R	$\langle\gamma(-3\omega;\omega,\omega,\omega)\rangle$ (esu) x 10^{-34}	λ_{max} (rim)	M.P. ('C)
Metal Free	Butoxy	0.3	774	125.0 – 126.0
	Decyloxy	0.3	775	0.0 – 25.0
Cu	Butoxy	1.8	744	79.0 -81.0
	Octyloxy	2.9	758	0.0 – 25.0
	Decyloxy	2.3	747	0.0 – 25.0
	Hexadecyloxy	3.5	755	65.0 – 66.0
Zn	Butoxy	1.7	740	114.5 -115.5
	Decyloxy	1.3	741	0.0 – 25.0
Ni	Butoxy	0.7	744	0.0- 25.0
	Octyloxy	0.7	745	<-11.0
	Decyloxy	0.8	739	-11.0 –0.0
	Hexadecyloxy	1.0	745	0.0 – 25.0
Co	Decyloxy	1.8	725	-11.0 -0.0
Pd	Butoxy	0.6	741	<-11.0
	Decyloxy	0.8	741	-11.0 –0.0

According to the results in Table I, a significant improvement in the third-order nonlinearity has been achieved by the introduction of metal, especially in the case of copper phthalocyanines. However, the dependence of nonlinearity on the functional groups is not significant.

Fig. 4 shows the measured output fluence verses input fluence for 1,4,8,11,15, 18,22,25 -octa substituted Cu, Zn, Ni, Pd and metal free phthalocyanines with butoxy substituent group at various concentrations in chloroform. The same range of concentrations was used with all samples and the measurements were made under the same conditions. According to the data, the behavior of nonlinear absorption is not the same in all materials. Even the metal free phthalocyanines show a strong RSA and material with Ni shows the least effect. Further studies are in progress to explain the RSA in these molecules in terms of optical absorption and transition between different energy levels, which possess different absorption cross sections, and relaxation time constants. The explicit figures of merit for materials which qualify to be a reverse saturable absorber include the ratio of excited-state optical absorption cross section at the operative wavelength to the ground-state optical absorption cross section and the first excited state life time.

Experiments werem, performed at various pulse repetition rates to examine whether any thermal effects had an influence on the observed behavior of the nonlinear absorption. Our results did not indicate any such effects, and therefore, all the measurements were made at the rate of 10 Hz.

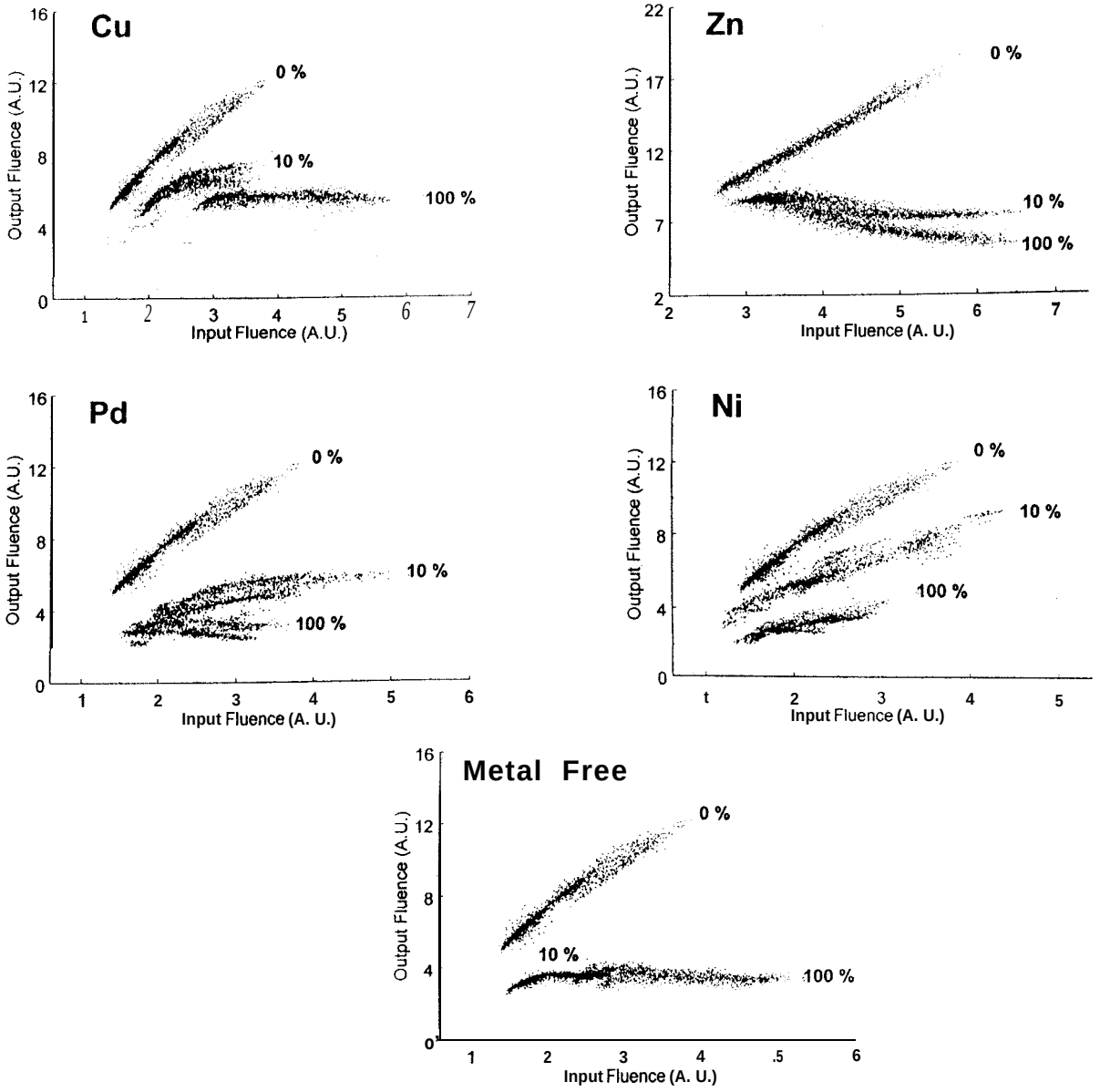


Figure 4. Measured optical fluence verses input fluence for octa-butoxy metal free phthalocyanines and metallophthalocyanines with Cu/Zn/ Ni / Pd at various concentrations. (100 % ≡ 6.8 x 10^{-4} M)

5. ACKNOWLEDGEMENTS

Support from the National Aeronautics and Space Administration (NAGW-4078, NAG8-1057 and NAG8-1288) is gratefully acknowledged. We thank Dr. C.C. Sung for helpful discussions and John Smith and Lu Loo for the assistance in the laboratory.

6. REFERENCES

[1] Giuliano, C.R. and Hess, L.D., J. *Quantum Electron,* **QE-3,** *338 (1967).*
[2] McKeown, N. B., Chambrier, I. and Cook, M. J., *J. Chem. Sot. Perkin Trans. I,* 1169 (1990).
[3] Barrett, P. A., Frye, D. A., and Linstead, R. P., *J. Chem. Soc.,* 1157 (1938).
[4] Sanghadasa, M., Wu, B., R.D. Clark, H. Guo, and B. G. Penn, in *Nonlinear Optical Properties of Organic Materials X,* M.G. Kuzyk, Editor, Proceedings of SPIE Vol. 3147, 185 (1997).

Characterization of the Sol-Gel Transition for Zirconia-Toughened Alumina Precursors

L. Moeti, E. Karikari, and J. Chen

Department of Engineering
Clark Atlanta University, Atlanta, GA 30314

ABSTRACT

High purity ZTA ceramic powders with and without yttria were produced using metal alkoxide precursors. ZTA ceramic powders with varying volume percents of zirconia were prepared (7, 15, and 22%). Aluminum tri-sec butoxide, zirconium propoxide, and yttrium isopropoxide were the reagents used. Synthesis conditions were varied to control the hydrolysis and the aging conditions for the sol to gel transition. FTIR analysis and theological characterization were used to follow the structural evolution during the sol to gel transition. The greater extent of hydrolysis and the build-up of structure measured from viscoelastic properties were consistent. Heat treatment was conducted to produce submicron grain fully crystalline ZTA ceramic powders. In all experimental cases a-alumina and tetragonal zirconia phases were confirmed even in the absence of yttria.

INTRODUCTION

There is currently a great interest in the development of new materials for advanced structural applications, for example engine components in high technology aerospace applications. These new materials should have improved properties such as strength, toughness, and wear resistance. Ceramic oxide composites are increasingly becoming desirable as materials for these applications. Tetragonal zirconia (ZrO_2) - toughened alumina (Al_2O_3) has become an area of increased technical interest in recent years [1-3]. By the introduction of dispersed ZrO_2 into an alumina matrix the resulting zirconia-toughened alumina (ZTA) has demonstrated improved toughness and strength when compared to pure alumina [4]. The stress induced tetragonal-to-monoclinic ($t \rightarrow m$) phase transformation and the stress induced microcracking [5-7] are the principal toughening mechanisms in ZTA ceramics. Factors such as particle size, particle size distribution, and the nature of the polymorph contribute to the toughening mechanisms and strength enhancement. These factors can be modified by the processing techniques used in the fabrication of ZTA ceramics.

The sliding wear resistance of ZTA ceramics [8] have shown improved properties when compared to Al_2O_3 or tetragonal ZrO_2 ceramics. An extensive amount of research has been conducted on the solid particle erosion wear of brittle materials such as Al_2O_3, silicon carbide/silicon nitride (SiC/Si_3N_4) composites, tetragonal zirconia, and whisker-reinforced composites [9- 12]. However, the erosion wear resistance of ZTA has only recently been investigated in a systematic fashion [13].

Most of the research conducted in the development of ZTA ceramics has been either by colloidal/powder processing [14] or by partial chemical routes [15]. The disadvantages of these methods include the impurities in the starting materials and the

difficulty in achieving good uniform distribution of the zirconia dispersed in the alumina matrix. In the present research a chemically based approach using high purity starting materials consisting of zirconium and aluminum metal alkoxides was used to control the precursor chemistry and particle sizes and also achieve improved uniform distributions of ZrO_2 dispersed in the Al_2O_3 matrix. By carefully controlling the precursor chemistry high purity ZTA ceramics with smaller grain sizes were prepared which should improve the toughness and wear erosion properties of ZTA ceramics. In the development of ZTA, yttria (Y_2O_3) additions are often used to stabilize the ZrO_2 in the tetragonal state. Without the presence of Y_2O_3 to stabilize the zirconia in the tetragonal state, ZrO_2 transforms from tetragonal to monoclinic below 1100°C. Extremely fine particles of ZrO_2 are known to be stable [16] in their tetragonal state even in the absence of solid solutions with other oxides such as MgO, CaO, Y_2O_3, or CeO_2.

EXPERIMENTAL APPROACH

High purity starting materials were used to synthesize ZTA ceramic powders with 7, 15, and 22 volume percent of zirconia. Aluminum tri-sec butoxide (ATSB), zirconium propoxide, and yttrium isopropoxide were the reagents used. Triethanolamine (TEA) was also used to stabilize the ATSB by the formation of chelating complexes between the ATSB and the TEA which reduced the reactivity of ATSB to water. The solvent used for all experiments done was sec-butanol (2-butanol). Precursor sols both with and without yttria to were prepared to determine if the zirconia phase would remain in the tetragonal state in the absence of yttria. Hydrolysis conditions were established such that sols could be gelled in several hours or days depending on water content and amount of TEA used. Table 1 shows the synthetic conditions used to prepare the ZTA precursors. FTIR spectroscopy was used to follow the structural evolutions in the precursor sol to gel transition. Theological properties of the precursor sols and gels were also monitored and correlated to the synthesis conditions. ZTA precursor gels obtained were heated to temperatures up to 1300°C and the crystallinity and microstructure of the final ceramics were examined by X-ray analysis and Scanning Electron Microscopy (SEM), respectively.

RESULTS AND DISCUSSION

Figure 1 shows the FTIR spectra of a ZTA (with 22 volume percent zirconia) precursor gel with yttria (6 mol percent to zirconia) after 2 hours of aging. The greater extent of hydrolysis of the gel in comparison to the sol was confirmed by the larger peak at 3100 cm-l wave numbers. This is consistent with the further extent of reactions and the build-up of structure in the gels as compared to the sols. The theological properties of the sol to gel transition were followed using both steady shear and dynamic viscosity. From steady shear experiments it was observed that the sols in the initial stages show slightly shear thinning behavior. From dynamic viscosity measurements the sols show a small elastic component during the early stages of aging. Aged viscous sols and the gels are observed to have a larger elastic component which is again consistent with the build up of structure. Figure 2 shows the storage (G') and loss (G") modulus of a ZTA precursor gel (with 22 volume percent zirconia and 6 mol percent yttria) after 3 days of aging. The theological data is in agreement with the observations of structural change as observed from FTIR characterization of the aging sols and gels. Heat treatment was conducted to prepare ZTA powders. Figure 3 shows a TG/DTA plot of the ZTA precursor gel (15 vol % zirconia) heated to 1400°C. A weight loss of 50% from the gel

to the final ZTA ceramic is observed. The exothermic peak at 300°C is due to the volatization of organics and the peak at 900°C is due to the formation of y-alumina. The peak beginning at 1125°C is attributed to the formation of α alumina The microstructure of the ZTA ceramic powders was examined using SEM and the crystalline phases were examined using x-ray analysis. Sub micron grain size, fully crystalline high purity ZTA ceramic powders were produced using the sol-gel processing approach. Figure 4 shows the x-ray plot of a ZTA powder with 15 volume percent zirconia without yttria heated to 1200°C. From x-ray phase analysis α alumina and tetragonal zirconia were the phases identified, indicating that tetragonal zirconia is retained even without the presence of yttria which was consistent with the observation [16] that fine particles of ZrO_2 are stable in their tetragonal state even without solid solutions such as Y_2O_3.

CONCLUSIONS

High purity ZTA ceramic powders with and without yttria were produced using metal organic precursors. Synthesis conditions were varied to control the hydrolysis and the aging conditions for the sol to gel transition. FTIR analysis and theological characterization proved useful in following the structural evolution during the sol to gel transition. Heat treatment produced submicron grain fully crystalline ZTA ceramic powders which was confirmed by SEM and x-ray analysis. Even without the presence of yttria, the tetragonal zirconia phase was retained.

ACKNOWLEDGMENTS

Support provided by the NASA Faculty Award for Research (FAR) Grant No. NAG4- 11 and by the Clark Atlanta University NASA supported High Performance Polymers and Ceramics (HiPPAC) Center Grant No. NAGW-2939 is gratefully acknowledged. The use of equipment provided by the Center for Environmental Policy, Education and Research (CEPER) under Environmental Protection Agency (EPA) Assistance ID No. CR 818689 and support provided by the Army Research Office Grant No. DAA L03-92-G-0380 is also gratefully acknowledged.

REFERENCES

1. N. Claussen, J. Am. Ceram. Sot., 59 [1-2] 49-51 (1976).
2. N. Claussen, J. Steeb, and R.F. Pabst, Am Cer. Sot. Bull., 56 [6} 559-62 (1977).
3. P.F. Becher, J. Am. Ceram. Soc.,**64** [1] 37-39 (1981).
4. H. Kamiya, M. Takatsu, and A. Hattori., J. Ceram. Jpn. 98 [5] 456-63 (1990).
5. A.G. Evans and K.T. Faber, J. Am. Ceram. Sot., 64 [7] 394-98 (1981).
6. R. McMeecking and A.G. Evans, , J. Am. Ceram. Soc.,**65** [5] 242-46(1982).
7. A.G. Evans and R. M. Cannon, Acts. Mettal., 34 [5] 761-800 (1986).
8. H. Kamiya and M. Takatsu, J. Ceram. Sot. Jpn. 98 [1] 13-21 (1990).
9. F.F. Lange and A.G. Evans, , J. Am. Ceram. Sot., 62 [1-2] 62-65 (1979).
10. S. Wada and N. Watanabe, J. Ceram. Sot. Jpn. 96 [5] 599-602 (1988).
11. S. Wada and N. Watanabe, J. Ceram. Sot. Jpn. 97 [2] 131-35 (1989).
12. S. Srinvason and R.G. Sparks, J. Am. Ceram. Sot., 73 [5] 1421-24 (1990).
13. A.G. Evans and T.R. Wilshaw, Acts. Metall., 24,939-56 (1976).
14. C.Hu and M. Rahaman, J. Am. Ceram. Sot. 77 [3] 815-19 (1994).
15. K. Ranjbar b. Rae, and T. Mohan, Am. Cer. Sot. Bull. 73 [2] 63-66 (1994).
16. B.E. Yoldas, J. Am. Ceram. Sot. 65,387 (1982).

TABLE **1.** Effect **of Synthesis Conditions on Gelation Time**

Sample	ATSB (mol 102)	ZRP (mol 102)	Y I P (mol 10~)	sec-butanol	H_2O/Total Alkoxide (ml)	TEA/Total Alkoxide (molar ratio)	ZrO_2 (volume percent)	Gel Time (hours)
1	**4.75**	2.231	0	30	2	0.302	7	8
2	**6.09**	2.856	0	40	2	0.321	7	24
3	**6.09**	2.856	0	40	2	0.332	7	48
4	**6.09**	6.693	0	40	2	0.333	15	10
5	**6.09**	6.693	0	40	2	0.344	15	24
6	**3.81**	6.827	0	30	2	0.284	22	1
7	**3.81**	6.827	0	30	2	0.301	22	6
8	**3.81**	6.827	0.804	30	2	0.328	22	10
9	**3.81**	6.827	0.804	30	2	0.345	22	40
10	**6.09**	6.693	0	40	4.5	0.388	15	192
11	**4.75**	2.231	0	30	4.5	0.347	7	0.5

ATSB Aluminum tri-sec butoxide
ZRP Zirconium propoxide
YIP Yttrium isopropoxide
T E A Triethanolamine

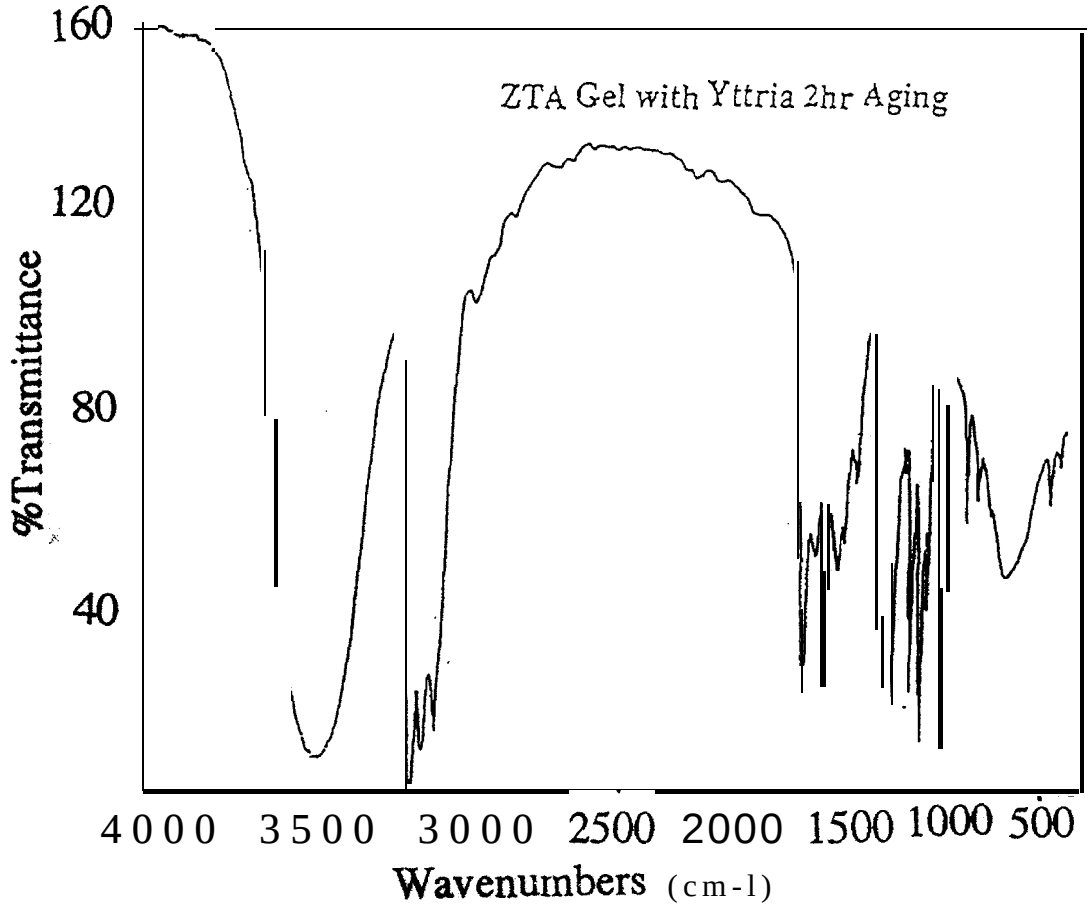


Figure 1 FTIR of ZTA Precursor Gel

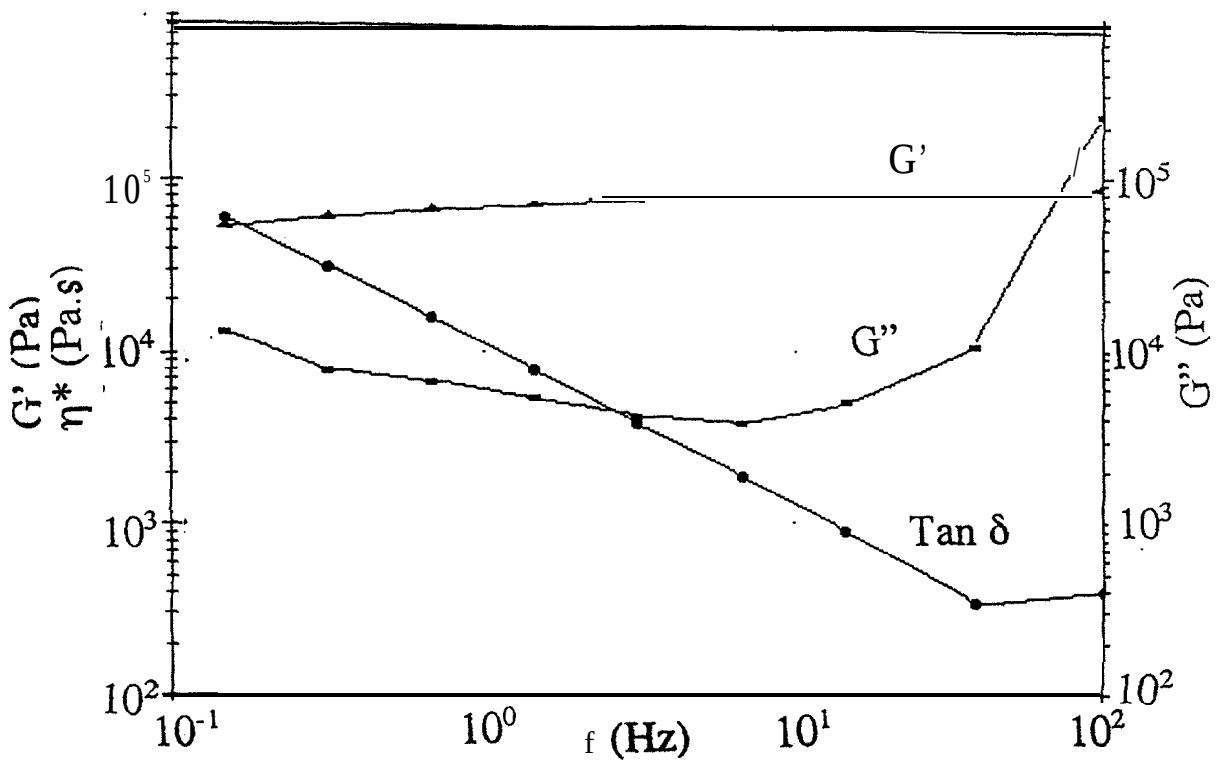


Figure 2 Viscoelastic Properties of ZTA Precursor Gel

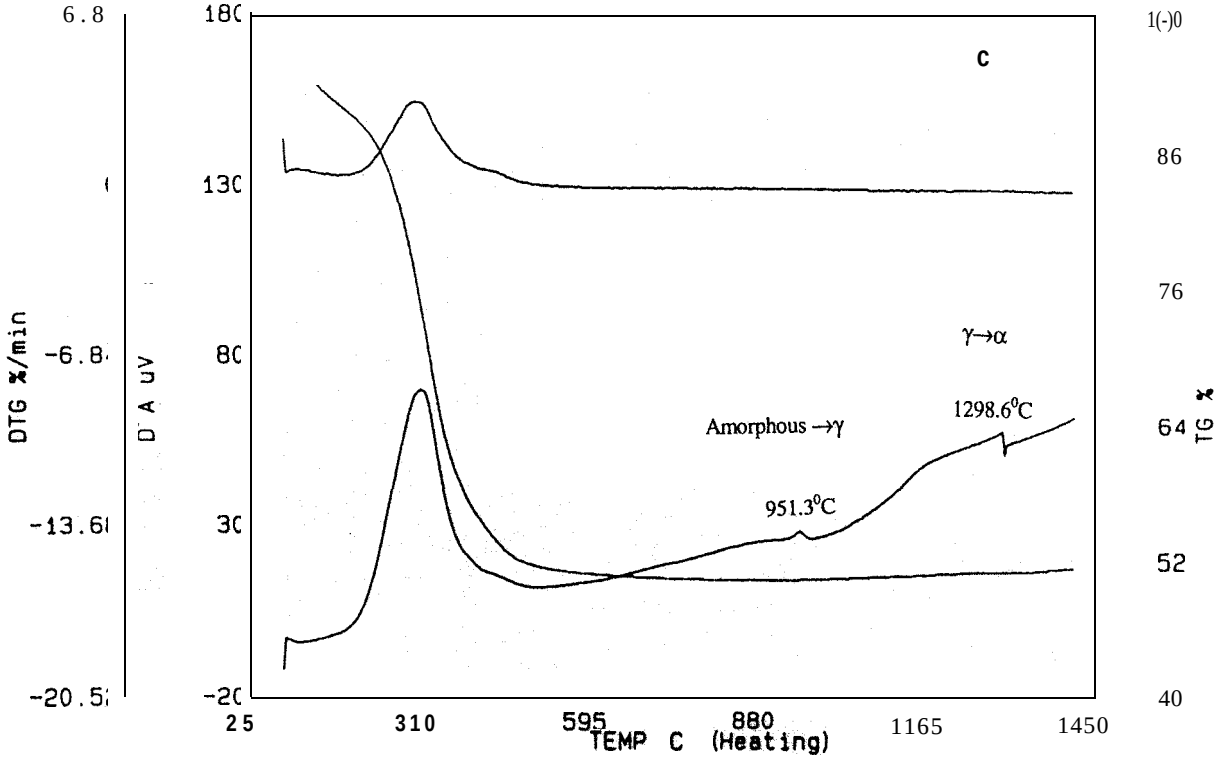


Figure 3 Thermal Analysis of ZTA Precursor Gel

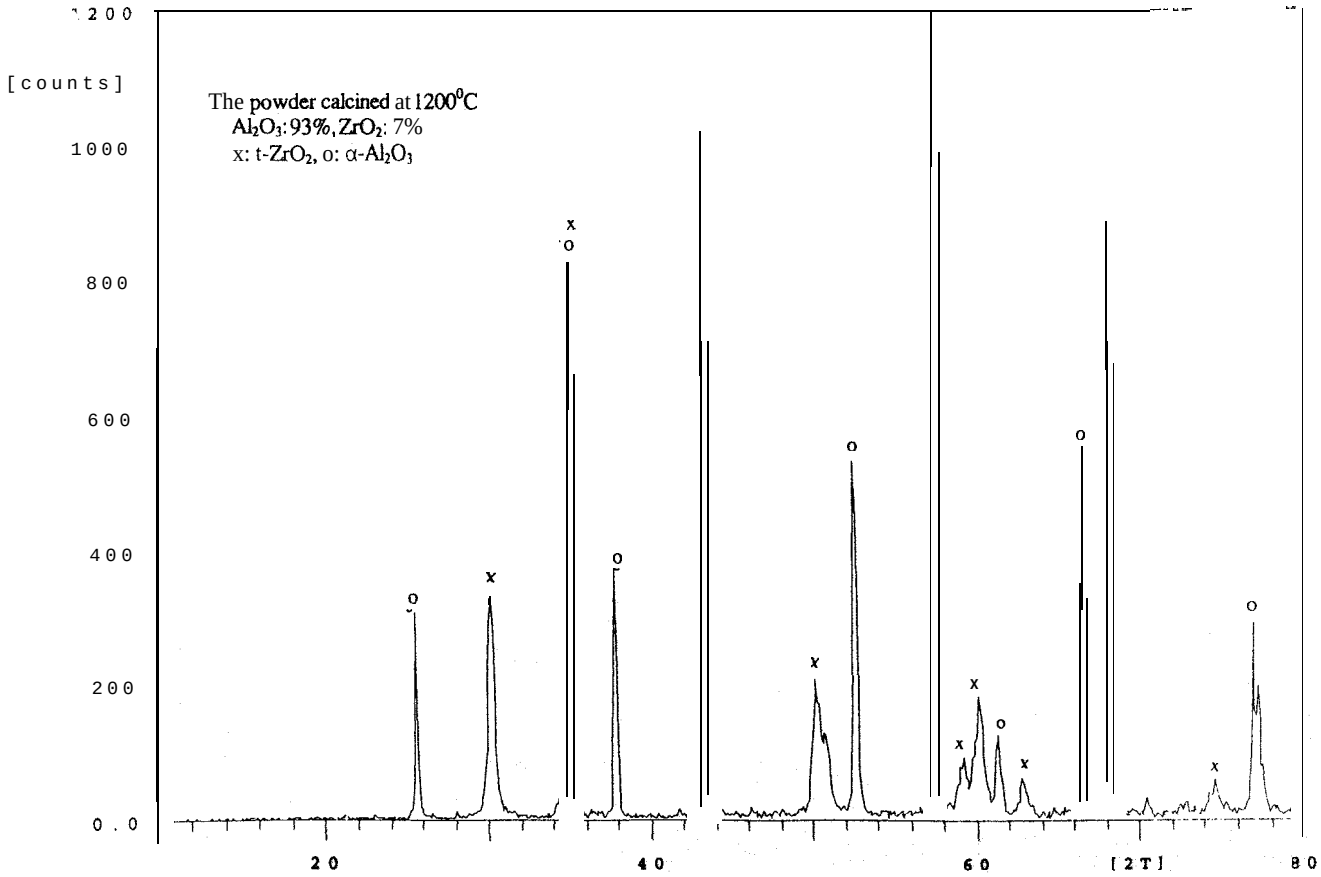


Figure 4 X-ray Analysis of ZTA Ceramic Powder

"Oxygen Impurities and Defects in Epitaxial Layer SiC and SiC Wafer characterized by Room and Low Temperatures FTIR"

W. J. Lu, W. E. Collins, D. T. Shi, Y. S. Tung
Center for Photonic Materials and Devices
Department of Physics
Fisk University
Nashville, TN 37208

D.J. Larkin
M.S. 77-1, NASA Lewis Research Center
Cleveland, OH 44135

Abstract

SiC as a highly promising semiconducting material has received increasing attention in the last decade. The impurities such as oxygen and hydrogen have a great effect in electronic properties of semiconducting materials.

In this study, the FTIR spectra were measured at room temperature (25 "C) and low temperature (-70 "C) for an n-type SiC substrate, a p-type epitaxial layer SiC, and patterned Ta on a p-type epitaxial layer SiC sample. The oxygen related IR peaks were measured for all three samples at room and low temperatures. The peak at 1105 cm-* is the result of a substitutional carbon and a interstitial oxygen in SiC. The concentration of the impurity oxygen increases in the SiC epitaxial layer during the CVD and electron beam processes. For the n-type SiC substrate, this peak does not appear. The peak at 905 cm^{-1} exists in the IR spectra only for two epitaxail layer on p-type SiC substrate samples. This peak is related to oxygen vacancy centers in SiC, which are introduced in the CVD epitaxial growth process. At low temperature, the peak at 1105 cm-l shifts down and the peak at 905 cm^{-1} shifts up for the epitaxial layer SiC samples. It can be explained that, at low temperatures, the stress increases due to the distorted bonds. The study shows that FTIR is a very effective method to evaluate low concentration impurities in SiC.

Key words: SiC, low temperature FTIR, and impurities in semiconductors.

1. Introduction:

SiC, as a high promising semiconducting material, has received increasing attention in recent years. SiC based devices are able to operate at high temperature, in high radiation environments, and has strong mechanical properties, Since the quality of SiC wafers is much less than the traditional semiconducting materials, the fundamental research of the characterization and evaluation of SiC wafers is a very important issue.

The bulk SiC crystal is grown in a physical vapor transport system [1]. Several groups (Cree Research, ATMI, and Northrop-Grumman Research Center) are working toward SiC substrate fabrication. Although the quality of SiC wafers have been much improved, a lot of defects such as micropipe (1 0-103 cm^{-2}), mosaic structures (10-15 cm^{-2}), dislocations (104-105 cm^{-2}), and poly-typism (3C, 4H, rhombohedral in 6H, and 3C, 6H, rhombohedral in 4H) exist for commercially

available SiC wafers. The impurities and dislocations have a great effect on electronic properties of semiconducting materials. The defects in SiC are very complicated. Micropipe and the mosaic structures can be observed by optical microscopy. Polytypes can be studied by Raman spectroscopy, Low temperature FTIR is a potentially excellent technique in studying the defects and impurities in SiC materials.

One attempt to improve the quality of SiC is to grow a homoepitaxial SiC layer on SiC substrate. Larkin *et al.* **[2,3]** at NASA Lewis Research Center has successfully fabricated a homoepitaxial SiC layer on the SiC substrate by "Site-Competition" chemical vapor deposition method. This technique has improved the surface state, and increased the doping concentration in SiC significantly.

The Fisk SiC research program, with research cooperation with Dr. Larkin at NASA Lewis Research Center, has been working on the characterization and fabrications of SiC epitaxial layers, metal thin film contacts and interracial diffusion on SiC, and SiC molecular spectroscopy for years. George *et al.* **[4]** analyzed a n+ SiC epitaxial layer using XPS combined with Ar sputtering. The n+ epitaxial layer has higher concentrations of oxides, and the oxides enhances adhesion between the metal thin film and the SiC substrate. The CVD process also introduced oxygen and other impurities in the exitaxial layer. Lu *et al.* **[5]** analyzed the SiC epitaxial layer and substrate samples by Raman spectroscopy. It was found that the SiC substrate and SiC epitaxial layer samples have a small peak at about 520 cm-l, which is assigned to transverse optical like mode of crystalline silicon [6]. Larkin *et al.* reported the detection of hydrogen in the CVD SiC epitaxial layers [7] since the PVT and CVD process are in the hydrogen environments.

Based on our previous studies, we are presenting a preliminary exploratory study of oxygen impurities in the epitaxial layer SiC and SiC substrates using FTIR.

2. Experimental:

Three types of silicon carbide samples were used in this study. All samples were provided by NASA Lewis Research Center.

(a) A n-type 6H-SiC substrate (sample # 1900) is commercially available [18].
(b) The n-type homoepitaxial SiC layer on 6H-SiC substrate is produced by chemical vapor deposition (CVD). The substrates were heated at 1400 to 1500 "C via the rf-coupled susceptor with an automatic optical pyrometer for temperature controlling (sample #1 895-6).
(c) A patterned tantalum on SiC sample was prepared by e-beam depositing tantalum on a n-type epitaxial layer on the p-type 6H-SiC substrate (sample #1 662-4).

Before experiments, the samples were cleaned by HF and acetone, then rinsed by deionized water, and finally, dried in nitrogen flow at 110 "C, so that the contaminated chemicals on the surface are removed.

FTIR spectra were obtained from the Boman MB 102 spectrometer with a resolution of 4 cm-l using a KBr beam-splitter and a MCT detector, The spectra were obtained by an average of 300 scans from 400 to 1500 cm^{-1} at room temperature and at approximately -70 "C.

3: **Preliminary Results and Discussions.**

FTIR has not been applied to detect the impurities in SiC wafers. Room temperature and low temperature FTIR have been successfully applied to measure the concentration of the impurity atoms and evaluate the quality of Si crystals [5, 9, 10]. The sensitivity of this method is very high. Pajot [10] provided an excellent review about oxygen and other impurities in Si wafers and other semiconductors.

Low temperature infrared spectroscopy has become a standard method to measure the oxygen and carbon concentrations in silicon for semiconductor industry [9]. The most common contaminants in silicon ingots made by the Czochlarski method are oxygen and carbon. The impurity ranges of 10-20 ppm atomic oxygen and 2-20 ppm carbon is expected from commercial silicon wafers. Mead *et al.* [8, 11] reported that the carbon and oxygen occupying substitutional positions in the silicon lattice can be detected by low temperature infrared spectroscopy. The peaks at 605 cm^{-1} and 1105 cm-l represent the vibrational spectra of the silicon-carbon and silicon-oxygen respectively. The detection limit is about 100 ppb. Shirakawa *et al. [12]* further pointed out that the peak at 1105 cm-l is composed of a substitutional carbon and a interstitial oxygen. At the liquid helium temperature, the interstitial oxygen and the precipitate oxygen peak can be distinguished. The interstitial oxygen peak shifts from 1105 cm-l to 1136 cm^{-1} and 1205 cm-l and becomes very sharp, the precipitate oxygen peak is still at 1105 cm-l [13].

Suttrop *et al.* [14] and Gotz *et al.* [15] studied the low-temperature infrared spectra of nitrogen doped 6H-SiC and 4H-SiC. To our knowledge, the infrared spectra of oxygen and hydrogen impurities in SiC has not been reported, although some basic research on SiC IR has been done [14,15, 16, 17].

Since the oxygen impurity is studied in this work, we only examined the IR spectra from 700 cm-l to 1200 cm-l. For the n-type SiC substrate (sample #1900), Figure 1 shows the IR spectra at room and low temperature. There is a broad peak at 979 cm-l, and two very broad shoulder peaks at about 1142 cm-l and 810 cm-l. These peaks show no differences at room temperature and low temperature.

For the epitaxial layer on p-type SiC substrate (sample # 1895-6), as shown in Figure 2, the IR spectrum at room temperature has a very similar shape, but a small peak shift with the spectrum at the low temperature. There are IR peaks at 905 cm-l, and 1105 cm-l at room temperature. At the low temperature, the peak at 905 cm^{-1} shifts to at 915 cm-*, and the peak at 1105 cm-l shifts to 1097 cm^{-1}.

The IR spectrum in Figure 3 is for the patterned Ta on n-type epitaxial layer SiC on p-type 6H-SiC substrate (sample #1662-4) at room temperature. As shown, it is similar to that of the epitaxial layer SiC (sample # 1895-6) in Figure 2, eg; there are two peaks at 1105 cm-l and 905 cm-l. At the low temperature, the peak at 905 cm-* shifts to 937 cm-l and the peak at 1105 cm-l shifts 1085 cm-l.

As stated, the peak at 1105 cm^{-1} is the result of interstitial oxygen atoms, also included substitutional carbon atoms [8, 10,11, 12] for Si crystals. In SiC wafers, it is reasonable to assume that the peak at 1105 cm-l also is the result of interstitial oxygen atoms, and substitutional carbon atoms, and the impurity atoms are associated with Si atoms in SiC structures. Our previous result

[4] showed that the oxygen concentration increased in the SiC epitaxial layer when analyzed by XPS combined with Ar sputtering. For the SiC epitaxial layer and Ta pattern on SiC epitaxial layer on p-type SiC substrate samples (Figures 2 and 3), this peak at 1105 cm-l is very clear. This can be explained if the oxygen impurity atoms were introduced in SiC during the CVD and electron beam processes, so the concentration of oxygen impurity increased. There is no peak at 1105 cm-l for the n-type SiC substrate sample (Figure 1). An obvious explanation is that the concentration of the oxygen impurity atoms in SiC substrate is much lower, and it can not be detected at room temperature.

To our knowledge, the peak 905 cm-l from SiC wafers has not been reported. From the studies of the impurity in silicon [1 O], the peak at 889 cm^{-1} maybe related to the oxygen vacancy centers [19, 20]. Brelot [21], Lindström *et al. [19]* and Stein [22] observed the peak at 936 cm-l in silicon crystals, and it may be caused by Si_iO_{1-i} (an interstitial Si atom trapped by an interstitial O atom). From our experimental results, the peak at 905 cm-l is shown in the SiC epitaxial layer and Ta pattern on SiC epitaxial layer on p-type SiC substrate samples (Figures 2 and 3). Both samples are on p-type SiC substrates. The n-type SiC substrate in Figure 1 shows a peak at 979 cm-l, not at 905 cm^{-1}. Therefore, the peak at 905 cm-l is possibly related to the doping conditions, and is the result of an increase of the concentration of the oxygen vacancies in the SiC epitaxial layer growth during the high temperature CVD and electron beam processes, which changed the oxygen concentration [4]. Once again, there is no peak at 905 cm-l for the n-type SiC substrate sample (Figure 1). The concentration of the oxygen vacancy in n-type SiC substrate is obvious much lower, and it cannot be detected at room temperature. Further study is needed to understand the nature of the peak.

It is very interesting to observe the shifts of the peaks at 1105 cm-l and 905 cm-l for the epitaxial layer SiC samples at low temperature (Figures 2 and 3), and the two samples are both p-type substrates. There are no shifts for the n-type SiC substrate sample (Figure 1). The impurity oxygen atoms are at the interstitial sites in SiC lattices, The O-Si bonds and the O-Si vacancies are distorted. The low temperatures increase the stress caused by the distortion, which results in the shifts in IR spectra. The electron beam deposition in Ta thin film fabrication process accelerates the formation of the oxygen-vacancies. This phenomenon does not occur with the n-type SiC substrate sample.

Comparing these three samples, the oxygen concentration for epitaxial layer SiC is larger than the SiC substrate. This result is consistent with our previous result using XPS and Ar sputtering [4]. The impurity oxygen atoms are obviously distorted in the SiC lattices at low temperature.

4. **Conclusions:**

This preliminary study is the first time to report the detection of the oxygen impurity atoms in SiC materials using FTIR. The peak at 1105 cm-l is the result of interstitial oxygen atoms, and substitutional carbon atoms, and the impurity atoms are associated with Si atoms in SiC structures. The peak at 905 cm-l is possibly related to the doping conditions, and it is the result of an increase in the oxygen vacancy concentration in the p-type SiC substrate. The CVD and electron beam processes increase the concentration of oxygen impurity in SiC. The lower temperature FTIR, at liquid nitrogen, will be performed in the future.

Acknowledgment:

This work was supported by NASA under Grant No. NAGW-2925, NAG3-2012, and NAG3-1430.

References:

[1] R. C. Glass, D. Henshall, V. F. Tsvetkov, and C.H. Carter, Jr., *MRS Bulletin,* March, 1997, p30.

[2] D.J. Larkin, P.G. Neudeck, J. A. Powell, L.G. Matus, *App.Phys.Lett., 65(13), 1659 (1994).*

[3] P. G. Neudeck, *J. Electronic Materials, 24(4), 1995.*

[4] M. A. George, D.J. Larkin, J. Petit, A. Burger, S.H. Morgan, W.E. Collins, *MRS Anneal Meetings,* Boston, MA, 1995.

[5] W. Lu, W. E. Collins, S.H. Morgan, D. Shi, D.J. Larkin, *The Eastern Regional Conference on Crystal Growth & Epitaxy,* Atlantic City, NJ, Sept. 28-Oct. 1, 1997.

[6] S.F. Yoon, R. Ji, J. Ahn, and W.I. Milne, *Journal of Materials Science, 32,* 1163(1997),

[7] D.J. Larkin, S.G. Sridhara, R.P. Devaty, and W.J. Choyke, *Journal of Electronic Science, 24(4), 289(1994).*

[8] D.G. Mead, *Applied Spectroscopy, 34(2), 171(1980).*

[9] B. G. Rennex, *"Certification of a standard reference material for the determination of interstitial oxygen concentration in semiconductor silicon by infrared spectrophotometry"*, National Institute of Standards and Technology, Washington, DC, 1994.

[10] B. Pajot, "Some Atomic Configuration of Oxygen", *Semiconductors and Semimetals*, Vol.42, F. Shimura (Ed.), Chapter 6, Academic Press, 1994, p191-249.

[11] D.G. Mead, and S.R. Lowry, *Applied Spectroscopy, 34(2), 167(1980).*

[12] Y. Shirakawa, H. Yamada-Kaneta, and H. Mori, *J. Appl. Phys.*, 77(1), *41(1995).*

[13] H. One, and T. Ikarashi, *Appl. Phys. Lett., 63(24), 3303(1993).*

[14] W. Gotz, A. Schoner, G. Pensl, W. Suttrop, W.J. Choyke, R. Stein, and S. Leibenzeder, *J Appl. Phys., 73(7), 3332(1993).*

[15] M.F. MacMillan, R.P. Devaty, W.J. Choyke, D.R. Goldstein, J.E. Spanier, and A.D. Kurtz, *J. Appl. Phys., 80(4), 2412(1996).*

[16] F. Engelbrecht, and R. Helbig, *Physical Review B, 48(21), 15698 (1993).*

[17] L. Chen, T. Goto, T. Hirai, *J. Materials Science, 25,4273 (1990).*

[18] Cree Research Inc., 2810 Meridian Parkway, Durham, NC 27713.

[19] J.L. Lindström, B.G. Svensson, *Mat. Res. Sot, ymp. Proc., 59,45 (1986).*

[20] H.J. Stein, *Appl. Pyhs. Lett., 48, 1540(1986).*

[21] A. Brelot, *In Radiation Damage and Defects in Semiconductors,* J. Whitehouse (cd.), Institute of Physics, London, p. 191.1973.

[22] H.J. Stein, *Appl. Pyhs. Lett., 55, 870(1989).*

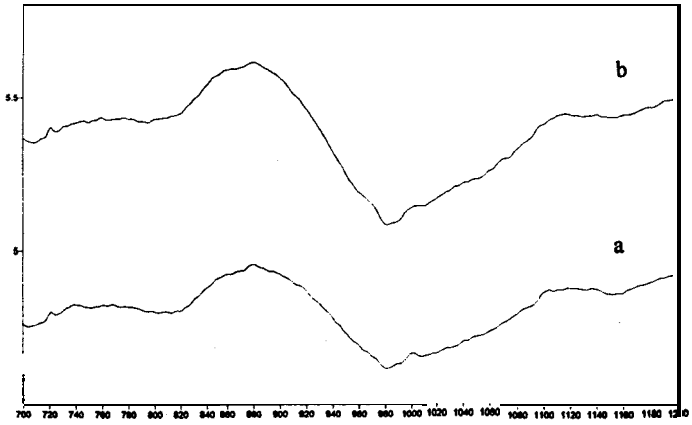


Figure 1. FTIR spectrum for n-type 6H-SiC substrate.
a. room temperature b. low temperature

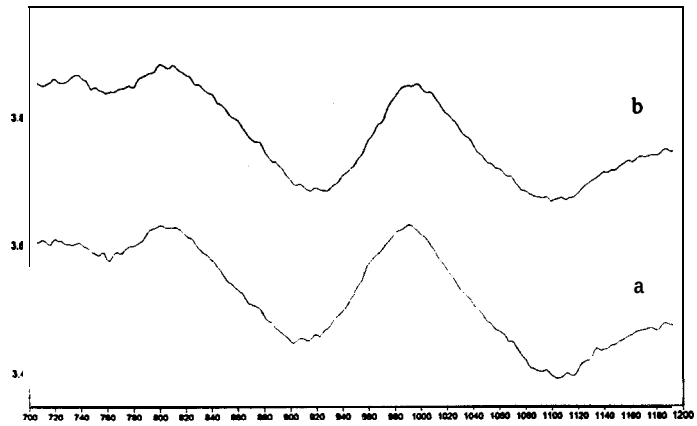


Figure *2.* FTIR spectrum for epitaxial layer 6H-SiC substrate.
a. room temperature b. low temperature

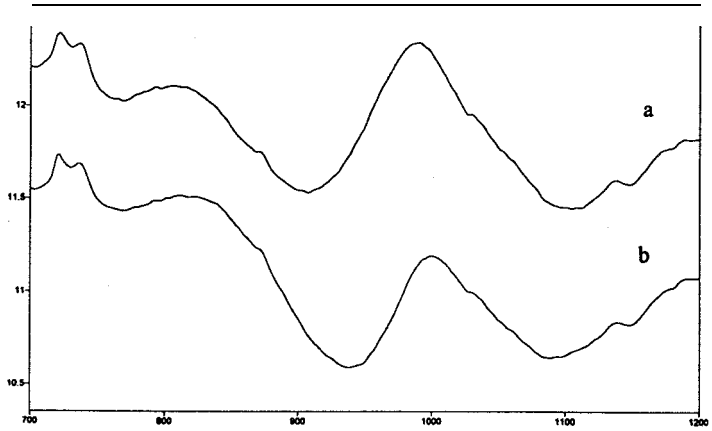


Figure *3.* FTIR spectrum for Ta pattern SiC epitaxial layer on 6H-SiC substrate.
a. room temperature b. low temperature

SYNTHESIS OF POLYMERS CONTAINING COVALENTLY BONDED NLO CHROMOPHORES

Xiaohua Deng[a], Mohan Sanghadasa[b], Connie Walton[c],
Benjamin B. Penn[d], Robert L.S. Amai[a], and Ronald D. Clark[a]

[a]Department of Physical Sciences, New Mexico Highlands University
Las Vegas, New Mexico 87701
[b]Department of Optical Physics, University of Alabama in Huntsville
Huntsville, Alabama
[c]Department of Chemistry, Grambling University
Grambling, Louisiana 71245
[d]Space Sciences Laboratory, NASA Marshall Space Flight Center
Huntsville, Alabama 35812

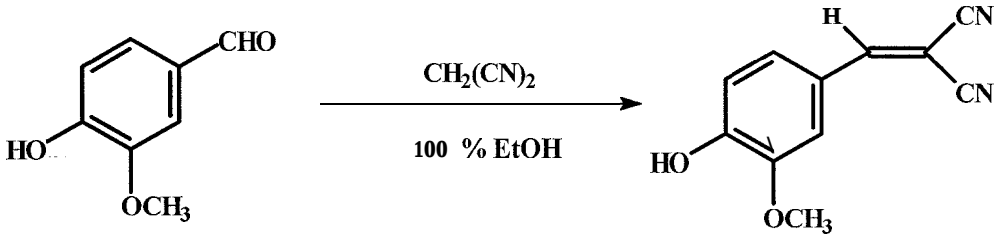
ABSTRACT

Polymers containing covalently bonded nonlinear optical (NLO) chromophores are expected to possess special properties such as greater stability, better mechanical processing, and easier film formation than their non-polymeric equivalent. For the present work, polymethylmethacrylate (PMMA) was selected as the basic polymer unit on which to incorporate different NLO chromophores. The NLO components were variations of DIVA { [(2-methoxyphenyl methylidene]-propanedinitrile} which we prepared from vanillin derivatives and malononitrile. These were esterified with methacrylic acid and polymerized either directly or with methyl methacrylate to form homopolymers or copolymers respectively. Characterization of the polymers and NLO property studies are underway.

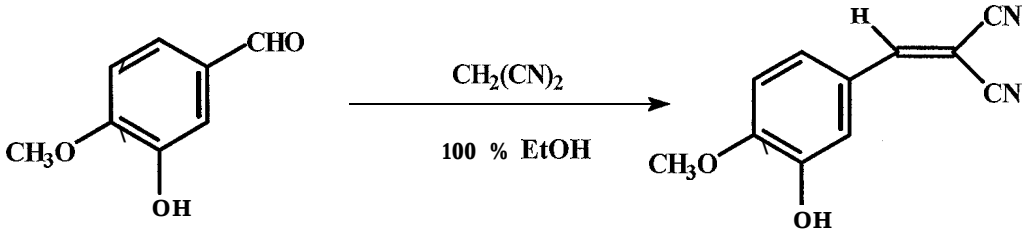
INTRODUCTION

Nonlinear optics is expected to play a significant role in photonics because of its potential applications in many areas of information and image processing technologies. Polymers containing NLO materials are of particular interest because of their ease of processing. Unfortunately, the number of polymer systems containing NLO materials connected to the polymer chain is relatively small.

We now wish to report the preparation of two methacrylate polymers containing derivatives of dicyanovinylanisole (DIVA) and the corresponding copolymers with methyl methacrylate and the attempted preparation of a third monomer. The specific NLO compounds and their method of preparation are shown on the next page. The procedure is basically that of Silver, Frandsen, Kelley, and Holmes[1] who prepared DIVA by treating o-anisaldehyde with malononitrile in ethanol.

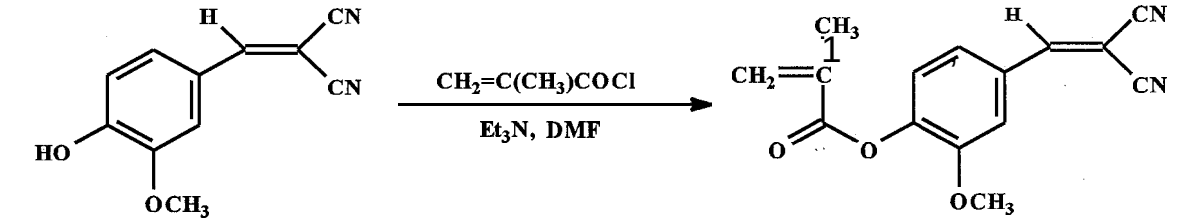


1

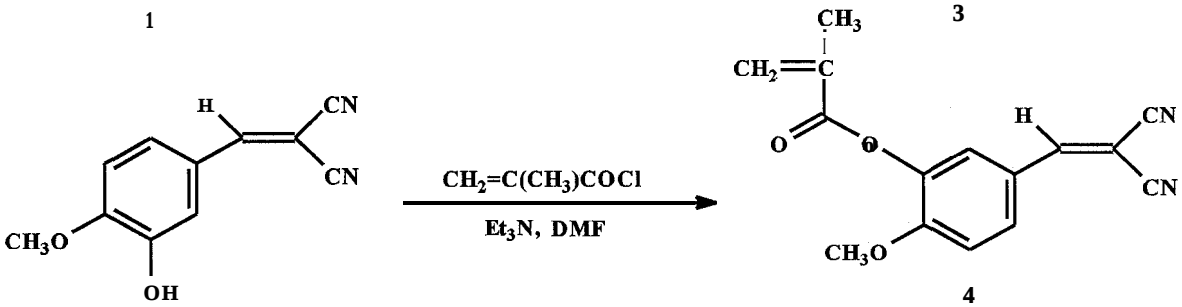


2

Compounds **1** and 2 were then treated with methacryloyl chloride to produce the methacrylate ester of the monomers.

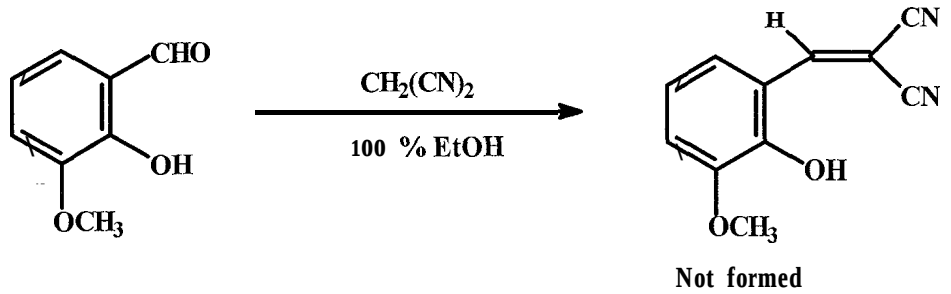


1 3



2 4

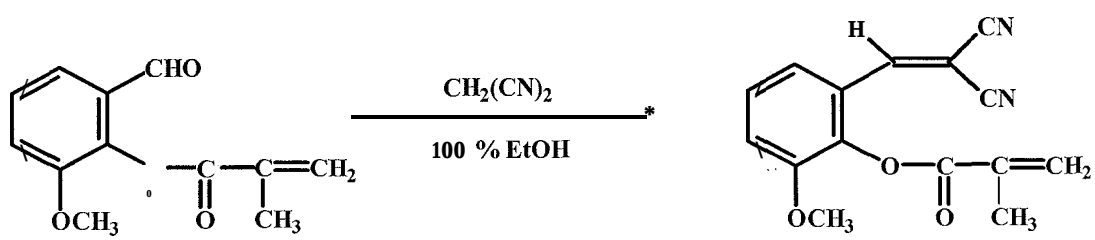
The condensation of 2-hydroxy-3-methoxybenzaldehyde with malononitrile was also attempted, but the product was insoluble, and was not identified.



Not formed

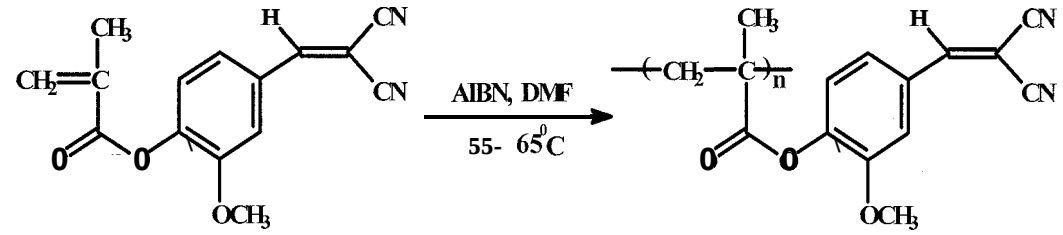
The methacrylate monomer, 6, of this material was finally prepared by reversing the order of synthesis.

5

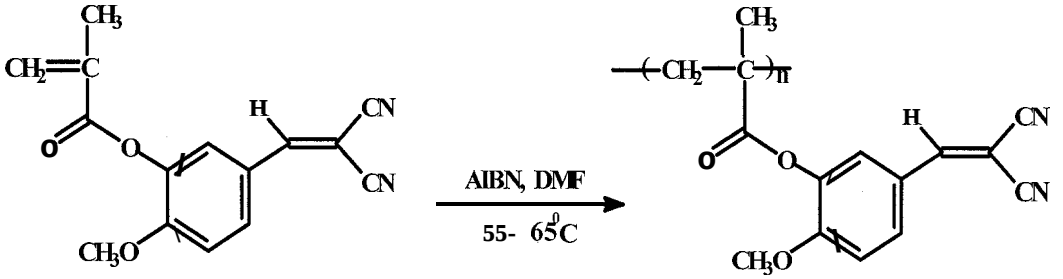


6

The polymers were prepared by polymerizing a dimethylformamide solution of the monomer with AIBN (2,2'-azobisisobutry onitrile) at 55-65°. The resulting polymers were light brown or yellow powders that could be dissolved in dimethylformamide and coated on a glass microscope slide.

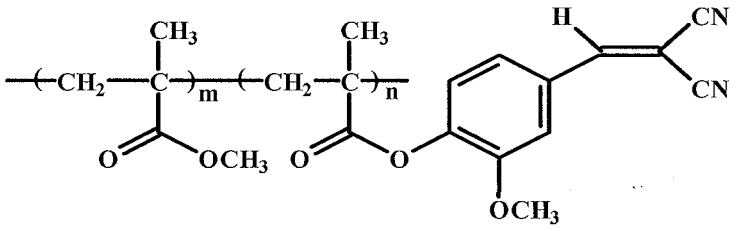


7

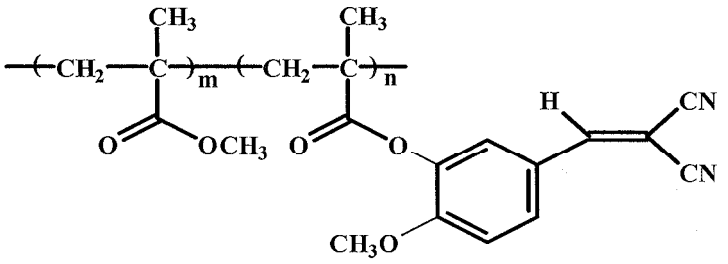


8

Copolymers of the two methacryate esters (3 and 4) with methyl methacrylate were prepared using a ratio of 1:1 for the two monomers. The physical properties of the resulting polymer were significantly improved over the homopolymers.



9



10

The corresponding polymers from 6 could not be prepared because the monomer was unreactive under the conditions used here.

Future work will involve a more detailed analysis of the polymers, modification of the copolymer ratios to optimize physical properties, and attempts to prepare poled films and thin fibers of the polymers.

EXPERIMENTAL

[(4-Hydroxy-3-methoxyphenyl-methylidene]-propanedinitrile, 1, and [(3-hydroxy-4-methoxyphenyl)-methylidene]-propanedinitrile, 2.- A 1:1 molar mixture of vanillin (3-methoxy-4-hydroxybenzaldehyde), or isovanillin (3 -hydroxy-4-methoxybenzaldehyde) and malononitrile in 100°/0 ethanol was refluxed in a hot water bath for 9-12 hours for **1,** and 3-4 hours for 2. The reaction mixture was then poured into an ice-water mixture and allowed to stand overnight. The solid product was collected by vacuum filtration and recrystallized. Product **1** was obtained as bright yellow crystals (mp 134-13 5°) in 85°/0 yield from a 10:1 mixture of benzene and ethanol. Product 2 was obtained as yellow crystals (mp 249-250°) in 90°/0 yield from a 6:1 mixture of n-propanol and pyridine. For compound 1: ^{1}H-NMR($CDCl_3$): δ 3.98 (-OCH_3), 6.44 (-OH), 7.03 (5-H), 7.31 (6-H), 7.64 (2-H), 7.71 [H-C=C($CN)_2$]. ^{13}C[^{1}H]-NMR($CDCl_3$): 656.2.78.0110.4, 113.6, 114.4, 115.2, 123.9, 128.9, 147.0, 152.1, 159.4. For compound 2: ^{1}H-NMR(pyridine-D_5): δ 3.81 (-OCH_3), 5.02 (-OH), 6.97 (5-H), 7.47 (6-H), 7.96 (2-H), 8.08 [H-C=C($CN)_2$]. ^{13}C[^{1}H]-NMR(pyridine-D_5): 656.0, 77.7, 112,1114.5, 115.6, 116.7, 125.4, 126.3, 148.8, 154.6, 160.5.

[(4-methacryloxy-3-methoxyphenyl)-methylidene]-propanedinitrile, 3, and [(3-methacryloxy-4-methoxyphenyl)-methylidine]-propanedinitrile, 4.- A mixture of compound 1 or 2, freshly distilled triethylamine (1 0-20 mole per cent excess),hydroquinone inhibitor 0.0 1mole per cent), and an organic solvent (Tetrahydrofuran for compound 3 and dimethylformamide for compound 4) was cooled to 0° in an ice-water bath with stirring under nitrogen. A solution of freshly distilled methacryloyl choride (10-20 mole per cent excess) in 10-20 ml solvent was then added dropwise. The mixture gradually turned from yellow to pale yellow, and a precipitate appeared. The reaction mixture was stirred at 0° for two hours and then at room temperature overnight. The precipitate was collected by vacuum filtration and washed twice with fresh solvent. The initial filtrate and washings were poured onto ice water and allowed to stand overnight to obtain additional product. The solid was recrystallized from a mixture of acetone and hexane togive the product in 73% yield. Compound 3 was a yellow powder (mp1 09-11 O°). Compound 4 was a cream yellow powder (mp 131-1320). For compound 3: ^{1}H-NMR($CDCl_3$): δ 2.06 (-CH_3), 3.88 (-OCH_3), 5.82 and 6.39 (CH_2=C, Z and E isomer), 7.22 (5-H), 7.38 (2-H), 7.70 (6-H), 7.74 [H-C=C($CN)_2$]. ^{13}C[^{1}H]-NMR($CDCl_3$): 618.3,56.1,82.2, 112.5, 112.7, 113.6, 124.0,125.6,128.3, 129.4, 134.9,145.2, 152.1, 159.0,164.6. UV-VIS($CDCl_3$): λ_{max} 241 nm,314 nm,355nm. FT-IR: v (cm^{-1}) 2976, 2228, 1750, 1753, 1509, 1322, 1268, 1101, 1027, 943, 869, 633. For compound 4: ^{1}H-NMR($CDCl_3$): δ 2.07 (-CH_3), 3.93 (-OCH_3), 5.81 and 6.38 (CH_2=C, Z and E isomer), 7.10 (5-H), 7.63 (2-H), 7.72 (6-H), 7.83 [H-C=C($CN)_2$]. ^{13}C[^{1}H]-NMR($CDCl_3$): 618.3,56.4,79.7, 112.6, 112.9, 114.1, 123.9, 125.1, 128.2, 131.5, 134.9, 140.5, 156.7, 158.1, 164.8. UV-VIS($CDCl_3$): λ_{max} 247 nm, 352 nm. FT-IR: v (cm-') 3391,2956,2223,1741, 11582,1508,11439,1283, 1124, 1010,946,823,614,561.

2-methacryloxy-3-methoxybenzaldehyde, 5.- A mixture of 0.01 mole of o-vanillin, 0.012 mole freshly distilled triethylamine and 20 ml N,N-dimethylformamide was cooled to 0° in an ice-water bath with stirring under nitrogen. A solution of 0.012 mole of freshly distilled methacryloyl chloride in 5-10 ml dimethylformamide was then added dropwise. The reaction mixture gradually changed color from green to pale white and a precipitate formed. The mixture was stirred at 0° for 1-2 hours and then at room temperature for several more hours. The mixture was poured into ice-water and allowed to stand overnight. The precipitate was collected by vacuum filtration and recrystallized from acetone/hexane. Colorless needles (mp 70-71°) were obtained in 52% yield. ^{1}H-NMR($CDCl_3$): δ 2.11 (-CH_3), 3.87 (-OCH_3), 5.84 and 6.43 (CH_2=),

7.23 (5-H), 7.36 (4-H), 7.50 (6-H), 10.16 (-CHO). $^{13}C[^{1}H]$-NMR($CDCl_3$): 618.4,56.3, 117.8, 120.5, 126.7, 128.1, 129.4, 134.9, 142.2, 151.8, 165.0, 188.6.

[(2-Methacryloxy-3-methoxyphenyl)-methylidene]-propanedinitrile, 6.- A mixture of compound 5 and malononitrile in a 1:1 molar ration in 100°/0 ethanol was refluxed for several hours. The reaction mixture was then poured into ice-water and allowed to stand overnight. The precipitate was collected by vacuum filtration and recrystallized from a mixture of acetone and hexane. The yield was 79°/0. ^{1}H-NMR($CDCl_3$): δ 2.09 (-CH_3), 3.86 (-OCH_3), 5.88 and 6.43 (CH_2=C, Z & E isomers), 7.21 (4-H), 7.36 (5-H), 7.85 (6-H), 7.88 [H-c=c$(CN)_2$]. $^{13}C[^{1}H]$-NMR($CDCl_3$): 818.3,56.3,84.9, 112.1, 113.6, 117.9, 125.2, 127.2, 129.0, 134.5, 140.3, 151.7, 153.3, 164.5.

Homopolymers of 3 and 4.- A solution of 3.34 g of monomer 3 or 4 and 0.1035 g AIBN (2,2'-azobisisobutyronitrile, 5 mole per cent) in 25 ml dimethylformamide was purged with nitrogen for 15 minutes at room temperature to remove dissolved oxygen. The solution was stirred under nitrogen at 55-69 in a water bath for 24 hours and then at 90° for 3-5 hours. The reaction mixture was poured into methanol and allowed to stand overnight. The precipitate was separated by centrifugation, redissolved in dimethylformamide and reprecipitated by pouring into stirred methanol. The precipated was isolated by centrifugation, washed with distilled water and dried in an oven at 90° until the weight was constant. Polymer 7 was a cream-yellow powder (t_g 405°) that swelled rather than dissolved in chloroform. Polymer 8 was an off-white powder (t_g 405°) that swelled rather than dissolved in chloroform. Both products were soluble in dimethylformamide.

Copolymers of 3 and 4 with methylmethacrylate.- A solution of 3.016 g of either monomer 3 or 4, 1.138 g methyl methacrylate and AIBN (1 mole per cent) in 225 ml dimethylformamide was purged with nitrogen for 15 minutes at room temperature to remove dissolved oxygen. The solution was stirred under nitrogen at 55-65° in a water bath for 24 hours and then at 90° for 3-5 hours. The reaction mixture was poured into methanol and allowed to stand overnight. The precipitate was separated by centrifugation, redissolved in dimethylformamide, and reprecipitated by pouring into stirred methanol. The precipitate was isolated by centrifugation, washed with distilled water and dried in an oven at 90° until constant weight.

ACKNOWLEDGMENTS

This work has been supported in part by NASA-IRA grant number NAG5-5 121 and NASA grant NAG8-1288. This work was taken in part from the Master of Science thesis of Xiaohua Deng, New Mexico Highlands University, 1997.

REFERENCES

1. R.F. Silver, K. Ann, P.D. Frandsen, S.J. Kelley, and J.L. Holmes, *Canadian J. Chem., 1967, 45,* 1001.
2. D.R. Robello, *J. Polym. Sci: Part A: Polym. Chem. Ed., 1990,28,* 1.
3. D.R. Robello, P.T. Dao, J.S. Schildkraut, M. Scozzafava, E.J. Urankar, and C.S. Willand, *Chem. Mater., 1995,7,284.*
4. W. Kohler, D.R. Robello, C.S. Willand, and D.J. Williams, Macromolecules, 1991,24,4589.

Analysis of Phase Separation in Czochralski Grown Single Crystal Ilmenite

R. Wilkins, Kirk St. A. Powell, Kieron R. Loregnard and Sy-Chyi Lin
Center for Applied Radiation Research
Prairie View A&M University
Prairie View, TX 77446

Jayakumar Muthusami, Feng Zhou and R. K. Pandey*
Center for Electronic Materials, Devices and Systems
Texas A&M University
College Station, TX 77843

Geoff Brown and M. E. Hawley
Center for Material Science, MS K765
Los Alamos National Laboratory
Los Alamos, NM 87545

INTRODUCTION

Ilmenite ($FeTiO_3$) is a wide bandgap semiconductor with an energy gap of 2.58 eV. Ilmenite has properties suited for radiation tolerant applications, as well as a variety of other electronic applications. These properties and applications have been discussed elsewhere [1].

Single crystal ilmenite has been grown from the melt using the Czochralski method. Growth conditions have a profound effect on the microstructure of the samples. Here we present data from a variety of analytical techniques which indicate that some grown crystals exhibit distinct phase separation during growth. This phase separation is apparent for both post-growth annealed and unannealed samples.

Under optical microscopy, there appear two distinct areas forming a matrix with an array of dots on order of 5 μm diameter. While appearing bright in the optical micrograph, atomic force microscope (AFM) shows the dots to be shallow pits on the surface. Magnetic force microscope (MFM) shows the dots to be magnetic. Phase identification via electron microprobe analysis (EMPA) indicates two major phases in the unannealed samples and four in the annealed samples, where the dots appear to be almost pure iron. This is consistent with micrographs taken with a scanning probe microscope used in the magnetic force mode. Samples that do not exhibit the phase separation have little or no discernible magnetic structure detectable by the MFM.

EXPERIMENTAL PROCEDURE

The Czochralski grown ilmenite crystal samples have been characterized using a variety of techniques to quantify the crystal quality. It is hoped that this analysis will provide the crystal growers with valuable information that will allow them to optimize their crystal growth and post-growth processing (e.g. cutting and polishing). For example, our

previous work using atomic force microscopy on similar ilmenite samples [1] indicated that surface polishing of cut crystals was producing micro-scratches on the surface. These scratches were on a scale that would have been difficult to detect using optical or electron microscopy.

Nevertheless, optical microscopy gives a simple and quick idea of the relatively large scale features on the crystal surface. An example of an optical micrograph of the crystals of interest is shown in Figure 1. These samples show a striking pattern of dots throughout the surface. Obviously, this inhomogeneity is undesirable and our work sought to establish the nature of the pattern and its possible causes.

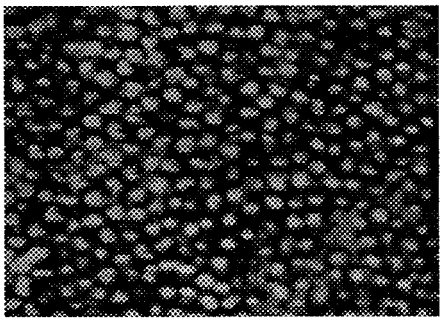


Figure 1: Optical micrograph of a phase separated ilmenite sample. The image is approximately 175μm across.

Crystal Growth: The Czochralski single crystal growth experimental set up is shown in the following figure [2]:

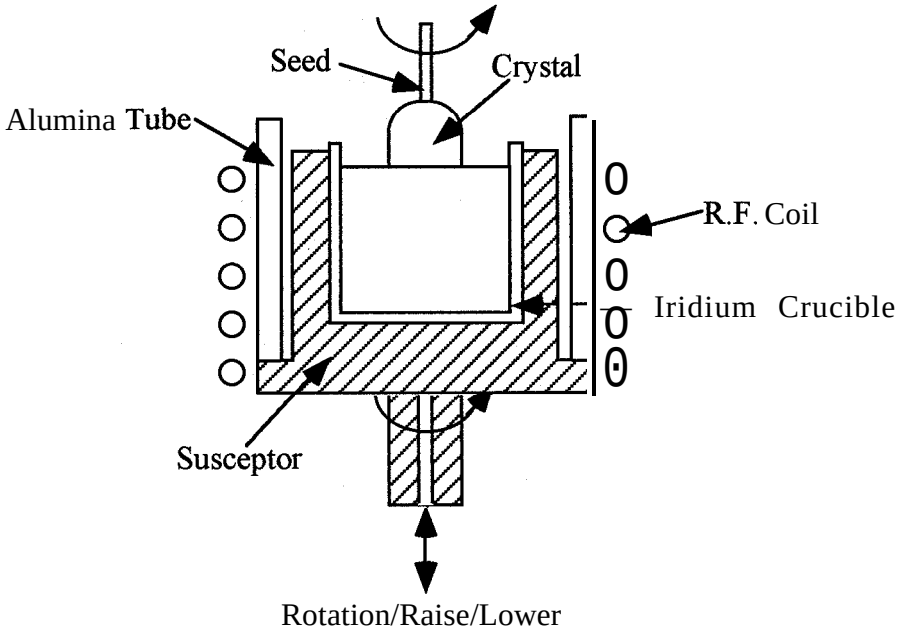


Figure 2: A sketch of the Czochralski single crystal growth.

The experimental set up is comprised of a water cooled chamber with a heating element in it. Our experiments used inductive heating to heat the inner crucible in which the charge ilmenite is placed. A graphite susceptor is used between the radio frequency coils and the crucible. The graphite susceptor provides a dual role as an efficient coupler of induced voltage and eddy current for heating and to protect the expensive iridium crucible. Iridium is used to contain the charge as ilmenite reacts with most other metals. The thermal energy from the graphite and iridium crucibles is transferred by conduction to melt the ilmenite charge. A seed rod that holds a previously grown and oriented single crystal seed of ilmenite is introduced from the top. The seed is made to contact the ilmenite melt and then slowly withdrawn from the melt. It is essential to maintain a proper temperature at the crystal-melt interface and pull rate for the growth process to continue. The seed rod is rotated to maintain uniform distribution of the charge carriers in the grown crystal. The chamber is filled with high purity nitrogen to provide a reducing atmosphere during growth. The important parameters that influence the growth of single crystal ilmenite are temperature at the crystal-melt interface, pull rate and rotation rate of the seed rod, and the power level.

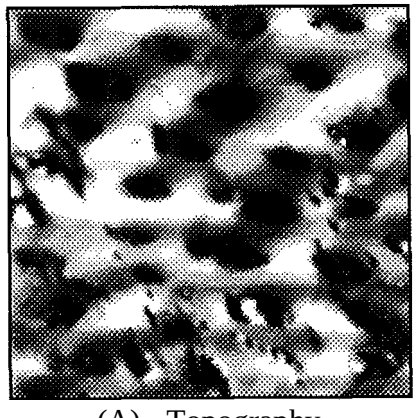
In the annealing process, the sample temperature increases from room temperature to 1,100 'C at a ramp-up rate of 1.5 °C/min. The temperature is then held at 1,100 °C for 36 hours before it cools to room temperature at a cooling rate of 1.0 °C/min.

Atomic Force/Magnetic Force Microscope: Atomic force microscopy (AFM) is a scanning probe technique in which surface topography is mapped by sensing the van der Waals interaction between nanometer sized surface features and a nanometer sized probe tip [3]. Using semiconductor device manufacturing techniques, a pyramidal probe tip, typically with < 10 nm radius, is formed on the end of a silicon cantilever with typical dimensions of 10 nm x 30 nm x 200 nm. This cantilever is then oscillated near its resonant frequency (70 kHz -300 kHz) and a laser-photodiode assembly is used to detect changes in the resonance characteristics as the tip interacts with the sample [4]. By feeding back to some chosen resonance parameter, the tip can be scanned over the surface to map out its topography. Because it is a scanned probe technique, AFM routinely provides 3-dimensional, quantitative surface data without extensive data processing or correction. The tip's small radius of curvature and the extremely short range van der Waals forces combine to give this technique sub-nanometer resolution parallel to the surface and sub-angstrom resolution perpendicular to the surface.

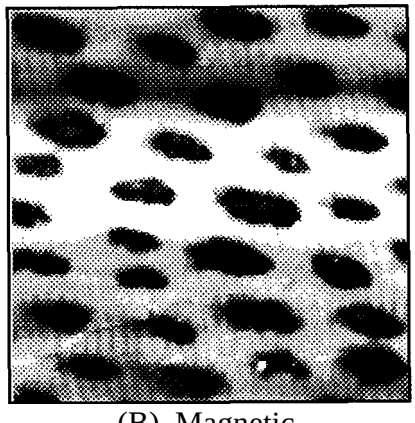
Magnetic force microscopy (MFM) is an extension of AFM in which an AFM tip with a layer of magnetic material is scanned over the surface, allowing it to sense magnetic interactions through their effect on the cantilever resonance characteristics [5,6]. Magnetic structure variations can be sensed independently of topographic influence by lifting the tip away from the surface during scanning. Very near to the surface, < 1 nm, van der Waals forces dominate and the feedback loop maps topographic variations. If the tip is lifted several tens of nanometers above the surface, changes in the resonance are primarily due to magnetic interactions. By interleaving topographic AFM scans and magnetic MFM scans as the raster image is built up, a one-to-one correspondence can be

made between surface physical and magnetic structure with little cross talk between the two signals. Since the tip must be lifted away from the surface and has a slightly larger radius due to the magnetic layer, the optimal resolution of MFM is about 20 nm and becomes worse as the tip is lifted higher.

Electron Microprobe Analysis (EMPA): This technique is based on scanning electron microscopy at moderate resolution, where the machine is optimized for the detection of characteristic x-rays produced by the electron beam interacting with the sample [7]. The resulting x-ray spectrum gives a map of the chemical composition of the sample. If the equipment is used in conjunction with a standard sample of known composition, the X-ray spectrum can be quantified to give stoichiometric information about the sample. We have analyzed both annealed and unannealed ilmenite samples; results are shown in Figure 3.



(A). Topography



(B). Magnetic

Figure 3: 50 μm x 50 μm scanning probe microscope imagines. (A) AFM topographic imagine. The gray scale presents a 100 nm range in the topography (black to white). (B) MFM imagine showing the corresponding magnetic structure.

RESULTS AND DISCUSSION

Figure 1 shows an optical micrograph of one of the ilmenite crystals. The most striking feature of the image is the pattern of bright dots on the surface. The array of dots is immediately suggestive of a phase separation in the sample. We found similar pattern on both annealed and unannealed samples, indicating that the phase separation associated with the dots was not the result of the annealing process. However, as we discuss below, the annealing process did cause additional phase separation in the samples. There is a range of diameters, but the typical value appears to be 5-10 μm , The overall structure of the dots is unclear from the optical micrograph because it is possible that the phase within the dots may have relatively high reflectivity. To investigate this structure, we have used scanning probe microscopy.

Figure 3 shows a typical image taken with a scanning probe microscope used in the atomic force mode (giving topographic information) and the magnetic force mode (giving information on magnetic structure) on an annealed ilmenite sample. The data were obtained simultaneously as described above. The topographic data reveals that the bright dots in the optical micrographs are indentations on the surface with depths of about 100 nm. The AFM also revealed other defects on the surface. On annealed sample, the indentations are surrounded by material with a parallel streak pattern which has been characteristic of other ilmenite samples [1].

Given that ilmenite is a ferrous compound, the possibility exists that the material may have a discernible magnetic structure. We applied the MFM to these ilmenite samples to elucidate any magnetic structure associated with the phase separation. The magnetic image in Figure 3 show clearly the dot pattern as is shown in the topographic image. Close inspection reveals that the MFM is detecting magnetic structure near the surface. Note that there is not a one to one correspondence between dot features in the topographic and magnetic images, showing that the MFM is indeed detecting magnetic structure. There is also some indication in our data that the MFM is detecting sub-surface magnetic features. Based on these magnetic images, we predicted the dots to be an iron rich phase of the material. To verify the prediction and identify the chemical composition of the phases, we performed electron microprobe analysis on these samples.

Figure 4 shows examples of the results from EMPA. Figure 4A is the micrograph for the unannealed sample and Figure 4B is the micrograph for the annealed sample. The analysis of both the annealed and unannealed samples indicate that the dots are very iron rich, in agreement with our prediction from the MFM. The micrographs also indicate that the annealing process caused further phase separation in the samples as compared to the unannealed ilmenite. The unannealed sample shows two distinct phases with an iron rich dot pattern in a background made up mostly of TiO_2. There is evidence of inter-diffusion upon annealing and additional phase separation. The dot pattern is still evident in the annealed sample, but the dots have been distorted and some have coalesced together (there was also some indication of this in the AFM/MFM images). It can also be concluded from the compositions inside the dots that iron atoms migrate into the dots from the background and/or the titanium atoms migrate out of the dots during the annealing process. The migration results in an increase in the Fe-Ti ratio from 36 to 113. In the

background, a distinct ilmenite phase is evident. Two other distinct phases are TiO_2 and a mixture of TiO_2 and Fe_2O_3.

At this time, the cause of the separation is not clear. More investigation will be conducted to study the phase separation. We propose here the possible reasons of the phase separation as following:

(1) Initial growth contained more pure ilmenite and then very little iron separation took place. This could be attributed to the use of alumina rod as the seed holder which did not provide any heat sinking. It takes a longer time for grown crystal to cool down and could attribute to the phase separation. At present, tungsten rod is used as the seed holder which would provide better heat sink to the grown crystal. The grown crystals with this set up are yet to be analyzed.

(2) Grown crystal of ilmenite is black in color and absorbs heat radiation from the surrounding crucible wall and the melt surface. This rises the temperature of the grown crystal to the region where phase separation can take place. Using resistive heating instead of radio frequency heating could minimize this problem. Also, the use of the tungsten rod as the seed rod holder could help to overcome this issue.

(3) The grown ilmenite crystal was in a high purity nitrogen ambient atmosphere (reducing). The pressure was held static in the chamber. Since the graphite crucible used as the susceptor absorbs oxygen from the high purity nitrogen, the partial pressure of oxygen in the chamber becomes very low. This could lead to the removal of oxygen from the ilmenite and not having the required stoichiometric ratio of oxygen. In the recent crystal growth experiment, we used a free flow of pre-purified nitrogen. This has reduced the iron precipitation problem. Ideally crystal growth of ilmenite requires closed control of the oxygen ratio in the grown chamber.

ACKNOWLEDGMENTS

This work is partially funded by NASA CARR and the State of Texas through the Advanced Technology Program. The electron microprobe analysis was conducted at the Texas Superconductivity Center of the University of Houston. We thank Ulh Woods and Shojah S. Ardalan for their help formatting the images.

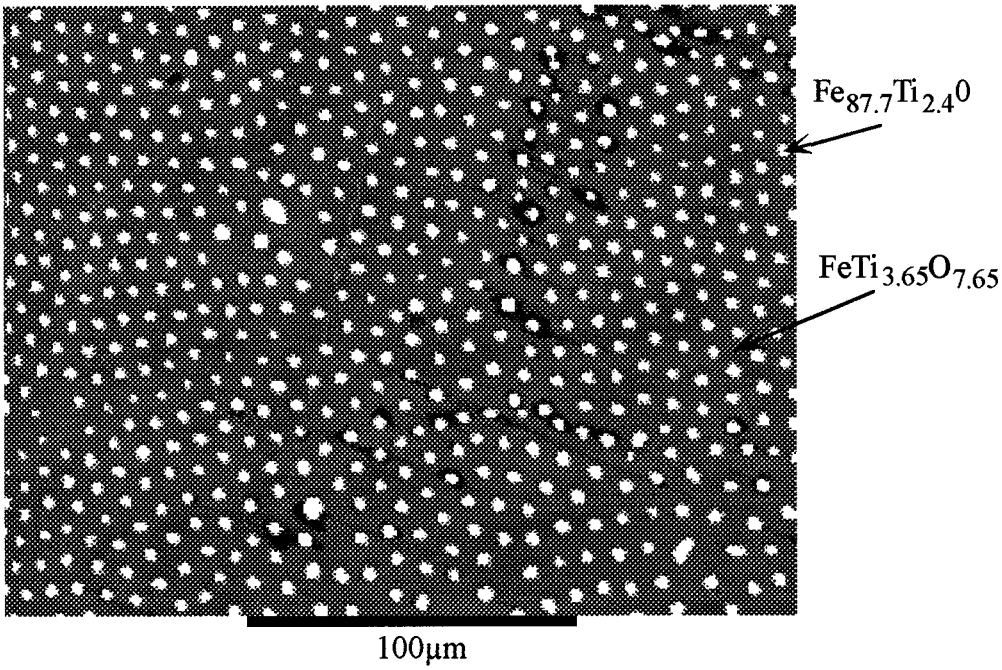
* Current address: Dept. of Electrical & Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487

REFERENCES

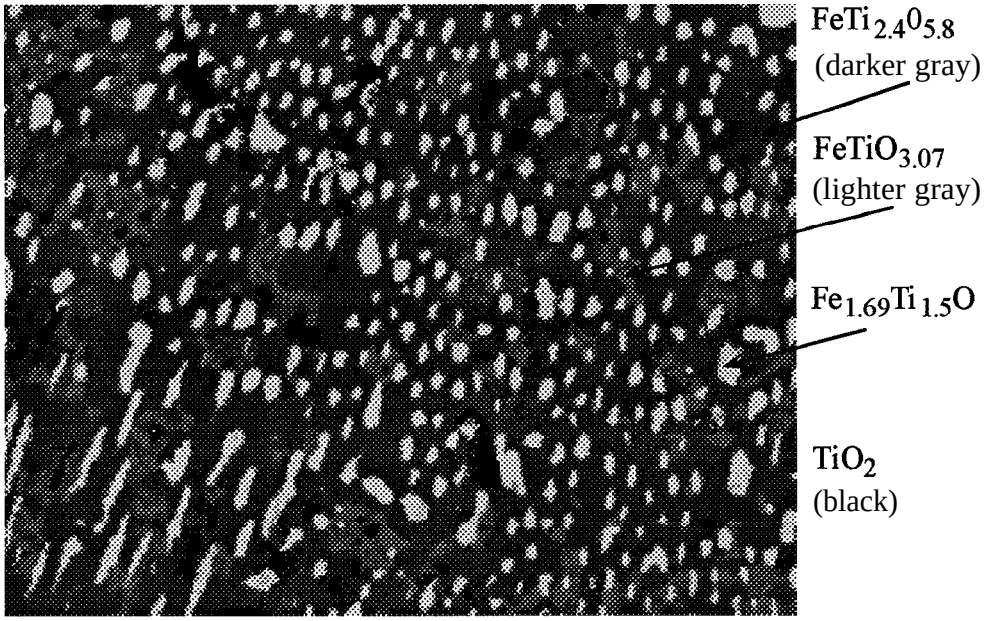
[1] R. Wilkins and Kirk St. A Powell, ***NASA URC Technical Conference Proceedings,*** University of New Mexico Press, Albuquerque (1997) pg. 797

[2] B.R. Pamplin, "Crystal Growth" 2nd Ed., Pergamon Press (1980).

[3] R. Wiesendanger, "Scanning Probe Microscopy and Spectroscopy", Cambridge University Press, Cambridge (1994).
[4] This section and the next one on magnetic force microscopy describe techniques implemented in the family of scanning probe microscopes made by Digital Instruments, Inc., Santa Barbara, CA
[5] K. Babcock, M. Dugas, S. Manalis, and V. Elings, Proceedings of the Materials Research Society, 355, p. 311 (1995).
[6] P. Grutter, MSA Bulletin, 24, p. 416 (1994).
[7] P. E. J. Flewitt and R. K. Wild, "Physical Methods for Materials Characterization", Institute of Physics, London (1994) pg. 266.



A. Unannealed sample.



B. Annealed sample

Figure 4: Electron microprobe analysis micrographs of both annealed (A) and unannealed (B) samples. The shades of gray represent different phases of the material. The local compositions of the phases are indicated. The scale shown with the unannealed sample is the same for the annealed sample.

BULK CRYSTAL GROWTH OF NONLINEAR OPTICAL ORGANIC MATERIALS USING INVERTED VERTICAL GRADIENT FREEZE METHOD

J. Choi, Magda Cruz, R. Metzl, W.S. Wang and M.D. Aggarwal
Department of Physics
Alabama A&M University, Normal, AL 35762

Benjamin G. Penn and Donald O. Frazier
NASA/Marshall Space Flight Center, Huntsville, AL 35812

ABSTRACT

A new process for producing large bulk single crystals of benzil (C_6H_5 COCO C_6H_5) is reported in this paper. Good quality crystals have been successfully grown using this approach to crystal growth. This method seems to be very promising for other thermally stable NLO organic materials also. The entire contents vycor crucible 1.5 inch in diameter and 2 inch deep was converted to single crystal. Purity of the starting growth material is also an important factor in the final quality of the grown crystals. The entire crystal can be very easily taken out of the crucible by simple maneuvering. Initial characterization of the grown crystals indicated that the crystals are as good as other crystals grown by conventional Bridgman Stockbarger technique.

INTRODUCTION

In an effort to grow organic nonlinear optical (NLO) materials (Nalwa, 1997, Gupta, 1997) for optical processing applications, we have successfully designed and fabricated a Czochralski crystal growth system in which one can view the crystal melt interface at various angles as and when desired (Aggarwal, 1992). To grow large size crystals of these NLO materials we tried to grow benzil crystals using Kyropoulos technique and by chance the entire top layer of the crucible solidified. Then it was left over the weekend. The next day we observed that a good quality single crystal was growing vertically downwards. After several runs, we were able to repeat the experiment for a particular programmed rate of lowering the temperature, and hence a new technique was developed.

A relatively well known approach to crystal growth of various materials is gradient freeze method (Hurle, 1994). In this method, normally a Bridgman type crucible is positioned in the furnace and a monotonically increasing temperature profile is imposed along the crucible. After melting the material and attaining the thermal steady state, temperature is ramped down slowly to solidify the melt. In the present method, we could convert the entire Czochralski type vycor crucible 1.5 inch diameter and 2 inch high filled with benzil into a single crystal. We might call this technique the Inverted Vertical Gradient Freeze (IVGF) method. This method combines several advantages of two common growth techniques, namely Kyropoulos and Czochralski methods. In the IVGF method, the hot zone has a characteristic axial temperature profile and as growth proceeds from the seed into the melt, the entire profile is shifted from the seed into the melt, the entire profile is shifted to lower temperatures. Unlike Czochralski method, no rotation and steep temperature gradients are required to induce crystallization. In addition, the crystal growth system has been fabricated using glass, so that the glass crucible containing the melt and the crystal melt interface can be clearly observed from any direction. The shape of the crystal melt interface is a function of the natural convection in the melt adjacent to the interface. In most cases, the crystal melt interface of the grown bulk benzil crystals maintains a slightly convex shape.

EXPERIMENTAL

We have successfully designed and fabricated Czochralski crystal growth system with automatic diameter control (Aggarwal, 1995) and a number of crystals have been grown using this technique. In the present IVGF method, we dipped a seed in the melt and temperature was slowly lowered to solidify the top surface. Then the temperature was lowered at a programmed rate to solidify the entire melt in the crucible. Although benzil is relatively an easy crystal to grow, this technique could prove useful for growing large crystals of other organic NLO materials. A Schematic diagram is shown in
Fig 1.

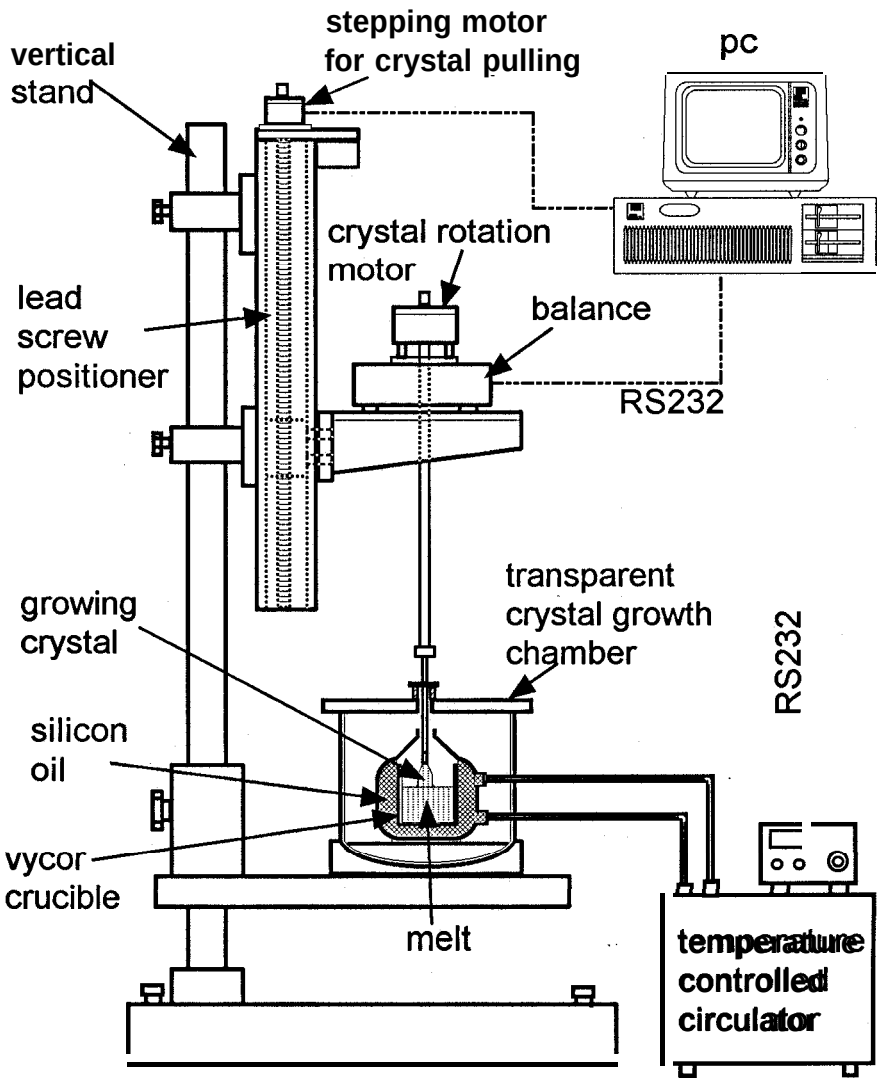


Fig. 1. Schematic diagram of automatic diameter controlled Czochralski puller for organic NLO materials.

The crystal growth chamber consists of a double walled oil circulating vessel designed and fabricated in our workshop. Its temperature is kept constant by circulating controlled temperature silicone oil using NESLAB high temperature circulating bath. This chamber is further seated in another glass vessel to avoid heat losses. This entire assembly is then used in conjunction with the Czochralski type pulling and rotation system shown in Fig. 1. In a typical experiment, a vycor crucible is filled with ultrapure benzil and melted. The temperature is equilibrated at 96 °C for 24 hours. The seed, mounted on the end of a glass rod, is then inserted in the melt surface. With proper adjustment of the vertical and horizontal gradients the entire surface is solidified. Growth is mediated by a steady decrease in the temperature of the system of 0.5 °C per day. The crystal begins to grow from the seed and spreads toward the crucible wall. The wall is reached by the growing crystal before it has grown more than 1-2 mm thick, so that growth after that is nearly uniform down into the remaining melt. As the crystal grows, the density difference between the liquid and solid cause the crystal to pull away from the crucible wall on one side. Impurities which remain in the melt cause the last solidified material to be deep yellow. After solidification is complete the crystal can be cooled at 0.5 to 1 °C per hour to room temperature. The bottom of the crucible can be removed by using a low speed diamond saw, and the crystal can be separated from the crucible wall by light manual pressure. The progress of the solid-melt interface can be observed in real time and is schematically shown in Fig. 2. The crystals grown using this inverted vertical gradient freeze method are shown in Fig. 3.

RESULTS AND CONCLUSIONS

Large size (1.5 inch diameter and 2 inch in length) Benzil crystals up to about 50 g have been successfully grown using an inverted vertical gradient freeze method. Crystals were characterized for their optical properties and found to have comparable values to the conventional crystals grown by Bridgman Stockbarger or Czochralski technique.

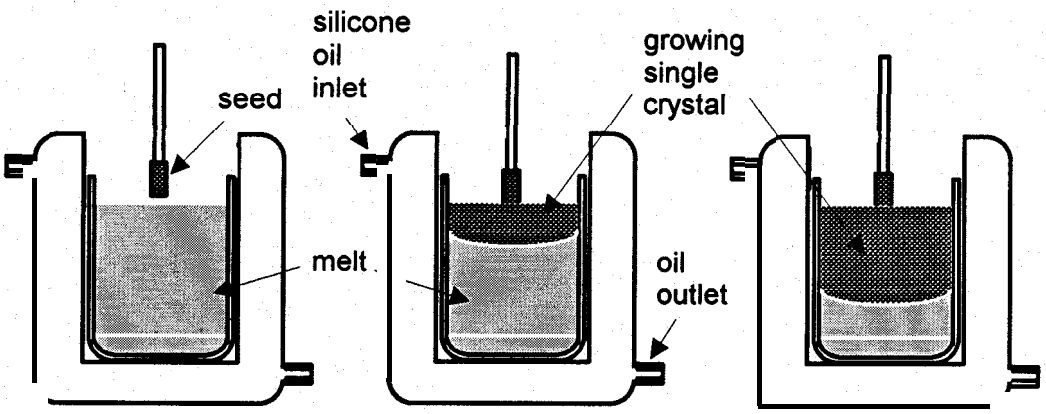


Fig. 2. Progress of the solid-melt interface in IVGF method.

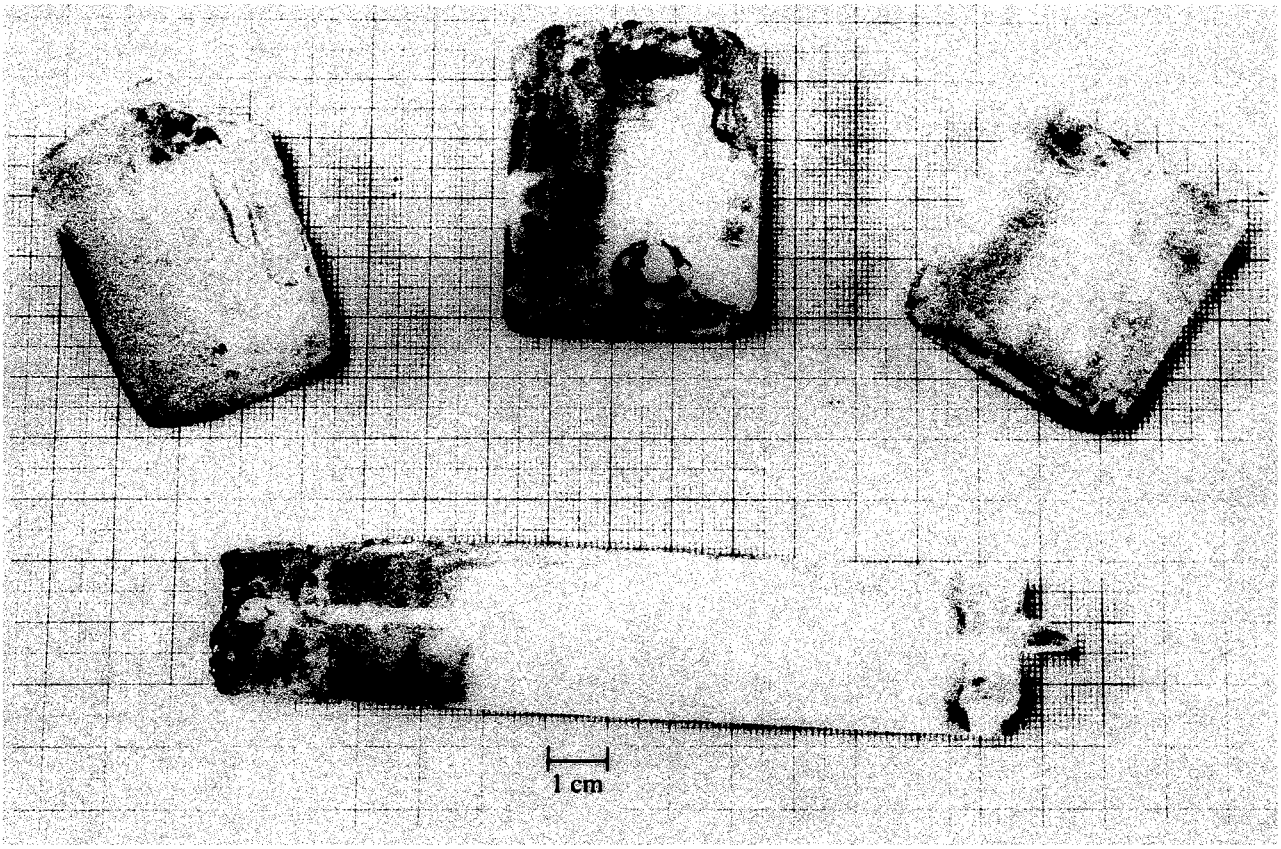


Fig. 3. The crystals grown using the inverted vertical gradient freeze method.

ACKNOWLEGEMENTS

This work is performed under NASA Alliance for Nonlinear Optics, NAG5-5 121, NASA grant NAG8-1390 and NSF grant HRD9353548. The authors wish to thank Dr Ronald Clark, Director for the NASA Alliance for Nonlinear Optics for his keen interest and encouragement during the progress of the work. The authors also thank Mr Jerry Johnson who helped us in the design and fabrication of our transparent growth system.

REFERENCES

Aggarwal, M.D. Metzl, R., Wang, W.S. and Choi, J. , Rev. Sci. Instrum. 66 (1995) 3939.

Aggarwal, M. D., Wang, W. S., Shields, A.W. Penn, B.G. and Frazier, D.O. Rev Sci. Instrum. 63 (1992) 1992.

Gupta, M.C. ,Ed., Handbook of Photonics, p. 155 (CRC Press, New York 1997).

Hurle, D.T.J. Ed. Handbook of Crystal growth, Vol 2a Basic Techniques, 1994 North Holland Publishing Co., Amsterdam).

Nalwa, H.S. and Miyata, S. Eds. Nonlinear Optics of Organic Molecules and Polymers (CRC Press New York) 1997 p. 89.

GROWTH OF BULK SINGLE CRYSTALS OF DICYANOVINYL-ANISOLE AND ITS DERIVATIVES FOR NONLINEAR OPTICAL APPLICATIONS

T. Gebre, J. Choi, W. S. Wang, R. Metzl and M.D. Aggarwal
Department of physics, Alabama A&M University, Normal, AL 35762

Melvin Romero and Ronald D. Clark
Department of Physical Science,
New Mexico Highlands University, Las Vegas, NM 87701

Benjamin G. Penn and Donald O. Frazier
NASA/ Marshall Space Flight Center, AL 38812

ABSTRACT

Bulk single crystals of a series of thermally stable nonlinear optical organic materials, Dicyanovinyl-anisole (DIVA) and their methoxy derivatives, have been successfully grown using the Bridgman-Stockbarger technique. The growth conditions are chosen to be temperature gradient of 5 to 10 °C/cm and lowering rate of 0.1 to 0.3 mm/h.. Single crystals of DIVA and its derivatives, of 8 x 8 x 50 mm^3 in size, have been grown while maintaining a flat solid-liquid growth interface.

INTRODUCTION

Organic materials exhibit extremely large nonlinear optical susceptibility compared to known inorganic materials. Organic molecular structure can be optimized in order to maximize nonlinear properties. Substituted-benzene derivatives are very promising organic materials for NLO applications (Grossman, 1989). One such compound, Dicyanovinyl-anisole or (2-methoxyphenyl)-methylene-propanedinitrile (DIVA) is thermally stable, and not excessively volatile. The microscopic second-order susceptibility (χ^2) of DIVA is comparable to that of nitroaniline-based molecules. Its melting point is 84.5 “C. Single crystal DIVA has the monoclinic point group symmetry 2 and space group $P2_1$ at room temperature. Vapor-phase and solution growth have been applied to grow single crystals of DIVA. However, vapor-phase growth can not be employed because of the low vapor

pressure of the DIVA crystals entailing prohibitive growth times. Solution growth method is also not applicable due to undesirable solvent/compound association occurs during growth. This formation of solvent inclusions can reduce the second-harmonic efficiency and other optical properties due to scattering. Mechanical properties are also reduced due to these solvent inclusions. In this paper, single crystals of DIVA and its derivatives are grown using the Bridgman-Stockbarger (BS) technique which is a melt growth method.

EXPERIMENTAL METHODS

The compounds of DIVA and its derivatives were synthesized and purified at New Mexico Highlands University for NLO applications. The molecular structure and melting point of these materials is shown in Fig. 1 and Table 1., respectively (Clark, 1995). In general, growth of single crystals of organic materials is more difficult than inorganic due to the tendency of supercooling and low thermal conductivity. A transparent BS growth chamber has been designed and fabricated to overcome the temperature limitation disadvantage of the previous set-up (Aggarwal, 1996). Two separate resistance heating elements are immersed in the growth chamber. Two types of silicone oils are used as heat transfer media. Dow Corning DS-704 diffusion pump oil is used as the cold-zone medium and Dow Corning fluid 200 is used in the hot zone. These oils form two separate zones without mixing with each other at any temperature. The circular heating elements are connected by copper wire to the power source. At the bottom of the chamber an optical window is mounted to provide access for the optical beam and other accessories. Growth chamber configuration is shown in Fig. 2. Two power control systems with PID control are used to maintain the hot and cold zone temperatures within ±0.1 ‘C. The growth chamber has been mounted on a vertical linear motion stage with a stepper motor and gear reducer. In this lowering system, minimum lowering rate of 0.01 mm/h can be obtained. The ampoule is fixed to the modified optical-mirror mount that permits it to be located at the center of the growth chamber. A fixed ampoule mechanism allows a simplified observation system because the reference point of the growth front is unchanged during

measurement of growth dynamics. A charge-coupled device (CCD) camera with telemicroscope lens is used to observe and record the growth process. The CCD camera images are recorded on a video recorder for future process analysis. Post processing will be conducted using a frame grabber and digital-signal processing software. A photograph of grown singles of DIVA is shown in Fig. 3.

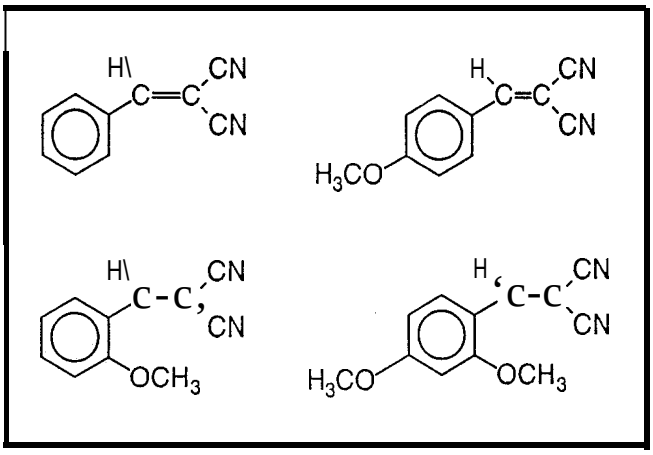


Fig. 1. The molecular structure of DIVA and its derivatives

Table 1. Melting point of DIVA and its derivatives.

Compound	Melting point (°C)
2-dicyanovinylanisole	84.5
phenylmethylene-propane dinitrile	85.0
p-methoxyphenylmethylene-propanedinitrile	115.0
2,4 dimethoxyphenyl methylene-propanedinitrile	143.0

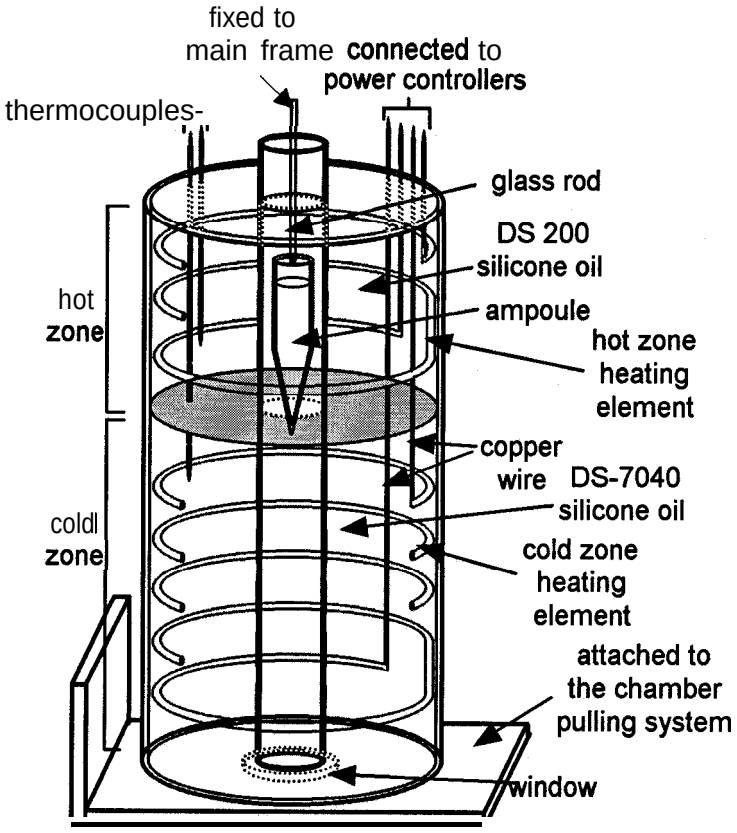


Fig. 2. Schematic diagram of the growth chamber configuration.

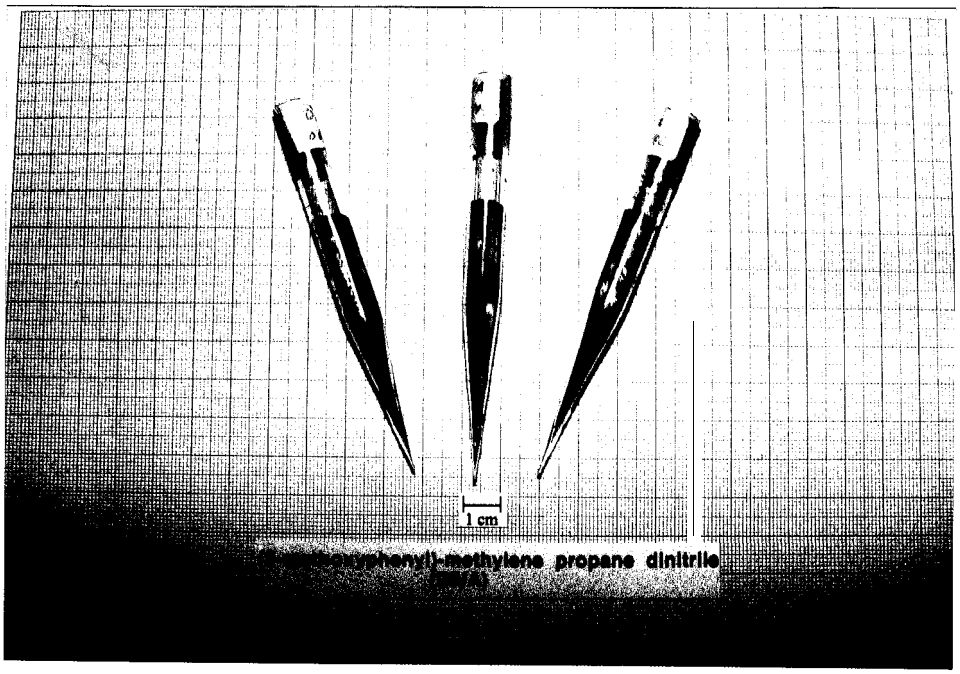


Fig. 3. A photograph of grown single crystals of DIVA.

SECOND-HARMONIC GENERATION OF DIVA CRYSTALS

Second-harmonic generation using an Nd:YAG laser has been conducted for grown DIVA and 2,4 methoxy DIVA single crystals. A Q-switched Nd:YAG laser (Continuum) with repetition rate 10 Hz has been setup in our laboratory. The peak power of each pulse (7 ns pulse width) is 3.5 MW/cm^2. A polished DIVA crystal is mounted on a goniometer which is set on the x-y translator. The fundamental wave at 1064 nm is separated by a high power harmonic separator. 15% of the fundamental wave is still mixed in the second-harmonic wave. Further purification of the second harmonic wave at532 nm is obtained by passing it through a prism. The fundamental and harmonic waves are separated by the prism. Since the pyroelectric detector is sensitive to infrared, the detector is placed far (- 1.5 m) from the prism to avoid detection of the fundamental wave. Harmonic power is monitored by a Molectron EPM 1000 joule/powermeter. The conversion efficiency is found to be 1.5 % with a 5 mm sample length. This efficiency has been found to be comparable to phase-matched potassium dihydrogen phosphate (KDP) crystal. The experimental values of SHG for DIVA must be taken only as a lower bound because the scattering from the crystal surface and the phase-matching direction observed may not be the most favorable for efficiency second-harmonic generation.

SUMMARY AND DISCUSSION

Single crystals of (2-methoxyphenyl)-methylene-propanedinitrile (DIVA) and its derivatives have been grown using the Bridgman-Stockbarger technique. A transparent growth chamber has been designed for observation of the growth process in real time. The morphology of the solid-liquid interface which is a major factor in determining the crystal quality has been monitored during the growth process. A high temperature gradient (>10 °C/cm) is required at the early stage of the growth process to prevent supercooling. Then the temperature of the cold-zone is adjusted in order to obtain a flatinterface. A planar or slightly convex interface shape is maintained during the growth of DIVA and its derivatives within growth conditions.

ACKNOWLEGEMENTS

This work was performed under NASA Alliance for Nonlinear Optics, NAG5-5121, NAG8-1390, and NSF grant HRD9353548.

REFERENCES

C.H. Grossman, T. Wada, S. Yamada, M. Kaji, H. Hakayama and A.F. Garito, In Nonlinear Optics of Organics and Semiconductors Vol. 36, Ed. T. Kobayashi, Springer-Verlag, Berlin, 1989.

Ronald D. Clark, Alliance for Nonlinear Optics, 1st semiannual project review meeting, 1995.

M.D Aggarwal, W.S. Wang, J. Choi, Tiffany Mayers, Ronald D. Clark, Benjamin G. Penn, Donald O. Frazier, Mohan Sanghadasa, J. Cryst. Growth 166, 542-544, 1996.

Dimension Reduction in AVIRIS and Landsat TM Images using Subset Selection Methods

Francisco Pagán, Gianna Fernández,
Advisors: Miguel Vélez-Reyes, and Luis O. Jiménez,
Tropical Center for Earth and Space Studies
Electrical and Computer Engineering Department
University of Puerto Rico Mayagüez Campus
P.O. Box 9042, Mayagüez, PR 00681-9042
Tel. (787) 832-4040 Ext. 3086, FAX (787) 831-7564
e-mail: mvelez@exodo.upr.clu.edu, jimenez@exodo.upr.clu.edu

1.. Summary

Observations from hyperspectral sensors lead to high dimensional data sets having hundreds of narrow spectral bands. A major challenge in handling such data sets is the removal of redundant information and to keep that information relevant to the application at hand. The standard approach to deal with such redundancy is the use of principal component analysis. In this approach the original hyperspectral image is transformed by means of linear combinations into a set of uncorrelated or orthogonal "images." A well known fact for multispectral and hyperspectral images is the fact that most of the spatial information content is summarized by the first few principal components [.61. A disadvantage of this approach is the inherent transformation of the original hyperspectral image (physical meaningful spectral data) into linear combinations of bands with, in many instances, little or no physical relation with the spectral information content on the original image. Therefore, it is of interest to look at methodologies that reduce the hyperspectral data set dimensional ity with no transformation of the data with the same performance as principal components.

In this paper, we will show the application of subset selection methods for the reduction of dimensionality in hyperspectral images and discuss potential applications to band selection and lossy compression. Subset selection is the problem of determining the most linearly independent columns in a matrix [2,3,4]. The reduction of dimensionality of the hyperspectral data set is achieved by exploring its near linear dependency structure. The linear dependency structure of the hyperspectral data set can be derived using the singular value decomposition (SVD) as described in [3,4]. The information contained in the SVD allows us to identify the most independent bands in the hyperspectral image. It can be shown, under certain conditions, that those bands selected using the SVD-based subset selection method of [3,4] will explain the same spatial variability as the principal components will do. It is shown in [2] that the canonical correlation between the principal components and the bands selected using the subset selection procedure will be close to 100% if the matrix associated with the hyperspectral data set is near rank deficient.

Numerical examples using AVIRIS and LANDSAT images are used to illustrate the potential applicability of the technique, wc will show how by using the SVD subset selection method of an AVIRIS image (224 bands) we can reduce the data set down to 7 bands that have a canonical correlation Of 99.9% with thc principal components. Also, due to thc simplicity of handling, we will use data from the LANDSAT multi spectral sensor to illustrate the concepts of subset selection and its combinatorial nature. In this case, results show that for the example at hand we can summarize all information in 1 principal component. Visual interpretation illustrates that the band selected with the subset selection method has most of the spatial structure present in the first principal component. In this case, the selected band has a 98% of cm-relation with the first principal component.

2. Principal Components

Principal Components [5] is a linear transform that has been widely used in data analysis and compression. The spectral bands of a multi-spectral image most commonly do not contain completely independent information. More likely there will be some degree of correlation between bands, indicating that they share elements of information in common. The principal components transform depicts the reflectance levels for a set of pixels by plotting their positions in what is commonly called band space. Each of the axes represents reflectance in thc spectral band indicated. When an image pixel is plotted in this space by placing its location at the intersection of its reflectance level on each band, a significant correlation between bands can be noticed which shows that they do not carry independent information and therefore that there is some degree of redundancy in the information they carry. It is this redundancy that Principal Components seeks to remove when it is used for the purpose of data compression. Unless they are perfectly correlated,

some independent information will always exist in each band. Thus removing data after transforming will always implies some loss of information. Principal Components Analysis uses the correlation among a set of variables to reformulate them so as to concentrate their variability into as few dimensions as possible.

Let X be the hyperspectral data set arranged in matrix form where each column of X corresponds to a sensor channel and each column is constructed by stacking the columns of the image. The singular value decomposition (SVD) of X is given by

$$X=U\Sigma V^T \tag{1}$$

where $U=[u_1, u_2, \ldots . u_n]$ and $V=[v_1, v_2, \ldots . v_n]$ are orthonormal matrices of the left and right singular vectors respectively, $\Sigma=diag\{\sigma_1, \sigma_2, \ldots . \sigma_n\}$, and σ_i is the i-th singular value of X. The singular values are ordered in according to magnitude $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. The matrix $\mathbf{X}$ is an mxn matrix where n is the number of channels and m is equal to the number of columns times the number of rows for each image. The columns of the matrix X are constructed by stacking the columns of the image at each channel.

For hyperspectral data, due to redundancy, it is the case that $\sigma_p >> \sigma_{p+1}$ for some $p<n$ which implies that the matrix X is near rank deficient. Therefore, most of the spatial information about the hyperspectral image is summarized in a p-dimensional subspace, as we shall see later, principal components is only one possibility for obtaining the summarizing information. It turns out that in most hyperspectral applications $p<<n$ resulting in a significant reduction of dimensionality.
The j-th principal component is given by

$$\sigma_j \mathbf{u}_j = \mathbf{X}\, \mathbf{v}_j, \quad j=1,2,\ldots,r$$

The percentage of variability explained by each individual principal component is given by

$$\%Var_{j-th\ component} \quad \frac{\sigma_j^2}{\sum_{i=1}^{n} \sigma_i^2} \tag{2}$$

Since U is orthonormal, i.e.

$$\mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 1 & i = j \\ O & i \neq j \end{cases}$$

data reduction using principal components is achieved by transforming the original image into a set of orthogonal images (or principal components), and then keep only the first p. For later reference we will denote by $\mathbf{U}_1$ the matrix formed by the first p principal components, i.e. $U = \left[\mathbf{U}_1 \quad \mathbf{U}_2\right]$ The total percentage of' variability explained by the first p principal components is given by

$$\%Var \cdot \frac{\sum_{i=1}^{p} \sigma_i^2}{\sum_{i=1}^{n} \sigma_i^2} \tag{3}$$

A clear disadvantage of this approach is the inherent transformation of the original hyperspectral image X (physical meaningful spectral data) into linear combinations of bands U with, in many instances, little or no physical relation with the spectral information content on the original image.

Therefore, it is of interest to look into data reduction methods where in the reduction process the original hyperspectral image is not transformed. What we would like to do then is to select p-bands from the original hyperspectral image that will explain about the same percentage of variability as the first p-principal components. From now on wc refer to this as *subset selection analysis* (SSA). The information contained in the SVD is used to estimate the effective dimensionality of the hyperspectral image. We will use methods for SSA based on the SVD of X.

3. Subset Selection

Subset Selection is the problem of determining the most linearly independent columns in a matrix [2,3,4]. The reduction of dimensionality of the hyperspectral data set is achieved by exploring its linear dependency structure. The linear dependency structure of the hyperspectral data set can be derived from its Singular Value Decomposition (1) as described in: [3,4]. The subset selection algorithm described in [3,4] allows us to identify the most independent bands in the hyperspectral image. The algorithms determines a permutation matrix P such that in the reordered matrix $\overline{\mathbf{X}} = \mathbf{XP} = [\overline{\mathbf{X}}_1 \overline{\mathbf{X}}_2]$, the submatrix $\overline{\mathbf{X}}_1$ will contain the p-most independent columns of X. Notice that the SVD of X and of $\overline{\mathbf{X}}$ are related as follows

$$\overline{\mathbf{X}} = \mathbf{XP} = \mathbf{U\Sigma V}^{\mathrm{T}}\mathbf{P} \tag{6}$$

It is shown in [2], that the sine of the angle between the subspaces spanned by the first p principal components $\mathbf{U}_1$ and the subspace spanned by the columns of $\overline{\mathbf{X}}_1$ is bounded above as follows

$$\sin\theta\left(\Re(\mathbf{U}_1),\ \Re(\overline{\mathbf{X}}_1)\right) \le \frac{\sigma_{p+1}}{\sigma_p}\left\|\overline{\mathbf{V}}_{22}^{-1}\right\| \tag{7}$$

where $\overline{\mathbf{V}}_{22}$ is a well conditioned (n-p) x [n-p) submatrix obtained from the bottom right block of the matrix $\mathbf{P}^{\mathrm{T}}\mathbf{V}$. For both spaces to contain the same information, they must be aligned, i.e. $\sin\theta\left(\Re(\mathbf{U}_1), \Re(\overline{\mathbf{X}}_1)\right) \approx$ o. This will happen if the gap condition is met, that is $\sigma_p >> \sigma_{p+1}$, and $\left\|\overline{\mathbf{V}}_{22}^{-1}\right\| \approx \mathbf{1}$. The first condition depends on near rank deficiency and the second on how the permutation matrix P is selected. In the following discussion, we will use the cosine of the angle since it is equal to the canonical correlation between the first p principal components and the p most independent columns.

4. Experimental Results

The images studied were selected from the LANDSAT Thematic Mapper and AVIRIS Hyperspectral Sensor. The LANDSAT image was from the Añasco Coast in Puerto Rico and the AVIRIS image is from a corn and soy field in Indian Pine, NW Indiana. The AVIRIS Image was provided by Purdue Laboratory for Applied Remote Sensing (LARS).

4.1 LANDSAT Thematic Mapper

Major characteristics and changes to the surface of the planet can be detected, measured, and analyzed using LANDSAT data. The effects of desertification, deforestation, pollution, cataclysmic volcanic activity, and other natural and anthropogenic events can be examined using data acquired from the LANDSAT series of Earth-observing satellites. The information obtainable from the historical and current LANDSAT data play a key role in studying changes to the Earth's surface. LANDSAT multispectral scanner (MSS) data provide a historical record of the Earth's land surface from the early 1970's to the early 1990's. LANDSAT Thematic Mapper (TM) data provide land surface information from the early 1980's to the present. LAND SAT 4 and 5 carry both the MSS and the TM sensors; however, routine collection of MSS data was terminated in late 1992. They orbit at an altitude of 705 km and provide a 16-day, 233-orbit cycle with a swath overlap that varies from 7 percent at the Equator to nearly 84 percent at 82 north or south latitude. These satellites were also designed and operated to collect data over a 195 km swath. The MSS and TM sensors primarily detect reflected radiation from the Earth surface in the visible and near-infrared (IR) wavelengths, but the TM sensor provides more radiometric information than the MSS sensor. The wavelength range for the TM sensor is from the visible (blue), through the mid-IR, into the thermal-IR portion of the electromagnetic spectrum. The TM sensor has a spatial resolution of 30m for the visible, near-IR, and mid-IR wavelengths and a spatial resolution of 120m for the. thermal-IR band. A LANDS-AT 4 and 5 TM scene has an instantaneous field of view (IFOV) of 30m square meters in bands 1 through 5 and band 7, band 6 has an IFOV of 120 square meters on the ground. [7]

4.2 AVIRIS Hyperspectral Sensor

The Airborne Visual Infrared Imaging Spectrometer (AVIRIS) main objective is to identify, measure and monitor constituents of the Earth's surface and atmosphere. All materials reflect, absorb or emit photons in ways characteristic of their molecular makeup. The intensity of radiation versus spectral wavelength shows a unique spectral signature for any given material. Absorption and emission bands specific to a given substance often occur across narrow bandwidths (10 nm or less). If a sensor has similarly tine spectral resolution, these details can be detected. Thus hyperspectral remote sensing might distinguish to a finer degree the spectral features of materials seen by the sensor.

AVIRIS has 224 bands (different detectors) in the visual and infrared part of the spectrum. Each band has a wavelength sensitive range (spectral bandwidth) of approximately 10 run, allowing a spectral coverage range of 400-2450" nm and a 17 m resolution (with some overlap between pixels). AVIRIS uses a scanning mirror to sweep back and forth ("whisk broom" fashion), producing 614 pixels for the 224 detectors each scan. The instrument flies aboard a NASA ER-2 airplane (a U2 plane modified for increased performance) at approximately 20 km above sea level, at about 730 km/hr. AVIRIS has flown all across the US, plus Canada and Europe. Research with AVIRIS is dominantly directed towards understanding processes related to the global environment and climate change. AVIRIS research areas include ecology, oceanography, geology, snow

hydrology, cloud and atmospheric studies. [1,8]

4.3 Experimental Results

The subset selection algorithm of' [3,4] was implemented in MATLAB™. A flowchart describing the implemented algorithm and preprocessing is shown in Figure 1. As part of the pre-processing these images were normalized so that each image matrix has values between zero and one in order to have a good visual display of the bands of these images. The command

$$[\mathbf{U},\mathbf{S},\mathbf{V}] = \mathrm{SVD}(\mathbf{X},0);$$

computes the "economy size" singular value decomposition. The matrix U contains the principal components associated with non-zero singular values, the matrix S contains the singular values, and the matrix V contains the singular vectors. By visual inspection of a semilog plot of the singular values of X we can estimate the effective dimension of the hyperspectral data set, p. Once p is determined the permutation matrix P is determined from the resulting pivoting matrix of the QR factorization with pivoting of the first p columns of the matrix of left singular vectors V. This is done using the command

$$[\mathbf{Q},\mathbf{R},\mathbf{P}] = \mathrm{QR}(\mathbf{V}(:,1{:}p));$$

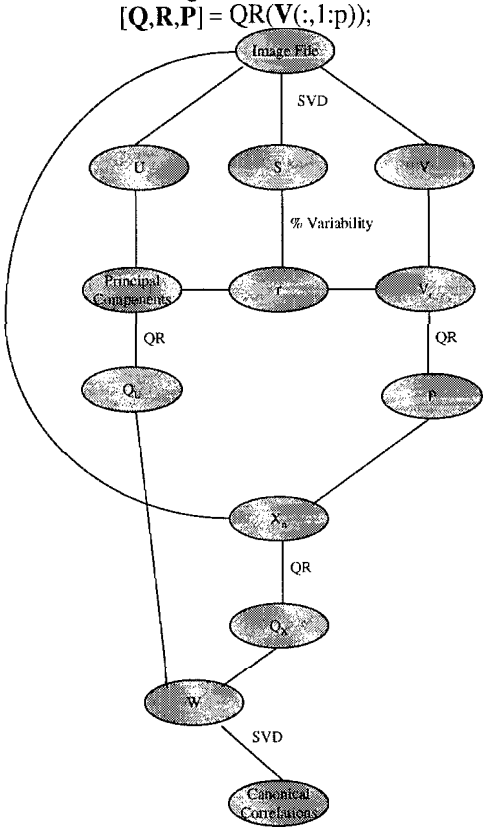


Figure 1: Algorithm flowchart.

5. Results

5.1 LANDSAT Thematic Mapper Image

The LANDSAT image used in this study is shown in Figure 1. After doing the principal components analysis for this image was noticed that the first principal component summarized most of the spatial information with 95.6% of' the total variation. The permutation matrix obtained using QR factorization indicated that the second band image was the band that had almost the same spatial variability y as the first principal component with 98'% of correlation. Visual interpretation illustrates that this second band selected with the subset selection method has most of the spatial structure present in the first principal component.

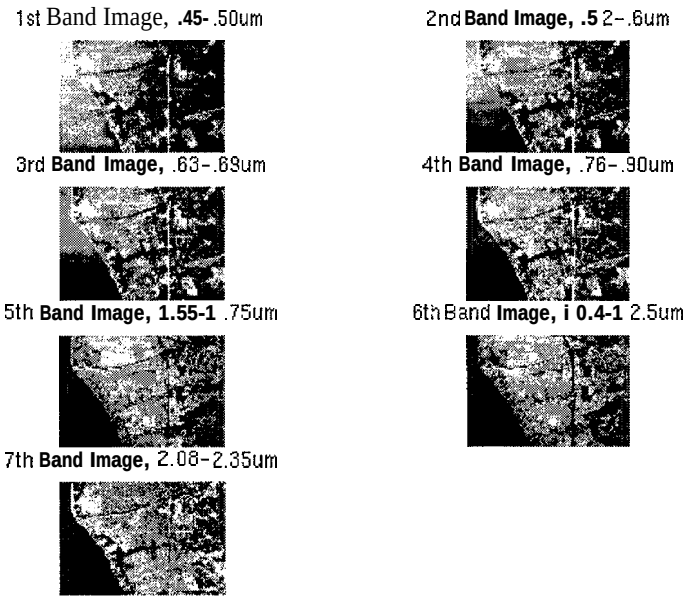


Figure 2: Landsat Image of the Mayagüez Bay.

6. AVIRIS Hyperspectral Sensor Image

The effective dimensionality of the AVIRIS data set was determined to be equal to 7. The first 7 principal components in this case explain 93.5% of the total spatial variability. Using the subset selection method, the seventh most independent bands were 1, 2, 3, 4, 5, 7 and 8. The correlation between these bands and the principal components was 99.99570.

7. Conclusions

This paper presented the application of subset selection to the reduction of multi/hyperspectral images. Results showed that if a gap condition is meet good approximations of the range space of the principal components can be achieved with images from the original data set. Experiments with LANDSAT TM and AVIRIS images showed that selected bands had a 99% canonical correlation with the principal components. The advantage of the subset selection method over principal components method is that the reduced data set is a subset of the original images and not a linear transformation of it like the principal components.

8. References

[1] Barge, L. M., *AVIRIS Data Facility.* http://www.ltid.inpe.br/html/pub/docs/html/aboutAV.htm
[2] Chan, T.F. and P.C. Hansen, "Some applications of the rank revealing QR factorization." In *SIAM J. Sci. Stat. Comput.*, Vol. 13, No. 3, pp. 727-741, May 1992.
[3] Golub G, V. Klema, and G.W. Stewart, *Rank Degeneracy and Least Squares Problems,* TR-456, Department of Computer Science, University of Maryland, College Park, MD, 1976.
[4] Golub, G. and C.F. Van Loan, *Matrix Computations,* John Hopkins University Press, Baltimore, MD, 1997.
[5] Jolliffe, I.T., *Principal Component Analysis,* Springer-Verlag, "1986.
[6] Richards, J.A., *Remote Sensing Digital image Analysis,* Springer-Vcrlag, New York, 1994.
[7] Thematic Mapper LANDSAT Data, http://edcwww.cr.usgs.gov/glis/hyper/guide/landsat_tm
[8] NASA. "The NASA hyperdeck a hyperspectral Remote sensing primer."

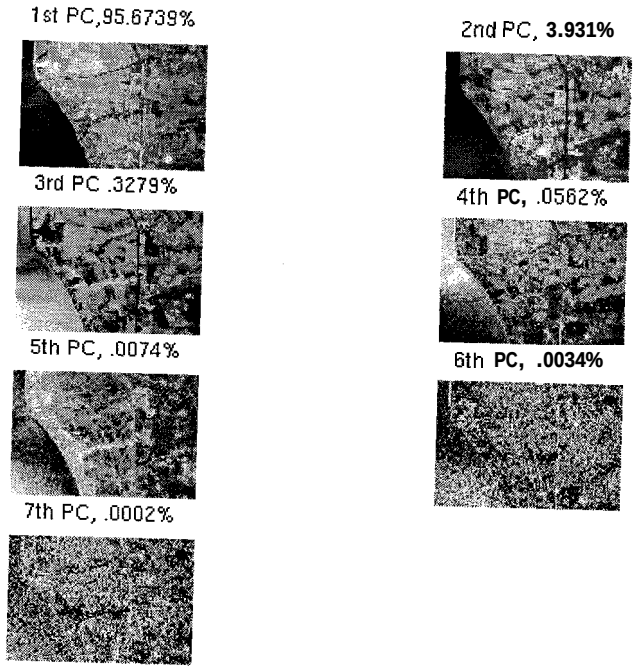


Figure 3: Landsat Image Principal components and their percentage of variability.

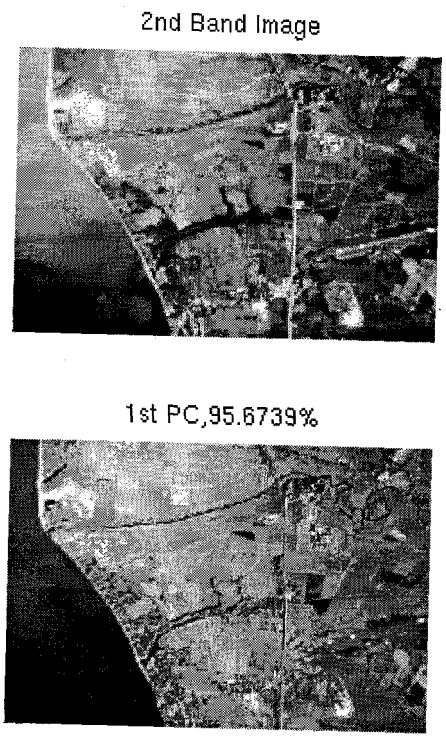


Figure 4: Visual comparison of Landsat band #2 and 1st principal component.

A Discussion of Issues in Integrity Constraint Monitoring

Francisco G. Fernandez Ann Q. Gates Daniel E. Cooke*
Department of Computer Science, The University of Texas at El Paso
ffernand@cs.utep.edu agates@cs.utep.edu dcooke@cs.utep.edu

Abstract

In the development of large-scale software systems, analysts, designers, and programmers identify properties of data objects in the system. The ability to check those assertions during runtime is desirable as a means of verifying the integrity of the program. Typically, programmers ensure the satisfaction of such properties through the use of some form of manually embedded assertion check. The disadvantage to this approach is that these assertions become entangled within the program code. The goal of the research is to develop an integrity constraint monitoring mechanism whereby a repository of software system properties (called integrity constraints) are automatically inserted into the program by the mechanism to check for incorrect program behaviors. Such a mechanism would overcome many of the deficiencies of manually embedded assertion checks.

This paper gives an overview of the preliminary work performed toward this goal. The manual instrumentation of constraint checking on a series of test programs is discussed, This review then is used as the basis for a discussion of issues to be considered in developing an automated integrity constraint monitor.

1 Introduction

Software development is a challenging task. Advances in technology have made it possible to solve problems which were technically unfeasible only a few years ago, and the trend still continues. Hardware advances are taking place at incredible rates, and the problems that we are technically capable of solving are becoming more complex. The difficulty is that the more complex the problems become, the harder it is to develop the solutions. The complexity of the solutions make it almost impossible for any single individual to entirely understand them.

An important part of developing a software solution to a given problem is identifying the key properties of the solution that indicate it is correct. Such properties can be identified at any stage of a software development cycle. For example, consider a waterfall software development model. Key properties of the software system may be specified explicitly during the requirements phase by domain experts, end-users and requirements analysts. During fictional specification, analysts identify new properties implicit in the requirements documents or may find properties while clarifying ambiguities in the requirements. In the design and implementation phases, the system designers may make assumptions about the system, such as the environment in which it will be used, and these assumptions can then become implicit or explicit properties about the execution environment of the system. Finally, *in* the maintenance phase, new properties may be added or previously identified properties may be altered modifying key properties around which the original solution was developed.

As the complexity of a problem increases, so to does the amount of knowledge required to understand and address its solution. As a consequence, it becomes more difficult for an individual to fully understand a complete solution system. This situation prompts the need for communication among solution developers which usually takes the form of a central repository of system documentation (i.e., requirements analysis documents, functional specification documents, interview recordings, and so on). These communication lines are potentially unreliable in the sense that as the solution complexity increases, key information becomes buried within the documentation and is easily lost or forgotten.

This paper looks at an approach, called *integrity constraint monitoring,* currently being investigated as a way to overcome these challenges. Section 2 gives an overview of the integrity constraint monitoring approach. Section 3 describes a study *in* runtime monitoring performed as part of the research. In section 4, the approach to developing an automated monitoring mechanism is outlined. Section 5 closes the paper with a brief summary and description of future directions for our work.

* This work was sponsored by NASA under contract NAG-1012 and NCCW-0089.

2 Integrity constraint monitoring

Integrity constraint monitoring [F96][G97] is an approach in which the key properties of the system are identified and captured in a central repository of integrity constraints and checked during execution of the program. Unlike previous approaches to constraint or assertion checking, an effort is being made to separate the constraints from the implementation. The integrity constraint repository is maintained separate from, but concurrently with, the development documentation. As software development proceeds, constraints are added to the central repository as they are identified. These constraints are also tagged with information about their source so that if justification of the constraint should become necessary, it can be traced back to its original source [D97] [S97]. This central repository of constraints not only provides a single communication point for the disbursement of key properties to all people involved in the development, but also provides information for a dynamic monitoring system that checks program correctness (with respect to the integrity constraints) at runtime. As a program executes, the execution is monitored against the integrity constraint repository to determine if the constraints are being enforced.

The procedures for identifying integrity constraints and for providing traceability of integrity constraints are discussed in other papers [G97][D97][S97]. This paper focuses on the work currently being done to develop a compiler that autonomously embeds constraint-checking code for runtime monitoring of the program.

3 A study on integrity constraint monitoring

Our approach toward developing a runtime monitoring mechanism first required us to perform studies on integrity constraint monitoring issues. In particular, we were interested in learning more about the interaction between a program's execution and the inserted integrity constraints. This section describes the study environment and surmises we have drawn from the information gathered to date.

3.1 Study environment

Our first concern in studying integrity constraint monitoring was in finding a pool of source programs to use in the study. An *Intensive Problem Solving* course, an independent study course intended to teach students strategies for problem solving and programming, served as the vehicle to provide a collection of programs. The lecture portion of the course provided strategies and approaches for solving problems, while a lab section gave students the opportunity to apply these strategies in practice. The lab sessions consisted of eleven lab assignments designed to expose students to a variety of common problem types such as numerical computations, string manipulations and graph problems.

The process behind the study was carefully designed and applied to learn as much about integrity constraint monitoring as possible. The first step was to precisely define a problem statement that would serve as a lab assignment. Essentially, the problem statement was developed around a specific problem class (e.g., string manipulation), Integrity constraints were identified from the problem statement and assumptions made about the program environment (e.g., the input values are positive integers). The next step was to develop a test suite to use in checking a solution's correctness. The identified integrity constraints were then reviewed to determine their appropriateness. This process determined incorrect or unusable constraints (e.g., those that were infeasible with respect to time or space), Any incorrectly defined constraints would be revised. At this point, the students would be given the problem statement and asked to provide a solution. Each student would submit his or her solution, which would be carefully tested using the test suites developed earlier. Any incorrect submissions would be saved for use in the study and returned to the student for correction. Accepted submissions were also kept. Finally, the last phase of the study involved taking the students' programs and manually embedding constraint checks, running the modified programs with the test suites given, and observing the results.

3.2 Example problem

To motivate the discussion in the next section, it may be helpful to have an example of the types of problems presented. The description of a problem statement and example of a constraint follows.

Problem: *Write a program to simulate transactions on a cash register. The cash register contains a collection of money of the following denominations: $20 bills, $10 bills, $5 bills, $1 bills, quarters, nickels, dimes and pennies. Transactions consist of a sale amount (given as a two-decimal-place fixed-point number) and a list of the currency given by the customer. For each transaction, the program should calculate the amount of change to be returned to*

the customer and remove that amount from the register using the least number of bills and/or coins from the currency available.

Sample constraint: *The amount of change returned to the customer is equal to the amount of money given by the customer minus the sale amount.*

Problem statements would include a detailed description of the formats for input and output. These descriptions were omitted here due to space restrictions.

3.3 **Results**

Although the work in manually embedding integrity constraints in programs from the *Intensive Problem Solving* course continues, some preliminary conclusions have been made. A suite of programs for the *cash register* problem described above has been completed (a total of 27 programs from 10 students, with approximately 2-3 submissions per student). In addition, at least three submissions for every problem given in the course also have had constraints embedded. From these tests, we have identified several issues which need to be addressed.

First, we identified the need to study the *not null* problem. Essentially, the problem here is determining if a memory location being examined contains a value that was either read into or assigned to that space. Without specialized hardware or software flags, there is no way to determine if a memory word has been initialized/assigned by a program. As a consequence, there is no way to determine if it is appropriate to check a constraint against such a memory location, since the location may be invalid. in addition, if the memory location represents a data value for a constraint that appears within a loop structure, it is often desirable to "invalidate" the value from one iteration of the loop to the next. Without hardware or software flags to indicate this, there is no straight-forward method for determining if a value is valid or "null".

Next, we realized that constraints often require special knowledge in order to perform a check on it. As an example, consider the cash register problem. In the problem definition, the program need only keep track of the state of the various denominations of money in the register at any time. One of the identified constraints requires checking a transaction by computing the difference between the money given by the customer (described only by denominations, not by the real value of the money) and the sum of the sale amount plus the change returned to the customer. In order to perform such a check, it is necessary to be able to convert a set of denominations into a real value to compute the difference. This type of knowledge is needed by, but is distinct from, the integrity constraint check itself.

A third important note is that the methodology for identifying and applying integrity constraint monitoring needs to be extended. Many of the problems required the program to display information in a pre-defined ordering or to output a specific message on the occurrence of a particular event. For example, in the cash register problem, the cash register contents were to be output in descending order by denomination. The problem that arises from such specifications is that any constraints defined to check these conditions rely on being able to watch the output and determine if it is correct; however, it is often the case that programs output results directly to an output device without first storing it in a buffer prior to output. The challenge is then to find a way to intercept the output and perform a check on it before it goes to the output device. This problem could easily be overcome if the integrity constraint methodology required all output to be stored in an output buffer prior to being sent to an output device.

Another issue discovered in the study was the difficulty of developing an automatic process for identifying the points in a program execution where integrity constraints should be performed. This issue is centered around a high-level classification of constraints identified in previous work [G96]. This classification distinguishes integrity constraints as being one of three types. A description of each follows.

1. *Immediate constraints* are integrity constraints that must hold during all states of a program's execution.
2. *Intermediate constraints are* integrity constraints that must hold after a sequence of state changes.
3. *Delayed constraints* are integrity constraints that must hold in the final state of program execution,

Immediate integrity constraint checks should be performed after any modification to a variable appearing in the constraint definition because they must hold during all states of a program execution,. Delayed constraint checks are performed at the end of program execution. Intermediate constraints, however, present an interesting challenge. Intermediate constraints are checked after a sequence of events *(i.e.,* variable modifications) occur. The problem is

that the events need ***not*** be performed in a fixed order. In addition, it is possible that unrelated program statements are executed in between related events. As a further complication, some events may be conditional, meaning they occur only when specific conditions exist. Thus, identifying that a sequence of events has or has not taken place is problematic. As a result, identifying the points in a program at which to perform an intermediate constraint check remains a problem to be overcome.

Finally, the need to develop a language for expressing integrity constraints was made apparent. English-language descriptions proved to be too ambiguous to properly translate into a program integrity constraint check. Furthermore, it is inadequate because of the inability to translate a description into an automatic check. First-order logic was tried, but several of the identified constraints were difficult to describe using this approach. In particular, integrity constraints that span several states of program execution were troublesome. Other forms of formal mathematical descriptions were also tried, but encountered similar difficulties. It was determined that what is needed is a language that follows the same model as integrity constraints themselves, i.e., a language that captures the notion of states of execution in a non-deterministic fashion. Again, this is important because the integrity constraints are for the most part implementation-independent.
These issues are being studied concurrently with the development of an automated integrity constraint embedding system. The development of this system is described in the next section.

4 Towards development of a constraint monitoring system

The approach currently being investigated for developing a constraint monitoring system is centered around a compiler that embeds integrity constraint checks in a given source program. The premise of this approach is the use of Petri nets to model the control- and data-flow of a program execution. It is believed that Petri nets can be used in conjunction with an integrity constraint specification to determine the points in program execution where constraint checks should be embedded.

The basic idea behind the approach is to compile a program to an intermediate representation, called a *quad representation* [A85]. A quad representation is a 4-tuple representation of an instruction with an operation, up to two operands, and a result. Once this intermediate representation is built, it is decomposed into *basic blocks*, which are sequences of instructions in which the control of the program is strictly sequential with no conditional execution within that sequence. The basic blocks can then be used to build a Petri-net representation of the program's control flow, where a basic block is represented as a transition. The Petri net can then be annotated with data assignment information, represented as places in the model.

This approach is being implemented on a simple Pascal-like language that provides only three types of control constructs:

1. Sequential: a series of program statements executed one after the other.
2. Conditional: an *if-then-else* construct. Program control depends on the truth value of a stated condition.
3. Loop: a while-loop construct. Program control loops over a give sequence as long as a condition holds.

Note that the three constructs given above are orthogonal to each other, so that one type of construct can be embedded within another type. Several programs in the language have already been modeled in Petri nets, and the characteristic Petri nets that exemplify loop and conditional control constructs are shown below.

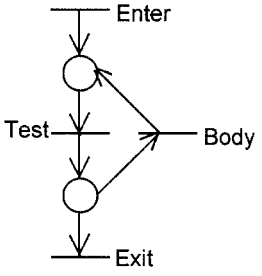


Figure 1. Characteristic while-loop Petri Net Construct

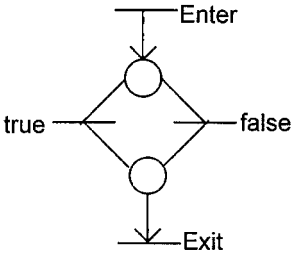


Figure 2. Characteristic if-then-else Petri Net Construct

In addition, the language provides for two types of data-flow constructs. These are the *read/readln* statement and the assignment statement.

Using these basic block Petri net representations, it is believed that the points of execution where an intermediate integrity constraint check should be embedded can be automatically determined, since the transitions of the model represent basic blocks, and basic blocks contain only one entry and one exit point. By abstracting a sequence of program statements into a single block and indicating the variable modifications (i.e., the events) that occur within that basic block, it may be possible to automate a process for identifying intermediate constraint check points. A key assumption of this approach is that at most a variable can be have a value assigned to it once within a basic block, and in at most two basic blocks. This restriction limits the ways in which a monitored variable can be modified and, therefore, limits the locations in which a constraint check need be placed. It is possible that by decomposing a program into basic blocks, a sequence of events will in fact occur within a single block and, as a result, an intermediate constraint check would need to be performed at the end of that block.

5 **Summary**

The integrity constraint monitoring approach shows promise as a powerful tool to aid in the development of software systems for solving the complex problems. Work on the approach is on-going, with research being conducted on several different aspects. At present, research is underway to develop an integrity constraint definition and traceability tool. The tool will provide a graphical-user interface that allows developers and system designers to define integrity constraints, and to link those constraints to documents that are the source of constraints. In addition, work is being done to develop a language for expressing integrity constraints in documentation. The language is also based upon the Petri-net representation of program control-flow modeling. Finally, work continues in the development of a compiler for automatically embedding integrity constraint checks in program source code for runtime monitoring of integrity constraints. Future work includes research into providing a separate, external monitor that could watch a program's execution from the outside and ensure that integrity constraints are being enforced. in addition, work will be done to look at ways of improving the performance of integrity constraint monitoring through the use of concurrent processes, including the possibility of using multi-processor systems for allowing parallel execution of a program and its associated constraint monitoring process.

6 **References**

[A85] Aho, A. V., Sethi, R. and J.D. Unman, *Compilers: Principles, Techniques and Tools.* Reading, MA: Addison-Wesley, 1985.

[D97] Delgado, N., and T. Watson, "An Approach to Building a Traceability Tool for Software Development," *Proceedings of the NASA University Research Center's Technical Conference on Education, Aeronautics, Space, Autonomy, Earth and Environment,* pp. 197--202, Albuquerque, NM, February 1997.

[F96] Gates, A., F. Fernandez, and L. Romo, "Building Systems with Integrity Constraints, " *Proceedings from the Second World Conference on Integrated Design and Process Technology Vol.1,* pp. 437--444, December 1996.

[G96] Gates, A., "On Defining a Class of Integrity Constraints", *Proceedings of the Eighth International Conference on Software Engineering and Knowledge Engineering,* Lake Tahoe, NV, June 1996, pp. 338-344.

[G97] Gates, A., C. Kubo Della-Piana, "The [identification of Integrity Constraints in Requirements for Constraint Monitoring", *Proceedings of the International Conference and Workshop on Engineering in Computer-Based Systems,* Monterrey, CA, March 1997.

[S97] Saenz, F., and T. Watson, "A Traceability Tool: Linking Knowledge and Programs," to appear *in Proceedings Of the National Conferences on Undergraduate Research-97 (NCUR),* Austin, TX, April 1997.

USING GENETIC ALGORITHMS FOR OPTIMAL CRANE CONTROL

Bahrain Kimiaghalam, Tel: (336) 334-7760 x259, Email: bahram@ncat.edu
Abdollah Homaifar, Tel: (336) 334-7760 x221, Email: homaifar@ncat.edu
Marwan Bikdash, Tel: (336) 334-7760 x225, Email: bikdash@ncat.edu
NASA-Autonomous Control Engineering Center
Department of Electrical Engineering
NC A&T State University, Greensboro, NC

Abstract: The nonlinear dynamics of a crane makes it a good test case for most of the available methods in control and motion planning, specially the relatively new methods such as Genetic Algorithm (GA). There are several approaches to solve a dock mounted container crane control problem using optimal control methods. Usually the necessary conditions for solving this problem require finding the initial costate vector. in this paper real value GA is used to the initial values of the costates of the system. Each individual gene has its own fitness value based on its ability to move the system to desired final states after a given time. in order to evaluate the fitness, we need to have a simulator of the system which keeps to simulate system trajectory. Here we will briefly cover the dynamics of the crane and present some initial results on the performance of GA.

KEYWORDS: Optimal control, Crane control, Genetic Algorithm, Real valued Genetic Algorithms

1. INTRODUCTION

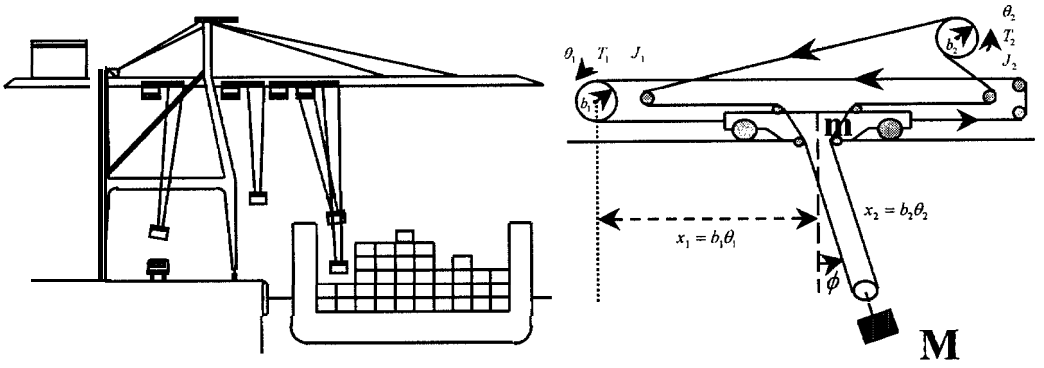
Controlling the motion of a load carried by a pantry crane is a difficult nonlinear control problem that is not easily solvable with conventional control methods. The solution of this problem by optimal control theory would lead to solving a two-point boundary value problem (Sakawa and Shindo, 1982). This problem has also been approached predictive fuzzy method (Yasunobu and Hasegawa, 1986). Since this is a difficult optimization problem with equality and inequality conditions, it is of interest to test whether GA can be used to solve this problem.

We try to solve the problem for the most difficult part of the trajectory of the load using optimal control theory with costates or lagrange multipliers method. To determine the initial co-states for the system, we use them as the elements of the gene.

2. CRANE DYNAMICS AND STATE SPACE MODEL

In this section, we present the dynamical equations governing the dynamics of the crane shown in figure 1.a. This figure shows the a dock based crane that we want to control. The operator's cabin is a part of the trolley in order to give the operator a good view of the load. Figure 1.b shows the trolley, ropes, and two motors of the crane and the pulleys making up the crane. The variables of the system are: θ_1, angle of rotation of trolley drive motor; J_1, total moment of inertia of the trolley drive motor mechanism which includes a brake, a drum and a set of reduction gears; b_1, equivalent radius of the drum of the trolley drive motor which is reduced to the motor side; θ_2, angle of rotation of the hoist motor; J_2, total moment of inertia of the hoist motor mechanism which includes a brake, a drum and a set of reduction gears; b_2, equivalent radius of the drum of the hoist motor which is reduced to the motor side; ϕ, the load swing angle; m, total mass of the trolley and the operators cab; M, total mass of the container load, the spreader and the attached equipment; T_1, driving torque generated by the trolley drive motor; T_2, driving torque generated by the hoist motor and g, the acceleration of gravity (9.81 ms^{-2}).

Following (Sakawa and Shindo, 1982), the model can be obtained using the Lagrangian approach under the following assumptions: The cable does not stretch, there are no damping effects, and the motor dynamics are negligible. The Lagrange's equations of the motion (Meirovitch, 1967) are:



a: The dock based container crane **b: Coordinates and** Torques

Figure 1: Crane geometry

$$T_1 = J_1 + (M+m)b_1^2\ddot{\theta}_1 + Mb_1b_2\ddot{\theta}_2\sin\phi + Mb_1b_2\theta_2\ddot{\phi}\cos\phi$$

$$T_2 = Mb_1b_2\ddot{\theta}_1\sin\phi + (J_2 + Mb_2^2)\ddot{\theta}_2 - Mb_2^2\theta_2\dot{\phi}^2 - Mgb_2\cos\phi$$

$$0 = b_1\ddot{\theta}_1\cos\phi + b_2\theta_2\ddot{\phi} + 2b_2\dot{\theta}_2\dot{\phi} + g\sin\phi$$

Where T_1 and T_2 are the torques of the trolley and hoist motors, Assuming the angle of swinging to be small (which is the case under proper control of swinging), we can simplify the equations by the following approximations: $\phi^\alpha\dot{\phi}^\beta = O$ for $a \geq 0, \beta \geq 0$ and $a + \beta \geq 2$ and $\cos\phi \approx 1$ and $\sin\phi \approx \phi$. With the following definitions:

$$X, = b_1\theta_1, \quad x_2 = b_2\theta_2, x_3 = \phi, \quad x_4 = \dot{x}_1, \quad x_5 = \dot{x}_2, \quad x_6 = \dot{x}_3,$$

$$v_1 = \frac{b_1T_1}{J_1 + mb_1^2}, \qquad v_2 = \frac{b_2(T_2 + Mb_2g)}{J_2 + Mb_2^2}, \quad \delta_1 = \frac{Mb_1^2}{J_1 + mb_1^2}, \qquad \delta_2 = \frac{Mb_2^2}{J_2 + Mb_2^2},$$

The state-space equations where xi are the state-variables, and V, and v_2 are the controls, are:

$$\begin{aligned} &\dot{x}_1 = x_4, \quad \dot{x}_4 = v_1 - \delta_1 x_3 v_2 + \delta_1 g x_3 \\ &\dot{x}_2 = x_5, \quad \dot{x}_5 = -\delta_2 x_3 v_1 + v_2 \\ &\dot{x}_3 = x_6, \quad \dot{x}_6 = -\frac{1}{x_3}[v_1 - \delta_1 x_3 v_2 + (1+\delta_1)gx_3 + 2x_5x_6] \end{aligned} \qquad (1)$$

3. CONTROL PROBLEMS

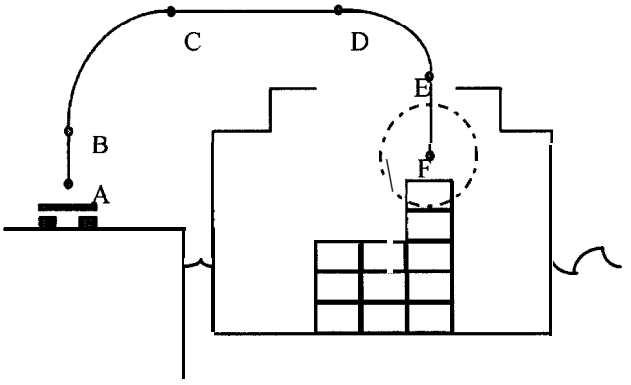


Figure 2: A" desired" trajectory

The objective of the control system is to move the load from one side of a fixed obstacle to a known destination on the other side, without having the load collide with the obstacle and in a relatively short time. We can make the

control problem easier by separating the operation into two phases: a hoist lifting-lowering phase and a translation phase. Then we could assume that the length is a known constant during the translation phase. It would save time if the two phases of operation could be done concurrently. As shown in figure 2.A, there is no collision with obstacles between points A and F in a typical desired trajectory of the load and there are concurrent vertical and horizontal motions from point B to C and D to E. Of course, swinging may cause a big delay in coming to rest at the destination. If we look at different parts of the trajectory there are five areas from point A to B, B to C, C to D, D to E and E to F. We can easily figure out that the most difficult parts to control are diagonal moves from B to C and From D to E because of concurrent changes in hoist and trolley speeds. Of course they have the same behavior, and just solving the control problem for one of them would be enough. We use the cost function:

$$J = \frac{1}{2}\int_0^{t_f}\left[\left(x - x_f^d\right)^T Q\left(x - x_f^d\right) + v^T R v\right]dt + \left(x_f - x_f^d\right)^T S\left(x_f - x_f^d\right) \quad (2)$$

x_f^d is the desired final state at time t_f , x_f is the final state at time t_f , and Q, R and S are arbitrary diagonal weight matrices. We simulate the system from $t = O$ to $t = t_f$, which is the expected time to go from B to C, and then we can calculate the cost for the applied control. The best control will have the least cost throughout the trajectory. The optimal control problem is to minimize the objective function subject to the state equations, initial and final conditions and the constraints. To minimize the cost we write the Hamiltonian as:

$$H(x, v, \lambda) = L(x, v) + \lambda^T f(x, v) \quad (3)$$

Where $L = \frac{1}{2}\left[\left(x - x_f^d\right)^T Q\left(x - x_f^d\right) + v^T R v\right]$

If there are no constraints on v and x, using necessary conditions we will have:

$$1 - \quad \dot{\lambda} = -\frac{\partial H}{\partial x} = -\left(\frac{\partial L}{\partial x} + \left[\frac{\partial f}{\partial x}\right]^T \lambda\right) \quad (4.a)$$

$$2 - \quad 0 = \frac{\partial H}{\partial \lambda} \Rightarrow \dot{x} = f(x, v) \quad (4.b)$$

$$3 - \quad \frac{\partial H}{\partial v} = Rv + \left[\frac{\partial f}{\partial v}\right]^T \lambda \Rightarrow v = -R^{-1}\left[\frac{\partial f}{\partial v}\right]^T \quad (4.c)$$

And then as a result:

$$\dot{\lambda} = \begin{vmatrix} -Q_{11}\left(x_1 - x_{1f}^d\right) \\ -Q_{22}\left(x_2 - x_{2f}^d\right) - \frac{1}{x_2^2}\left[v_1 - \delta_1 x_3 v_2 + (1 + \delta_1) g x_3 + 2x_5 x_6\right]\lambda_6 \\ -Q_{33}\left(x_3 - x_{3f}^d\right) + \delta_1 \lambda_4 \left(v_2 - g\right) + \delta_2 v_1 \lambda_5 - \frac{\left[-\delta_1 v_2 + (1 + \delta_1) g\right]\lambda_6}{x_2} \\ -Q_{44}\left(x_4 - x_{4f}^d\right) - \lambda_1 \\ -Q_{55}\left(x_5 - x_{5f}^d\right) - \lambda_2 + \frac{2x_6\lambda_6}{x_2} \\ -Q_{66}\left(x_6 - x_{6f}^d\right) - \lambda_3 + \frac{2x_5\lambda_6}{x_2} \end{vmatrix},$$

And

$$v = R^{-1}\begin{bmatrix} -\lambda_4 + \delta_2 x_3 \lambda_5 + \dfrac{\lambda_6}{x_2} \\ \delta_1 x_3 \lambda_4 - \lambda_5 + \dfrac{\delta_1 x_3}{x_2}\lambda_6 \end{bmatrix},$$

$x(0)$, x^{df} and $x(t_f)$ are given. Then $\lambda(0)$ can be interpreted as the minimizing value. Hence we minimize $J(\lambda(0))$. To do so, we compute the trajectory for some assumed $\lambda(0)$, by integrating equations (1) and (4.a) with $x(0)$ given and v given by equation (4.c). Then J is computed from equation (2). To apply GA, we have to code $\lambda(0)$ as an individual and maximize a fitness function related to $J(\lambda(0))$. Here we simply take $F = \dfrac{1}{1+J}$ as fitness.

If we add the constraints on the control, equation (4.c) will change to:

$H(x^*, v^*, \lambda^*) \le H(x^*, v, \lambda^*)$ For admissible x and v

4. GA APPROACH TO SOLVING OPTIMIZATION PROBLEM AND STRUCTURE OF THE GENE AND FITNESS FUNCTION

In section 3 we evaluated the proper controls of the system in terms of x, the states and λ, the co-states. Once we have the initial conditions of x and proper initial conditions for λ, we can move the system from that point on using the control inputs calculated from equation (4.c). This problem is a two-point boundary value problem and what GA is going to search for are the initial values for λ.

Selecting the structure of the gene and the fitness function are the two most important setups for a GA implementation. Here we use real-valued GA. The co-states of the system will construct a gene with length six. We use real-valued GA because of its improved accuracy and because there are no determined ranges for the variables. Binary GA needs to search in a pre-determined area to find the optimum values and each binary code should be assigned to one value within the range. Since we don't know the nature of the co-states and we can't find a reasonable range for them, we choose real values to construct genes. Each gene will be a vector of six real numbers representing an initial value set for co-states. Using GA we search for a gene that when we use it as initial conditions of the system and we simulate it for the specified time we can have, the least possible cost.

5. SELECTION, CROSSOVER AND MUTATION

Choosing the appropriate GA operators to be used for any GA problem is very important. We chose tournament selection to keep the variety. A relatively very good individual should not necessarily exterminate most of the bad individuals, as it is in roulette wheel selection. And because we start from the area that can be very far away from the final answer, the difference between the best and the average fitness in one generation can be too large, hence case must be exerted as not to eliminate below average individuals prematurely.

We use a modified Radcliffe crossover in which each component of parent A (in our case, λ_k^A, one of the six components of the initial condition) is randomly crossed with the corresponding component, λ_k^B, of parent B. According to the formula:

$$\lambda_k^c = (1-\alpha_k)\lambda_k^A + \alpha_k \lambda_k^B$$

Where α_k is random number generated as $\alpha_k = rand + 0.15 * randn$, Where rand is uniformly distributed between O and 1, and randn is Gaussian with zero mean and unity standard derivation (as in MATLAB notation).

The Gaussian part of α_k *is* the extrapolation from the range. If α_k *is* the same for all the pixels of an individual and with no extrapolation then the possibilities for the child of the parents A and B is a straight line connecting A and B, as shown in picture 3 for a two dimensional individual. Different α_k for each variable with no extrapolation will extend the chance of the child to be in the dotted area. Adding the extrapolation will increase the area to the bigger box in the picture. This will let the child have the chance to go beyond the range of difference between the parents. Practically, extrapolation is used in place of mutation and the operation will be a considered as a combined crossover-mutation operator. Probabilities of crossover and mutation in this combined operator will be the same.

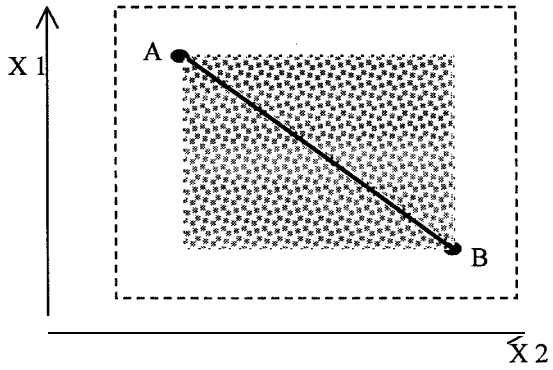


Figure 3: Possibilities for the children of point A and B

By designing the gene structure, fitness function, selection method, and the crossover and mutation operators, the set up is complete and GA can start the search. Choosing the probability of crossover-mutation, initial population size,

6. RESULTS

We have created all the operators needed to implement the GA in MATLAB. For this particular problem the result within generations shows fast increase in the fitness. But after 15 generation still the result is not satisfactory and because of using simulation of the crane for each individual the process takes a long time to come up with fitness. This can be because of high sensitivity to the costates. Since changes in different parameters in GA can change the speed of the convergence, we are trying to adjust them to get to the fastest convergence speed. On the other hand we are trying to find some ways to avoid some of the extra computations to save time and get to the final answer in a reasonable time. Steady-state GA may be a good alternative that we may apply to it in near future.

7. CONCLUSION

GA is a search method and has lots of potentials to be used for different applications. To solve a problem using GA, one should understand the problem clearly and be able to translate it to a search problem. Then the problem would be searching for desired values for some variables to maximize a fitness function. GA as a search method can be used for different control problems but because of its nature that is searching within different possibilities, it would be difficult to use it for on line controls. Decision making process using GA needs processing power and usually it is time consuming. In this application we didn't involve any external forces and still amount of processing power needed is incredible. Adding disturbances to the system makes the problem more difficult and maybe not appropriate for GA method. This study searches for the power of GA in control applications and however the final result hasn't obtained yet but preliminary results show that the direction of the search is satisfactory and this work could be helpful to others who want to use GA for their application in any area.

8. ACKNOWLEDGEMENT

The U.S. Office of Naval Research through contract No. N000 14-96-1-1123 supports this work. The authors would like to thank them.

REFERENCES

1. Sakawa, Y., and Y. Shindo "Optimal Control of Container Cranes," ***Automatic,*** Vol. 18, **No. 3 (1 982),** pp. ***257-266***
2. Moustafa, Kamal and A. M. Ebid "Nonlinear Modeling and Control of Overhead Crane Load Sway," ***Transaction*** *of the ASME*, Vol. 110 (September 1988), pp. 266-271.
3. Filess, M. and J. Levine and P. Rouchon "A Simplified Approach of the Crane Control Via a Generalized State-Space Model," ***Proceedings of the 30th conference on decision and control,*** Brighton, England (December 199 1), pp. ***736-741.***
4. Meirovitch, L., ***Analytical Methods in Vibrations,*** Macmillan, New York (1 967).
5. Yasunobu, S., and T. Hasegawa "Evaluation of an Automatic Container Crane Operation System Based on Predictive Fuzzy Control," *Control-Theory* ***and Advanced Technology,*** Vol. ***2, No. 3*** (September 1986), pp. 419-432
6. Yasunobu, S and T. Hasegawa (1987), "Predictive Fuzzy Control and its Application for Automatic Container Crane Operation System," ***Preprints of second IFSA Congress,*** Tokio, July 20-25 (1987).
7. Lee, C.C. "Fuzzy logic in control systems: fuzzy logic controlled Part I & Part II, " ***IEEE*** *Trans. Syst. Man and* ***Cybernetics,*** Vol. 20 (March /April 1990), pp. 404-435.
8. Davis, L., ***Handbook of Genetic Algorithms,*** Van Nostrand Reinhold, New York (1991).
9. Goldberg, D. E., ***Genetic Algorithms in search, optimization, and machine*** *learning*, Addison-Wesley, Menlo Park, California.

DESIGN OF A FAST FUZZY CONTROLLER*

F. Vainstein, telephone: (336) 334-7760 x222, email: feodor@ncat.edu
C.V. Curtis, telephone: (336) 334-7761 x259, email: cvc@ncat.edu
The NASA Autonomous Control Engineering Center
Department of Electrical Engineering, North Carolina A&T State University

Abstract - The paper presents a new approach to designing a fast fuzzy controller. We propose to use modular arithmetic for computation of the controller's output. The theoretical foundation provided by the Chinese Remainder Theorem, shows that some modular rings are isomorphic to the direct sum of smaller modular rings. In this case, no carry is produced during the process of addit ion and multiplication, thus making parallel computations feasible. Genetic Algorithms is used for hardware minimization.

keywords: fuzzy controller, fast adders, fast multipliers, chinese remainder theorem, modular arithmetic, genetic algorithms, hardware design, optimization methods, control theory.

I. INTRODUCTION

A typical fuzzy controller is shown schematically in Figure 1 [3]. It consists of n sensors, fuzzification circuitry, a memory storing Rule Base and a computational block.

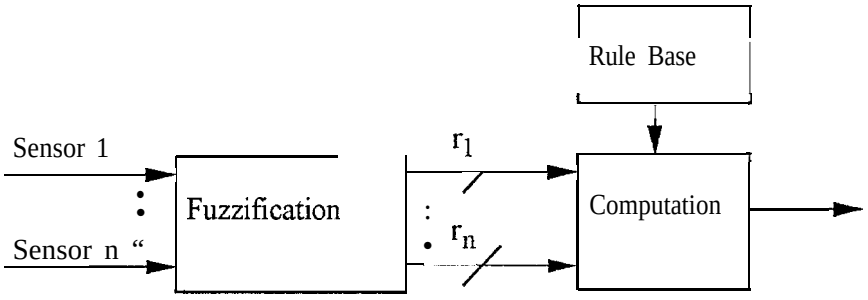


Figure 1. Fuzzy Controller

The i-th output of the fuzzification block (i =1,... n) is a vector (a^i_j) , where j = 1,2,... r_i , of membership functions corresponding to i-th sensor. The output is produced in the computational block according to the formula

$$y = \frac{\sum_{j_1 j_2 \cdots j_n} a^1_{j_1} a^2_{j_2} \cdots a^n_{j_n} y_{j_1 j_2 \cdots j_n}}{\sum_{j_1 j_2 \cdots j_n} a^1_{j_1} a^2_{j_2} \cdots a^n_{j_n}}$$

where the values $y_{j_1 j_2 \cdots j_n}$ are stored in the Rule Base

*This research has been partially supported under grant number ACE-48 146 from the NASA Autonomous Control Engineering Center.

The number of additions and multiplications required for computation using this formula grows exponentially with the number of sensors, while only one division is necessary. Therefore, the problem of speeding up the computational process becomes increasingly important. Traditional methods of speeding up computations, due to their generality, will not provide us with the fastest possible calculations. The use of modular arithmetic based on the Chinese Remainder Theorem (CRT) has been suggested in [1]. In this case, no carry is produced during the process of addition and multiplication, thus making parallel computations feasible.

In short, this method can be explained as follows:
Given a set of mutually prime numbers ml,..., m_k, any integer k, $O \le k \le M$, where $M = (m_1 . m_2 \ldots : m_k) - 1$ is uniquely represented by a string of n numbers $k \bmod m_1, \ldots, k \bmod m_n$. This representation preserves operations of componentwise addition and mult implication.

Example Let $m_1 = 2$; $m_2 = 3$; $m_3 = 5$; $m_4 = 7$;
The integers from O to $M = (2.3.5.7) - 1 = 209$ can be uniquely represented. Table 1 shows the representation of the integers from O through 15.

	mod 2	mod 3	mod 5	mod 7
0	0	0	0	0
1	1	1	1	1
2	0	2	2	2
3	1	0	3	3
4	0	1	4	4
5	1	2	0	5
6	0	0	1	6
7	1	1	2	0

	mod 2	mod 3	mod 5	mod 7
8	0	2	3	1
9	1	0	4	2
10	0	1	0	3
11	1	2	1	4
12	0	0	2	5
13	1	1	3	6
14	u	2	4	0
15	1	0	0	1

Table 1. CRT Number Representation Table

According to Table 1, $2 \mapsto 0222$, $5 \mapsto 1205$. Let us perform the addition and multiplication of the numbers 2 and 5 using the table above.

$2 + 5 = 7 \mapsto 1120$; $0 + 1 \bmod 2 = 1$; $2 + 2 \bmod 3 = 1$; $2 + 0 \bmod 5 = 2$; $2 + 5 \bmod 7 = 0$;

$2 \cdot 5 = 10 \mapsto 0103$; $0 \cdot 1 \bmod 2 = 0$; $2 \cdot 2 \bmod 3 = 1$; $2 \cdot 0 \bmod 5 = 0$; $2 \cdot 5 \bmod 7 = 3$;

The idea of using CRT for computations is not new- [4]. The disadvantage of this approach is that the operation of division is not easily performed, However, a fuzzy controller has to execute the operation of division only once, so for division traditional methods can be used.

The problem of converting CRT numbers to and from binary (or decimal) can be easily solved. Let us denote the function that performs conversion by $\varphi(k) : \varphi(k) = (k \bmod m_1, \ldots . k \bmod m_n)$.
Clearly $\varphi(c_0 + c_1 \cdot 10 + \cdots + c_d \cdot 10^d) = \varphi(c_0) + \varphi(10) \cdot \varphi(c_1) + \cdots + \varphi(c_d) \cdot \varphi(10)^d$ and this formula can be used for converting from decimal to CRT numbers. Conversely, a CRT number $a_1, \ldots . a_k$ can be

presented by $\sum_{i=1}^{k} a_i e_i$, where $e_i = 0, \ldots, 1, \ldots, 0$ (the string consists of all O's, except one 1 in the *i-th* position). The inverse transform, φ^{-1}, can be determined according to the following equation:

$\varphi^{-1}\left(\sum_{i=1}^{k} a_i e_i\right) = \sum_{i=1}^{k} \varphi^{-1}(a_i) \cdot \varphi^{-1}(e_i)$, where the numbers of $\varphi^{-1}(e_i)$ can be precomputed.

We want to acknowledge that the CRT numbers can be obtained by using specialized Analog-to-Digital converters in the initial phase of data processing b y a fuzzy controller. The conversion of CRT numbers to binary (or decimal) is performed only once in the final stage of the computation for division.

Let us compare the speed of CRT adders and multipliers with the traditional fast adders and multipliers. The CRT adder and multiplier are, basically, a set of modulo *m*, adders and multipliers, Even for the case of relatively small m_i, a significant value of *m* can be obtained. For example, if $\{m, \} = \{11,13,17,19,23,25,27,28,29,31\}$, then $M \approx 1.8.1013 \approx 2^{44}$. Modulo m_i adders and multipliers can be implemented using two or three gate level circuits. The result of this comparison is shown in Table 2 (for a two gate level implementation).

	Time to perform operation	Time to perform operation for n=32
Carry Lookahead Tree Adder	$2\lceil \log_2 n \rceil \tau$	10τ
Wallace Tree Multiplier	$2\lceil \log_2 n \rceil^2 \tau$	50τ
CRT Adder	2τ	2τ
CRT Multiplier	2τ	2τ

n = number of bits, τ = average delay on a gate

Table 2. Comparison of Chinese Remainder Adders/Multipliers with Traditional Fast Adders/Multipliers

III. GENETIC ALGORITHMS

The implementation of the hardware for the fast fuzzy controller is discussed in [2]. In this paper, we introduce some improvements on the crossover operator.

The crossover operation is now defined as follows. Suppose we have two chromosomes, $\{h_1^1, h_2^1, h_3^1\}(t_1^1, t_2^1, t_3^1, t_4^1, t_5^1,) = (H^1, T^1)$ and $\{h_1^2, h_2^2, h_3^2\}(t_1^2, t_2^2, t_3^2, t_4^2, t_5^2,) = (H^2, T^2)$. These two parents will always have two offspring (H^1, T^2) and (H^2, T^1). The children have different heads, which they inherited from their parents, and a tail, which is obtained from a special one point crossover that wc shall now describe.

Just as the crossover site is chosen for the simple crossover operator, we shall choose a random crossover site for each string. So we have $T^1 = (t_1^1, t_2^1, t_3^1, t_4^2, t_5^2,)$ and $T^* = t_1^2, t_2^2, t_3^2, t_4^1, t_5^1, .)$ Once our two parents have been crossed, we then check for repeating numbers in each string. If we encounter a repeating number, we replace that number with the parent string's original number. This is continued until the entire string is evaluated. In modulo 5 for example, we are given the first parent string

as *{6, 1, 7)(2,0,5,4,3)* and the second parent string as {7,5, 0}(1, *7,2,6, 4)* to perform a crossover on the tail. We first randomly select a cross site and cross or (,2,0,12, 6,7) and (1,71,5,4,3). Now we must perform our check. This will change our offspring to (2, O,\ 5, 6,7) and (I, 71,2,4,3). This revised operator causes our genetic algorithm to spend less time evaluating illegal children.

IV. RESULTS

We have applied our genetic algorithm to evaluate modulo 5, 7, and 11 adders. Our objective is to use the genetic algorithm to find the number representation, or chromosome, that will give us the smallest number of logic gates. We have defined our fitness function, ψ, as the number of AND gates of the adder implementation. We know that the cardinality of the domain of the fitness function for designing the modulo 5 adder is equal to $\binom{8}{5} \cdot 5! = 6{,}720$. This cardinality for modulo 7 adder is equal to $\binom{8}{7} \cdot 7! = 40{,}320$, and equal to $\binom{16}{11} \cdot 11! = 174{,}356{,}582{,}400$ for modulo 11 adder. In our analysis, We found that the standard fitness for modulo 5 is $\psi_{standard} = 18$, for modulo 7 is $\psi_{standard} = 36$, and for modulo 11 is $\psi_{standard} = 77$. Using the genetic algorithm along with Espresso[6], we determined that the minimum fitness for modulo 5 is $\psi_{min} = 14$ and $\psi_{min} = 24$ for modulo 7. In the following graphs, the upper line shows the maximum fitnesses for the population. The middle line shows the mean of the fitnesses for the population and the lower line represents the minimum fitness values for the

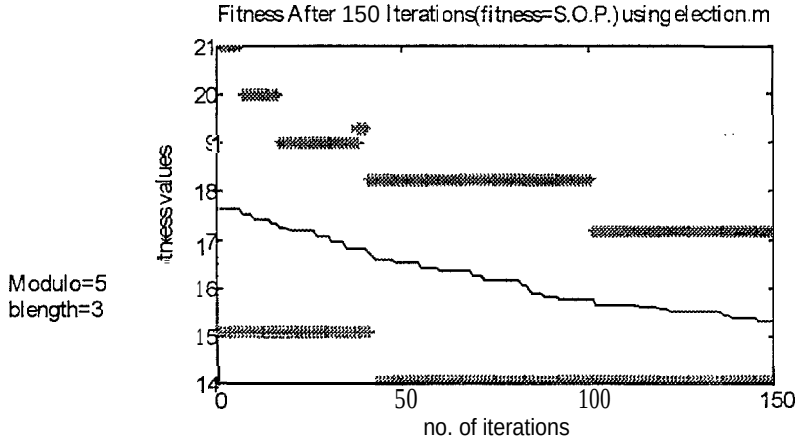


Figure 2 First example of a SSGA for modulo 5.

Figure 3 Second example of a **SSGA** for **modulo** 5.

populations. Looking at Figure 2 and 3, we can see that the Steady State Genetic Algorithm (SSGA) converges before the fiftieth iteration. The crossover probability was set at $p_c = 0{,}85$ for the modulo 5 adder, which is relatively high. At $p_m = 0.1$, the mutation operator value for the modulo 5 adder was

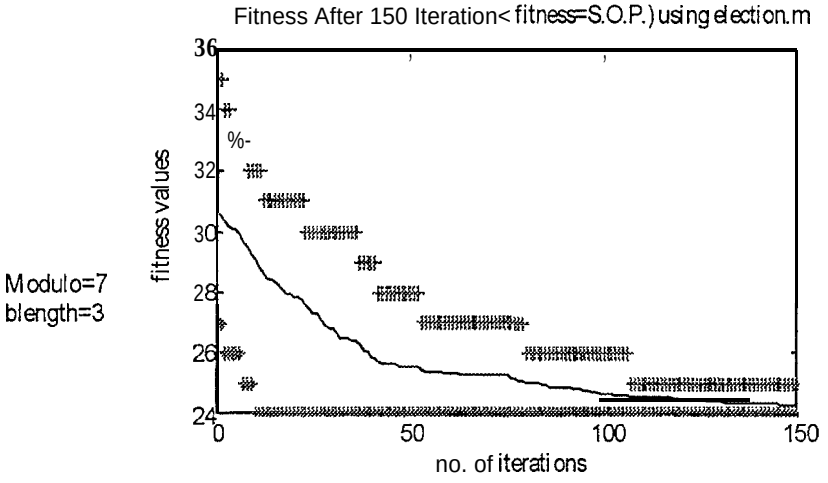


Figure 4 Example of **SSGA for modulo** 7.

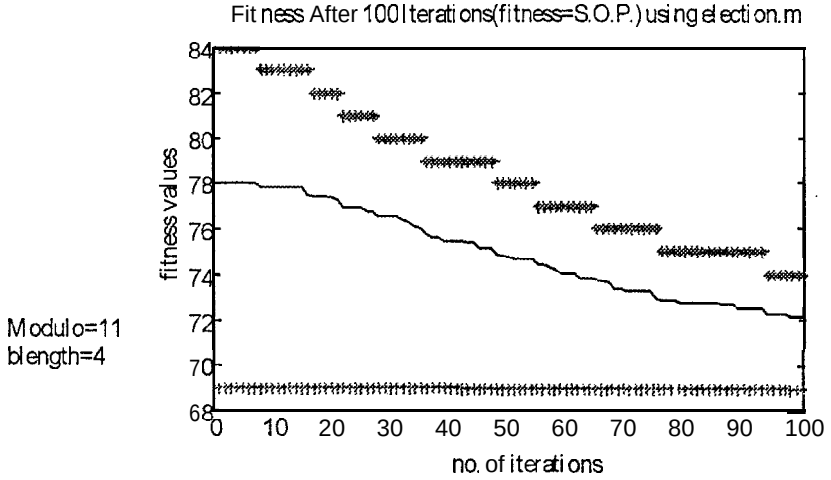


Figure 5 Example of SSGA for **modulo** 11.

high as well. For modulo 7 and 11, our crossover and mutation probabilities were reduced to $p_c = 0.75$ and $p_m = 0.01$, respectively. In Figure 4, we see one example of how the genetic algorithm has evaluated the modulo 7 adder. The SSGA minimized the fitness value to 24. In Figure 5, the SSGA converged in the first iteration to the fitness values of 69. For modulo 11, initial runs of the genetic algorithm have found that the string $\{2,5,8,10, 12\}(13, 9,0,11,7,15, 4,1,14, 3,6)$ yielded a fitness of $\psi = 69$. There is a penalty function included in the genetic algorithm. The function is as follows:

Penalty=2^(b length)+(pwr(1+modulo)). Simulations are continuously being done for modulo 11, 13, 17, 19, 23, 29, and 31 adders,

IV. CONCLUSION

As stated earlier, results for the modulo 11 through modulo 31 adders are being produced at this time. As seen in this paper, the number of logic gates has been reduced by 22°/0 for modulo 5 adder, by 33°/0 for modulo 7 adder and by 10% for modulo 11 adder. We expect our future results to be as promising as the results presented above.

A method to reduce the hardware necessary to build a fast fuzzy controller has been formulated and proved. This is the first of many steps in the process towards building a physical model of the controller; Procedures are in place to design the Digital-to-CRT, CRT-to-Digital, Analog-to-CRT and CRT-to-analog converters that are also needed for this problem. We plan to continue working in this area and present results in future publications.

V. REFERENCES

[1] Vainstein, F. and C.V. Curtis, "Acceleration of a Fuzzy Controller Using the Chinese Remainder Theorem," Proceedings of the NASA URC-TC '97, pp.747-752, 1997.

[2] Vainstein, F. and C.V. Curtis, "Acceleration of a Fuzzy Controller by an Algebraic Method," Proceeding of the Second International Symposium on Soft Computing for Industry 1998.

[3] A. Homaifar and E. McCormick, "Simultaneous Design of Membership Functions and Rule Sets for Fuzzy Controllers Using Genetic Algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 3,no.2, pp. 129-139, May 1995.

[4] J. L. Hennessey and D. A. Patterson, *Computer architecture: a quantitative approach.* San Francisco: Morgan Kaufmann, 1995.

[5] L. L. Dornhoff and F. E. Hohn, *Applied Modern Algebra. New* York: Macmillan, 1978.

[6] R. Rudell, A. Sangiovanni-Vincentelli,"Espresso-MV: Algorithms for Multiple-Valued Logic , Minimization," *Proc. Cust. Int. Circ. Conf.*, Portland, May 1985.

Measurements of the plasma parameters and low frequency oscillations in the Fisk Plasma Source

Edward Thomas, Jr.,[a] Kent Wallace, Gregory Lampkin, and Michael Watson

NASA/Fisk University Center for Photonic Materials and Devices
Department of Physics, Fisk University
Nashville, TN 37208
[a] Contact: etjr@dubois.fisk.edu

A new plasma device, the Fisk Plasma Source (FPS), has been developed at Fisk University. This plasma device is used to study the physics of low temperature plasmas and plasma-materials interactions. The FPS device is a stainless steel vacuum 6-way cross vacuum vessel with at 10-inch inner diameter. Low temperature argon plasmas are generated using DC glow discharge and thermionic filament techniques. Spatial profiles of the plasma density, plasma potential, and electron temperature are measured using Langmuir probes. We present initial experimental measurements of density and temperature profiles in the FPS device. Experimental and theoretical studies of low frequency oscillations observed in the FPS device are also presented.

1. Introduction

The application of plasma technologies in the production of semiconductor devices has grown dramatically in the past decade. Plasmas are commonly used in the preparation and cleaning of surfaces,' etching of electronic components,[2] and low energy ion implantation.[3] One of the primary benefits of plasma processing, as compared to chemical processing techniques, is the minimization of waste products. 1 n this manner, plasmas offer an economically and environmentally attractive method of production.[4] In spite of the widespread application of plasma processing technologies, fundamental understanding of plasma-material interactions as well as the physics of the plasma itself remains elusive. This paper focuses on the characterization of plasmas used in processing experiments.

Most models of' plasma behavior assume that the plasma is sufficient y ionized so that the interaction between the plasma ions and the remaining neutral particles is negligible. When the plasma parameters are measured using probes, this condition is summarized by:

$$\lambda_D << a << \lambda_{mfp} \quad (1)$$

where λ_D is the Debye shielding length, a is the scale length of' a probe in the plasma, typical] y the radius, and λ_{mfp} is the mean free path. However, most plasma processing system are operated at pressures ranging from $p_0 = 1$ to 500 mTorr with ionization fractions $n/n_0 < 0.1\%$; where n is the plasma density and n_0 is the neutral particle density. In these eases, the relationship between the Debye length, probe scale length, and the mean free path will become:

$$\lambda_{mfp} < \lambda_D \sim a \quad (2)$$

Consequently, the presence of the neutral particles has a significant impact on the plasma parameters and the increases the difficulty in understanding the underlying physical mechanisms that govern the behavior of the plasma.

The Plasma Sciences Laboratory is a newly established research effort in the Fisk University Center for Photonic Materials and Devices (CPMD). This laboratory supports the development of new materials and new semiconductor devices by providing plasma processing capabilities to the CPMD. However, the primary focus of the laboratory is the development of a more fundamental understanding of plasma transport and plasma parameter control.

This paper discusses the initial characterization of argon plasmas produced by a DC glow discharge technique, Measurements of the spatial profiles of the plasma density and electron temperature are presented. In the course of these characterization experiments, a regime of low frequency ($f < 1$ kHz) plasma oscillations was observed. A characterization of these plasma oscillations is also presented.

2. Facilities

The centerpiece of the Plasma Sciences Laboratory is the Fisk Plasma Source (FPS). The FPS device is a highly flexible, multiple configuration plasma device that produces moderate density, low temperature plasmas using DC glow discharge, thermionic filament, and, in the near future, RF/microwave techniques, The plasma source has

been operational at Fisk University since Summer, 1997.[5] All of the measurements presented i n this paper are in plasmas produced using the DC glow discharge technique.

The major hardware component of the FPS is a stainless steel 6-way cross vacuum vessel. The vacuum vessel has a large volume; it has a 10" inner diameter with 13.5" - ASA 8 flanges and a length of 20" to 24" along each axis. The large volume combined with the low pumping speed of the current vacuum system currently limits the base pressure in the FPS to a minimum of 1 to 5 mTorr. Upgrades to the vacuum systems over the next six months are expected to reduce the base pressure to $p_0 \sim 10^{-5}$ to 1 0^{-6} torr. A schematic of the FPS device is shown in Figure 1.

To aid in the confinement of the plasma, two electromagnets are wound directly onto the vacuum vessel walls to produce a mirror magnetic field with a peak field strength of 1.50 to 200 Gauss within the vacuum vessel. Models of the magnetic structure, as well as simulations of single particle orbits, in the FPS device are performed using a version of the Integrable Field Torsatron (IFT) code from Auburn University that has been adapted to model linear plasma devices.[6,7] The code will be used to utter the magnetic configuration of the FPS device in order to Facilitate spatial control over the density and temperature profiles of the plasma.

Measurements of plasma parameter on the FPS have been made using Langmuir probes and by optical emission spectroscopy. Preliminary experiments have been performed to demonstrate that the FPS device can be used for materials modification. These plasma processing results are discussed in the paper by Lampkin, *et. al.* in these proceedings.

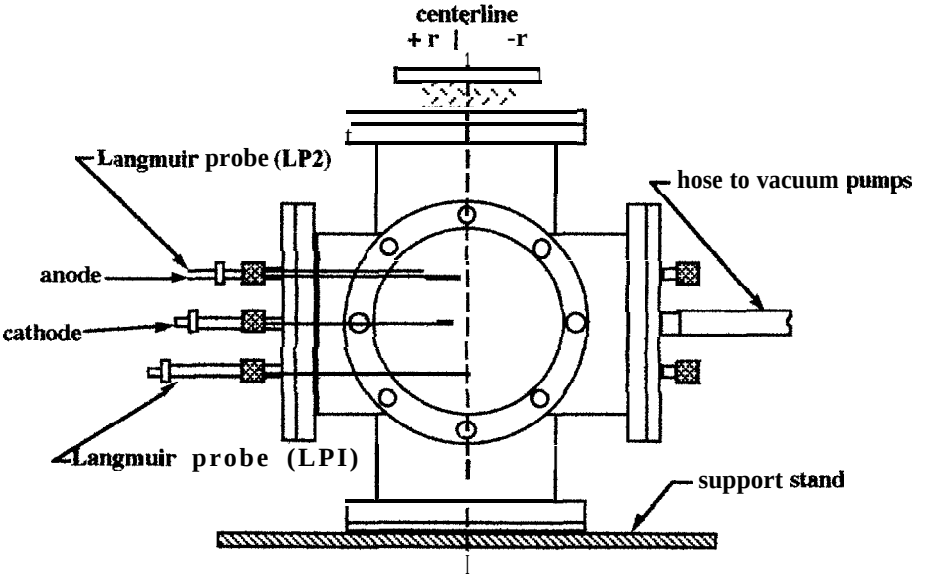


Figure 1: Schematic of the Fisk Plasma Source

3. Plasma parameter measurements

Spatial profiles of the plasma density and temperature arc measured using Langmuir probes. Langmuir probes are one of the oldest and simplest techniques used to diagnose a low temperature, moderate density plasma.[8] A Langmuir probe simply consist of a wire inserted into a plasma and connected, through a power supply, to ground. A bias voltage (V) is placed on the wire and the current (I) that flows to/from ground is recorded. This current-voltage (I-V) trace is a non-linear curve that is analyzed at each radial and vertical position to obtain the plasma density, electron temperature, and plasma potential. However, many factors, such as the ratio of the ion to electron temperature, the presence of a large neutral particle population (i.e., high collisionallity), the size of the probe, etc., contribute to the Langmuir probe I-V trace and make the interpretation of the plasma parameters difficult.

In these experiments on the FPS device, the Langmuir probe labeled LPI (in Figure 1) is moved radially across the plasma. The vertical position of LP1 can be altered by changing the shape of the probe as shown in Figure 2. It is then possible to obtain a 2-dimensional profile of the density and temperature of the plasma. The results presented in this paper represent our on-going measurements of the 2-D plasma parameter profile.

The radial plasma density and electron temperature profile of an argon plasma, at $p_0 = 100$ mTorr, is *shown* in Figure 3. The anode is biased at $V_A = 168$ V and the cathode is biased at $V_C = -282$ V. Both the anode and cathode are located at a radial position of $r = 5.0 \pm 0.2$ cm. The LPI probe was moved radially across the plasma at

a vertical position 4.5 cm below the cathode: positive and negative radii are indicated in Figure 1. The peak density in the plasma is shown to be $n \sim 3.5 \times 10^{14}$ m^{-3}. The electron temperature is in the range 5 to 15 eV.

Because of the high neutral density present in the FPS system, there is a great deal of difficulty in interpreting the measurements taken by Langmuir probes. Recent discussions suggest that the effect of ion-neutral collisions have not been sufficiently incorporated into previous theories of Langmuir probes.[9] The values of the electron temperature and plasma density presented here are based upon a revision of probe theory that explictly includes the effects of collisions. The I-V Langmuir probe trace is interpreted according to the following equation.

$$I = enA_p (T_e/m_i)^{1/2} \{ (m_i/2\pi m_e)^{1/2} \exp[e(V-V_p)/T_e] - (\lambda_{mfp}/\lambda_D)^{1/2} \exp(-1/2) \} \quad (3)$$

where A_p is the crass-sectional area of the probe, V_p is the plasma potential, m_e is the electron mass. and m_i is the ion mass. The factor $(\lambda_{mfp}/\lambda_D)$ is the new factor in the probe theory that includes the effects of ion-neutral collisions and corrections for the estimated value of the plasma density and electron temperature.

Nonetheless. while the absolute values of the density and temperature may change, the spatial profiles are not expected to be substantially affected. The measured profiles are consistent with the fact that the anode and cathode are generating the plasma at -positive radii and that the density falls rapidly away from this region. We plan to measure the density and eletron temperature using a double Langmuir probe since double probes arc generally less sensitive to neutral effects than are single Langmuir probes.

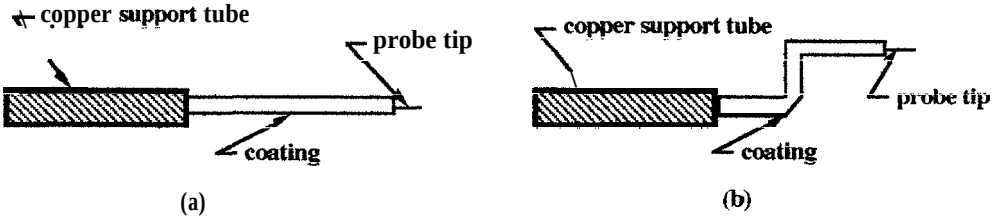


Figure 2: (a) Linear configuration of a Langmuir probe. (b) Vertical configuration of a Langmuir probe.

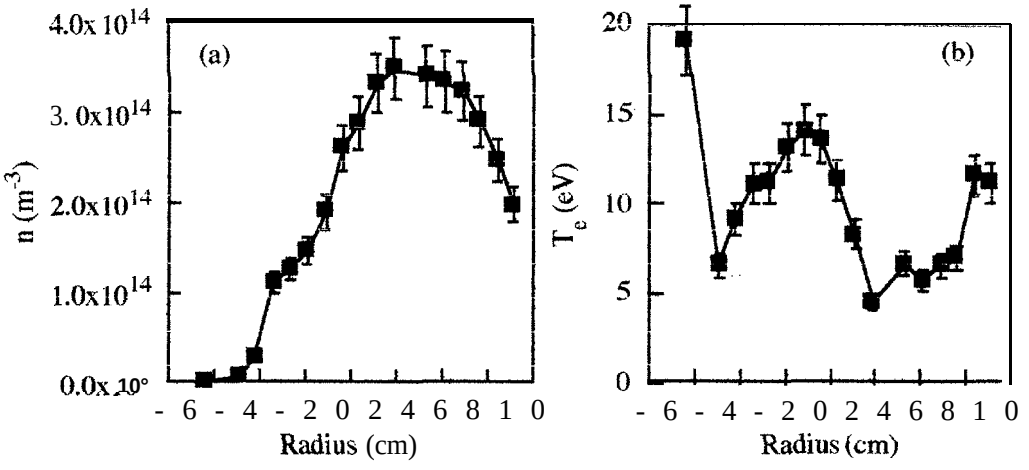


Figure 3: Measurements in an argon DC glow discharge plasma at a pressure of $p_0 = 100$ mTorr. (a) Radial profile of the plasma density. (b) Radial profile of the electron temperature.

4. Plasma oscillations

Because most plasmas are in a state of unstable equilibrium. microscopic disturbances of the plasma can lead to large scale plasma oscillations provided there is sufficient free energy in the system to sustain the growth of the instability. In many applications. the appearance of plasma oscillations significantly reduces control over the plasma parameters and can cause substantial damage to plasma exposed surfaces due to reductions in particle confinement, [10]

In experiments on the FPS device, the plasma oscillations were first observed in DC glow discharge argon plasma while potential difference between the anode and the cathode was increased. The oscillations were measured by both Langmuir probes LP1 and LP2. The current flowing from the probe tip to ground was measured as the

voltage drop across the 1 MΩ input impedance of the oscilloscope. The probe signal was displayed both on the oscilloscope and captured digitally on a computer.

These oscillations have very low frequencies, typically under 1 kHz, and have very large amplitudes as shown in Figure 4. In Figure 4 the voltage measured by the probe LP 1 is shown as a function of time. Both curves were measured in an argon plasma at p_0 =55 mTorr and the bias voltage on the cathode was held fixed at V_C = -284 V. In the lower, heavier curve the bias voltage on the anode was V_A = 170 V. At this potential difference between the cathode and anode, the plasma was fairly quiescent. In the upper curve. the bias voltage on the anode was raised to V_A = 235 V. Large amplitude oscillations at a frequency off ~ 100 Hz are observed.

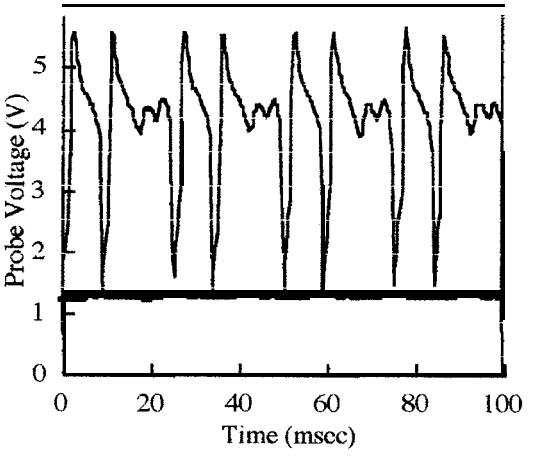


Figure 4: The voltage measured on a Langmuir probe as a function of time. The lower, darker curve shows a quiescent plasma state. Here V_A = 170 V and V_C = -284 V. The upper. lighter curve measured low frequency, large amplitude oscillations in the plasma. Here the cathode bias voltage was fixed and the anode bias voltage was raised to V_A = 235 v.

Because of the low frequency of the plasma oscillations, it was necessary to ensure that the oscillations were independent of fluctuations in the power supplies. A comparison of the voltage measured by probe LP1 and the AC component of the bias voltage on the anode are shown in Figure 5(a). In this experiment, an argon plasma was generated at p_0 =55 mTorr with a cathode bias of V_C = -301 V and an anode bias of V_A = 291 V. The thicker line is the measurement of the power supply voltage; the thinner line is the measurement taken by the Langmuir probe. A fast Fourier transform (FFT) was performed on the measured voltage signals to obtain the frequency components present in the signal. The FFT's of both the measured probe and power supply voltages are shown in Figure 5(b). Again, the thicker line is the measurement of the power supply voltage; the thinner line is the measurement taken by the Langmuir probe. A comparison of the two FFT's demonstrates that the dominant frequency components in the power supply voltage signal and the dominant frequency components in the Langmuir probe voltage signal are not the same. This suggests that the observed plasma oscillations are not due to fluctuations in the power supply.

In an effort to understand the observed plasma oscillations. a simple model has been developed to describe the source of free energy that drives the oscillations. We consider the energy gained by a charged particle in the plasma between collisions. The work done on the particle by the electric field (i.e.. the electric field between the anode and cathode) is given by:

$$W = qE\lambda_{mfp} \tag{4}$$

where W is the work, q is the charge of the particle, and E is the magnitude of the electric field. This work done on the particle is the energy that is gained between collisions. The mean free path is given by:

$$\lambda_{mfp} = 1/n_0\sigma = kT_0/p_0\sigma \tag{5}$$

where n_0 is the neutral particle density (given by kT_0/p_0), p_0 is the neutral pressure, T_0 is the temperature of the neutral particle, typical] y taken to be room temperature. k is Boltzmann's constant, and σ is the collision cross-section. Therefore, the total energy, U, gained by particles between collisions is given by:

$$U = (qkT_0/\sigma)\,(E/p_0) \tag{6}$$

It is shown in Equation 6 that the energy gained by a particle between collisions is proportional to the electric field strength and inversely proportional to the pressure. Experiments were performed to test the validity of this model.

In one experiment, an argon DC glow discharge plasma was generated at a pressure of $p_0 = 55$ mTorr and the cathode voltage was held fixed at $V_C = -282$ V. The anode voltage was raised from 230 V to 315 V. The fluctuation level of the plasma is characterized by: $dV = (V - \langle V_{DC} \rangle)/\langle V_{DC} \rangle$ where $\langle V_{DC} \rangle$ is the average DC voltage. The plasma fluctuation level as a function of the anode voltage is shown in Figure 6(a). As suggested by the previous results, as the potential difference between the anode and cathode is increased (i.e., electric field strength), the fluctuation level in the plasma rises. This measurement is generally consistent with the prediction made in the model.

In another experiment. the electric field in the glow discharge plasma was fixed while the neutral pressure was varied from 50 to 100 mTorr. In this experiment. the cathode voltage was $V_C = -288$ V and the anode voltage was $V_A = 240$ V. The plasma fluctuation level as a function of the neutral pressure is shown in Figure 6(b). It is shown that as the pressure decreases. the fluctuation level of the plasma increases. Again, this is generally consistent with the predictions of the model.

However, it is observed that the fluctuation level does not continually rise with increasing electric field or with decreasing pressure. Rather, it reaches a maximum value then begins to decline. This suggests that there is a resonant mechanism that leads to the formation of the plasma oscillations. Resonant physical phenomena that occur in high neutral pressure plasmas such as moving, striations[11] or global ionization instabilities[12] may he responsible for the observed oscillations.

We have begun to develop a numerical model for the oscillations using a combination of particle-in-cell (PIC) and Monte Carlo techniques. in these models. it is the interaction between the electrons and the background neutral particles that gives rise to the plasma oscillations. The goal is to develop a self-consistent model for the plasma oscillations based upon fluctuations in the neutral density.

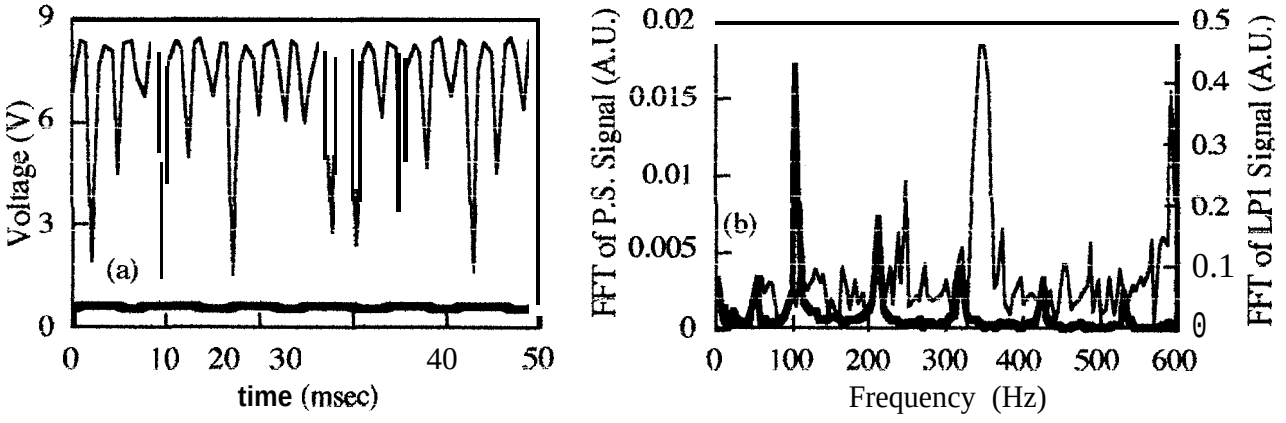


Figure 5: (a) Voltage vs. time plot for Langmuir probe LPI (thinner line) and the AC component of the power supply (thicker line). (b) FFT of the LPI voltage signal (thinner line) and the power supply voltage signal (thicker line) as a function of frequency. The vertical *axes are* in arbitrary units.

5. Conclusions

The Fisk Plasma Source is a newly developed. flexible plasma device. Measurements using Langmuir probes, evaluated using a standard probe theories, yield electron temperatures T_e -5 to 15 eV and moderate densities $n \sim 1013$ m^{-3} in argon plasmas are pressures ranging from $p_0 = 50$ to 100 mTorr. However, a new! y developed model for interpreting measurements from Langmuir probes in high neutral pressure plasma systems suggest that the while the measurements of the density and temperature profiles are correct, the values of the density and temperature need to be reevaluated. Work on the plasma parameter measurements is on-going.

Low frequency plasma oscillations are observed in DC glow discharge argon plasmas The oscillations are shown to be independent of power supply fluctuations. A theoretical model has been developed to describe the source of free energy that drives the oscillations. Experimental measurements of the plasma fluctuation level as a function of pressure and electric field strength are generally consistent with this model. However, an apparent resonant feature in the experimental data is not described by the model. To describe this resonant feature, PIC and Monte Carlo simulations are being developed.

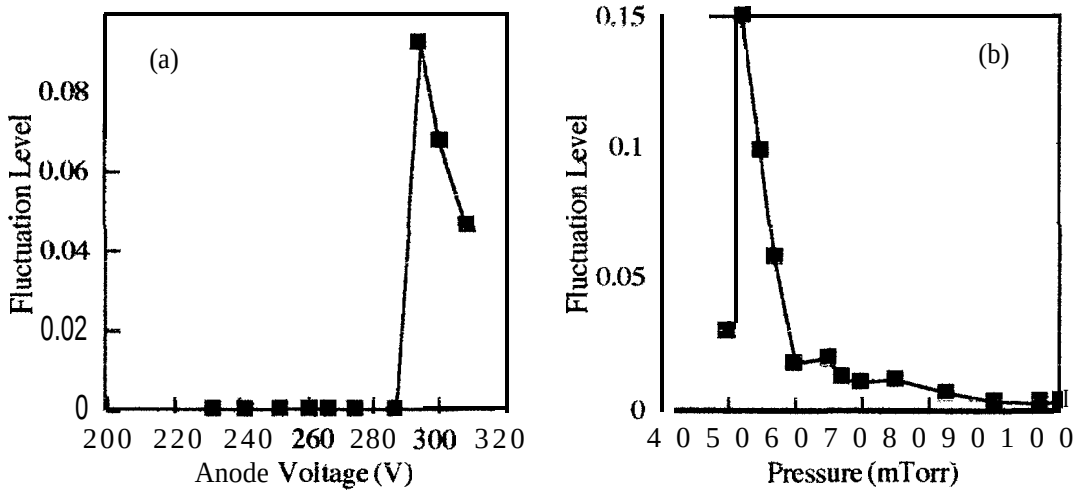


Figure 6: (a) Voltage vs. time plot for Langmuir probe LP1 (thinner line) and the AC component of the power supply (thicker line). (b) FFT of the LP1 voltage signal (thinner line) and the power supply voltage signal (thicker line) as a function of frequency. The vertical axes are in arbitrary units.

Acknowledgments

The authors of this paper would like to acknowledge the support of the members of the Fisk University Center for Photonic Materials and Devices without whom this work would have been impossible. The CPMD is supported by NASA Grant No. NCC-8-133.

[1] Z. Lu, M. T. Schmidt, R. M. Osgood, Jr., W. M. Holber, D. V. Podlesnik, J. Vat. Sci. Tech. A. 9 1040 (1991).
[2] J. Asmussen, J. Vat. Sci. Tech. A., 7883 (1989).
[3] J. Mantese, I. G. Brown, N. W. Cheung, and G. A. Collins, MRS Bulletin, 21, No. 8, 52 (1996).
[4] G. Collins and D. J. Rej, MRS Bulletin, 8 26(1996).
[5] for further information, refer to the following WWW page http://www.fisk.edu/~etjr/psl.html
[6] J. R. Cat-y, Phys. Fluids, 27, 119 (1984).
[7] J. R. Cary and J. D. Hanson, Phys. Fluids, 27,767 (1984).
[8] H. H. Mott-Smith and Irving Langmuir, Phys. Review, 28, 727 (1926),
[9] N. Hershkowitz, Personal communications, November, 1997.
[10] J Reece Roth, *Industrial Plasma Engineering, Vol 1.*, (Inst. of Physics Publishing, Bristol, 1995), Chap. 9, p. 289.
[11] *ibid*, p.308.
[12] B. Song, N. D'Angelo, and R. L. Merlino, J. Phys. D: Appl. Phys., 24, 1789 (1991).

NOVEL LINEAR AND NONLINEAR OSCILLATOR ARCHITECTURES

Catharine Rice[1] *and J. Ramirez-Angulo*[2]

[1]NASA Center for Autonomous Control Engineering
Department of Engineering
New Mexico Highlands University
Las Vegas, NM 87701
[2]Klipsch School of Electrical and Computer Engineering
New Mexico State University
Las Cruces, NM 88003

I. Introduction.

Operational transconductance amplifiers can be used to implement oscillators with control over a wide range of oscillating frequencies. An advantage of these oscillators, besides programmability, is that they can operate at much higher frequencies than op-amp based oscillators. In this paper we report on novel linear and nonlinear oscillator structures using operational transconductance amplifiers. Two of the structures using a novel approach denoted OTA-OP-AMP approach have not been reported previously.

II. OTA based Oscillators

The building blocks of Figs. 1 reported in [1] can be used for the implementation of linear and linear oscillators which we discuss next:

11.1 Nonlinear oscillators:

The diagram of Fig. 2a shows the scheme of a triangular/square wave oscillator using an inverting Schmitt trigger and an inverting integrator. Fig. 2b shows the OTA implementation, Fig. 2c the OTA op-amp implementation. The oscillation frequency and the amplitude of the square and triangular waveforms are given by:

$$V_{amp}=I_{abc}R \text{ where } R=R1+R2$$

and

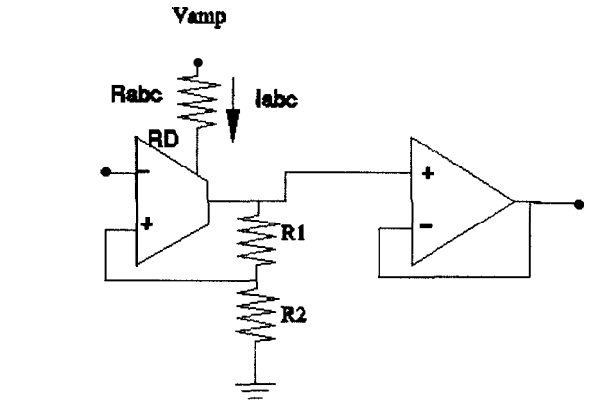
$$f_{osc}=G_{meff}/(2\pi RCI_{abc})$$

respectively

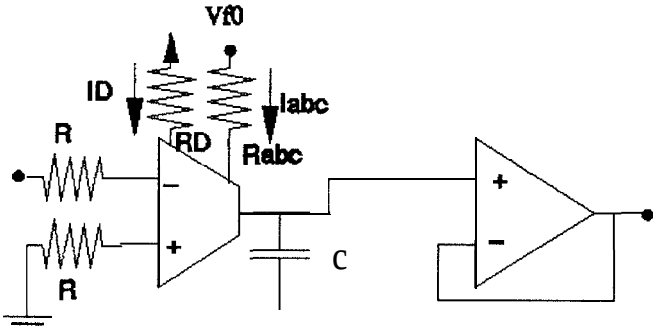
11.2 Linear quadrature oscillators:

The diagram of Fig. 3a shows the scheme of a sinusoidal oscillator using a two integrator loop. Fig. 3b and 3c show implementations using OTA and OTA-op amp integrators respectively. The circuits include Zener diodes for amplitude limiting purposes and in both cases the frequency of oscillation is given by

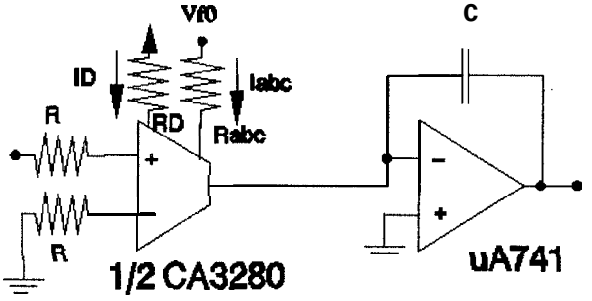
$f_{osc}=(1/2\pi)(G_{meff}/C)$ with G_{meff} *given* by (l).



(a)



(b)



(c)

Figure 1. Building blocks of OTA based voltage controlled linear and nonlinear oscillators: (a) Schmitt Trigger (b) Conventional OTA integrator (c) Novel OTA-op amp integrator

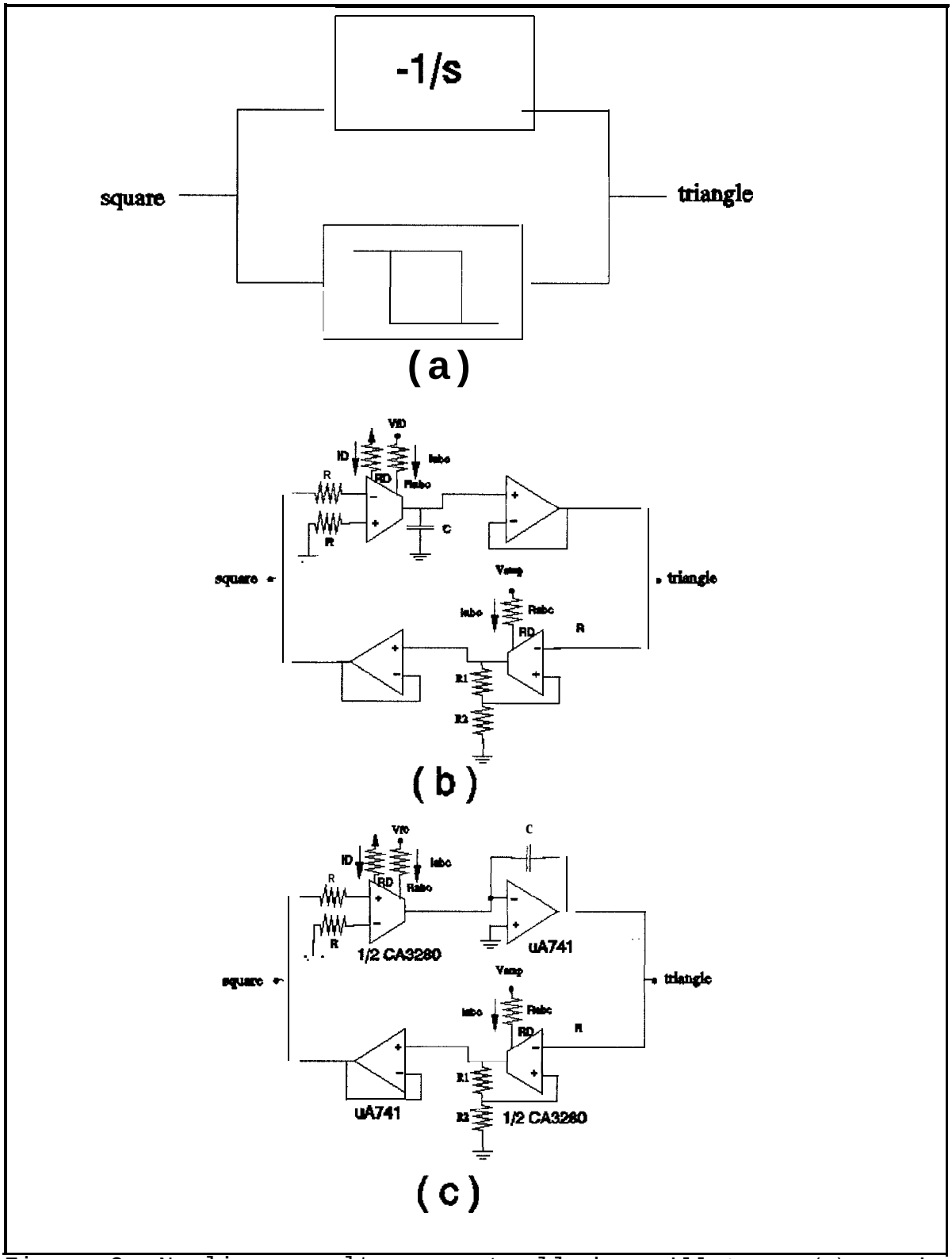


Figure 2 Nonlinear voltage controlled oscillators (a) Basic scheme (b) Conventional OTA implementation (c) Novel OTA op-amp implementation

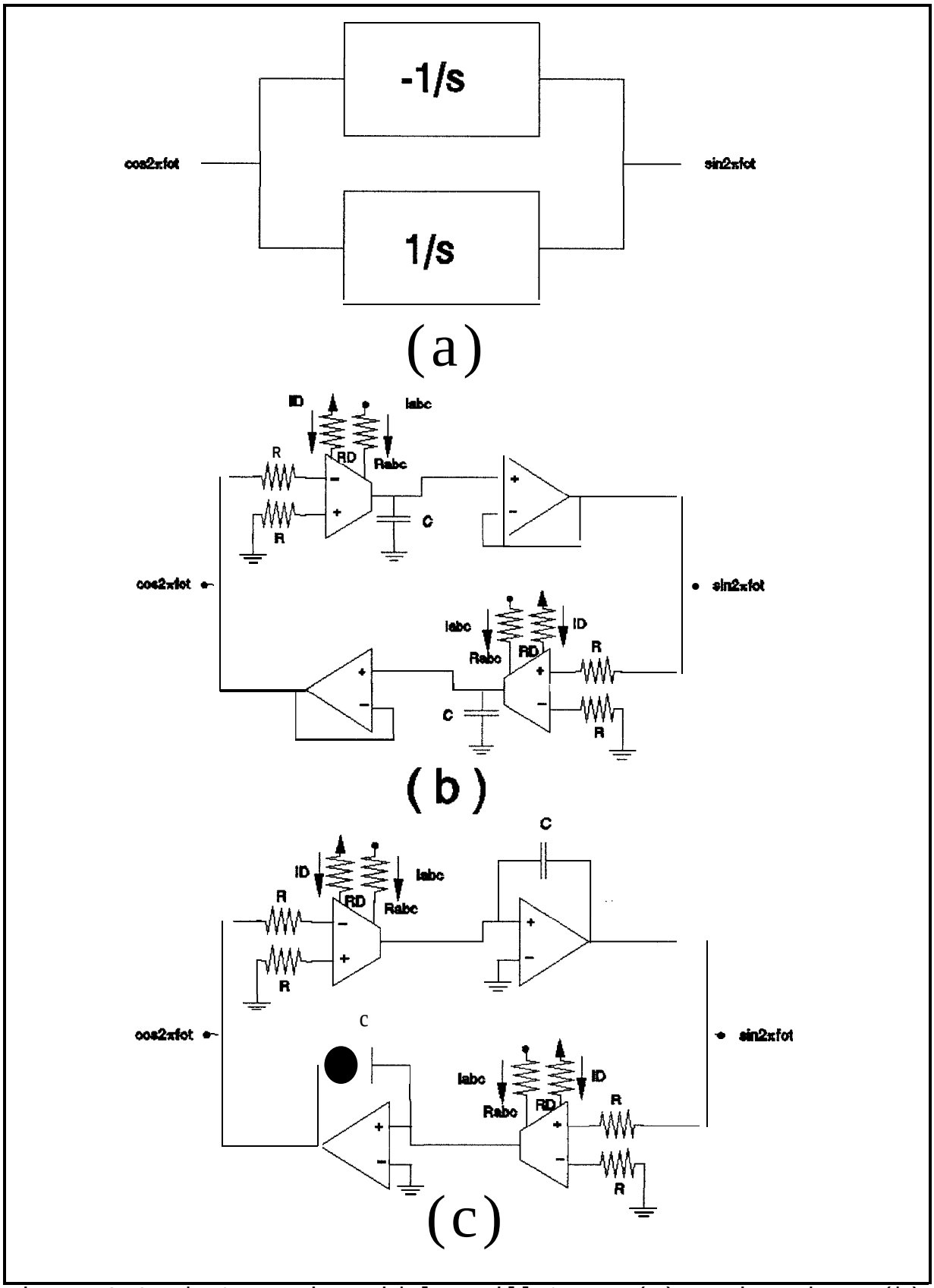


Figure 3 Quadrature sinusoidal oscillators: (a) Basic scheme (b) Conventional OTA implementation (c) Novel OTA-op amp implementation

III. Experimental results

Experimental results that verify the functionality of the proposed oscillator structures are available. These results verify the wide range programmability of the oscillation frequency as wells the operation well into the MHz range. These results were obtained from breadboard prototypes using low cost commercial OTAS and op-amps (CA3280 and LM741 respectively) .

References

[1] "Building Blocks of Linear and Nonlinear circuits Based on Operational Transconductance Amplifiers," Anthony Esquibel, Spyros Andreou and J. Ramirez-Angulo, submitted *to* the 1998 NASA ACE conference.

[2] "Low Voltage High Frequency Continuous time Filters Based on Simple Transconductors and Miller Integrators", J. Ramirez-Angulo, *Analog* ***Integrated Circuits and Signs Processing,*** April 1997.

[3] "Large f0Q Second Order Filters Using Multiple Output OTAs", Jaime Ramirez-Angulo and E. Sanchez-Sinencio, ***IEEE Transactions on Circuits and Systems,*** Vol. 41, No. 9, pp. 587-592, September 1994

Application of Model Identification in Robust Controller Design for Power System Damping NASA-Autonomous Control Engineering Center

Fereshteh Fatehi
Electronics & Computer Technology Department, NC A& T State University
fatehi@aurora.ncat.edu

Abdolah Homaifar
Electrical Engineering Department, NC A& T State University, Greensboro, NC
homaifar@ncat.edu

ABSTRACT

This paper presents recent results on investigation into the use of model identification methods for the design of robust controllers to enhance the damping of interarea oscillations in multimachine power systems. This method of model identification and reduction is called Prony model identification method. The linear reduced model which can be obtained by this method will be used to design the controller. μ- *synthesis*, which produces a robust controller satisfactory for a range of operating conditions, is suggested as a control design technique.

KEYWORDS: power system, model identification method, robust control, fuzzy rules

I. INTRODUCTION

Power systems commonly exhibit low-frequency electromechanical oscillations particularly when operated under conditions where large power transfers are being made over long transmission lines. With utilities increasing power exchanges over a fixed network, the **use** of new equipment in the transmission system to aid the damping of these oscillations is being seriously considered.

Recently, a considerable amount of effort had been dedicated to research into the use of power electronics to enable fast switching of the series compensation, resulting in improved system damping, stability, and increased power flows. It is well known that Flexible AC Transmission Systems (FACTS) controllers can be utilized with the proper modulation, to dampen the oscillatory local and inter-area electromechanical modes in multi-machine power system [1,2].

This paper presents recent results on investigation into the use of model identification methods for the design of robust controllers to enhance the damping of interarea oscillations in multimachine power systems. This method of model identification and reduction which is described in this paper is called Prony model identification method. The linear reduced model which can be obtained by this method will be used to design the controller. μ- *synthesis*, which produces a robust controller satisfactory for a range of operating conditions, is suggested as a control design technique [3].

The paper is organized as follow. In section 2, Prony model identification and reduction method is described. μ- *synthesis* control design is described in section 3 . The power system model and test results are presented in section 4 and conclusion in section 5.

II. PRONY MODEL IDENTIFICATION METHOD

This section explains the application of model identification in power systems using deterministic input signals and recently developed signal processing techniques.

The order of any transfer function in a large power system could be as large as the number of states that are needed to model the system. For most large systems this would mean many thousands of states. However, most of the system modes cannot be observed in any one signal in the system. This situation allows one to identify effective reduced-order transfer function between any input and output signals. An input signal applied to a power system will usually excite many modes which are then exhibited to differing degree in different output signals [4].

The transfer function between the input and output signals contains the poles (or eigenvalues) and zeros which are acted upon by the input signal to produce an output signal, A transfer function is usually written as a ratio of polynomials but can be also expressed in parallel form as a sum of residues over first-order poles. The conversion from the rational polynomial form to the parallel form is the well known process of partial fraction expansion.

This method of model identification and reduction use in this paper, is called Prony Transfer Function Identification (TFI) method. The method is very similar to the direct ARMA method in the way it uses the same numerical procedures to solve for an over-determined set of equations. The main differences is in how the equations are set up.

In the Prony TFI method the equations are set up in pole residue form. This is also sometimes referred to as parallel **form.** Given an s-plane signal description, expressed as a ratio of polynomials where the denominator is of higher order than the numerator, the parallel form is obtained by partial fraction expansion. This can be expressed as

$$\hat{G}(s) = \sum_{i=1}^{n} \frac{\hat{R}_i}{s - p_i} \qquad (1)$$

A corresponding zero-order-hold z-plane representation is G(z):

$$G(z) = \sum_{i=1}^{n} \frac{R_i}{z - \lambda_i} \qquad (2)$$

where

$$R_i = \frac{\hat{R}_i}{p_i}(\lambda_i - 1) \qquad (3)$$

and T is a constant sampling period. $\hat{G}(s)$ and $G(z)$ are system transfer functions, the p_i's are s-plane eigenvalues, the λ_i 's are z-plane eigenvalues, and R,'s and R_i 's are transfer function residues. Finding a linear model for a given system means to solve for $\hat{R}_i$'s and *pi's* in (l).

Suppose the input is applied at time k= 1 up to k=d and the time k=d+1, is the first sample value for which the input is no longer being applied. When no input is being applied, then the signal "ring down" can be described by a difference equation model where the output is dependent only on past outputs and not on any inputs. This type of model is commonly referred to as an auto regressive model or AR model. The Prony TFI method uses the AR method to solve the first set of over determined equations. The results are coefficients of difference equations. Once these coefficients are obtained the characteristics polynomial associated with the difference equations can be rooted to get the transfer function eigenvalues, the λ_i 's. Then the solution can be written in the form

$$y(k) = \sum_{i=1}^{n} \overline{R}_i \lambda_i^k \qquad (4)$$

where the λ_i 's are the eigenvalues of the discrete transfer function G(z) and the $\overline{R}_i$'s are the signal residues of the ring down portion of the output signal. These residues are a function of the input signal as

will be discussed shortly. The residues are obtained by setting up a second set of over determined equations to solve for $\overline{R}_i$'s.

At this point, knowledge of the input signal can be utilized to calculate transfer function residues from the signal residues which have been identified. In order to find the system transfer fiction, we need to know the poles and the zeros of the system, or the poles and residues associated with the impulse response of the system. If the input signal is restricted to be block pulses whose heights and duration are known then the determination of the transfer function is simplified. For the sake of clarity we can take for an example the case where the input is a single rectangular pulse of height U_0 and d+ 1 samples in duration, The z-transform of this input is

$$U(z) = U_0 \sum_{j=0}^{d} z^{-j} \qquad (5)$$

Applying this input to a linear transfer function yields the output signal Y(z)

$$Y(z) = G(Z)\,U(z) = U(z) \sum_{i=1}^{n} \left[\frac{R_i}{z - \lambda_i}\right] \qquad (6)$$

where the right hand part of this expression contains the transfer function in partial fraction form. The R_i 's in (6) are transfer function residues which can be mathematically determined from the output signal residues, which are the $\overline{R}_i$'s in (4). Combining equation (6) with the previous expression for U(z) we can get

$$Y(z) = U_0 \sum_{j=0}^{j=d} z^{-j} \sum_{i=1}^{n} \left[\frac{R_i}{z - \lambda_i}\right] \qquad (7)$$

The inverse z-transform gives the following sampled output signal

$$y(k) = U_0 \sum_{j=0}^{d} \sum_{i=1}^{n} R_i (\lambda_i)^{(k-j-1)} U_s^{(k-j-1)} \qquad (8)$$

where u_s is the unit step function. This expression is actually a discrete time convolution of the sampled input signal with the system impulse response. When the input is no longer being applied to the system then the output can be expressed as

$$y(k) = U_0 \sum_{i=1}^{n} R_i (\lambda_i)^k \sum_{j=0}^{d} (\lambda_j)^{-(j+1)} \text{ for k>0} \qquad (9)$$

The last summation in this expression can be rearranged to yield

$$\text{y(k)} = \sum_{i=1}^{n} \overline{R}_i (\lambda_i)^k \qquad \text{for k>0} \qquad \textbf{(lo)}$$

This last expression illustrates the difference between transfer function residues R_i in (9) and the signal residues $\overline{R}_i$ given by (4). The relationship for this case is

$$R_i = \frac{\overline{R}_i}{U_0} \left(\frac{-1 + \lambda_i}{1 - \lambda_i^{-(d+1)}}\right) \qquad (11)$$

Thus by using the information obtained from the ring-down signal description obtained by applying the basic Prony method and utilized known information about the input signal, it is possible to solve for the system transfer function between the input signal and the output of the system. The transfer function can be checked for accuracy by applying the same input to this model and calculating the resulting output and comparing this signal to the actual system output. If the calculated system response is very close to

that of the actual system response, this is a positive indication that the Prony method has successfully located the dominant modes of the system and the correct damping terms for these modes. It is also a good indication that the transfer function is accurate.

III. μ-*SYNTHESIS* CONTROL DESIGN

This section contains an approach to *p-synthesis* that can be used to design a controller. The problem is posed as a robust performance problem, with multiplicative plant uncertainty at the plant input and plant output weighted sensitivity function as the performance criterion. The reduced linear model which can be obtained by Prony model identification method is used to design the controller.

A diagram for the closed-loop system, that includes the feedback structure of the plant and controller, elements associated with the uncertainty models and performance objectives, are shown in Fig. 1. The dashed box represents the "true" power system model, with associated transfer function G. Inside the box is the nominal model of the system, which is the linear model that had been found using Prony TFI method (G_{nom}), and two elements, W_{del} and A $_G$ which parameterize the uncertainty in the model. This type of uncertainty is called multiplicative uncertainty at the plant input. The performance objective is that the transfer function from d to e be small, in the $\|\|_\infty$ sense, for all possible uncertainty transfer function G. The weighting function W_p is used to reflect the relative importance of various frequency ranges for which performance is desired. The approach for finding these weighting functions will be discussed in the later sections.

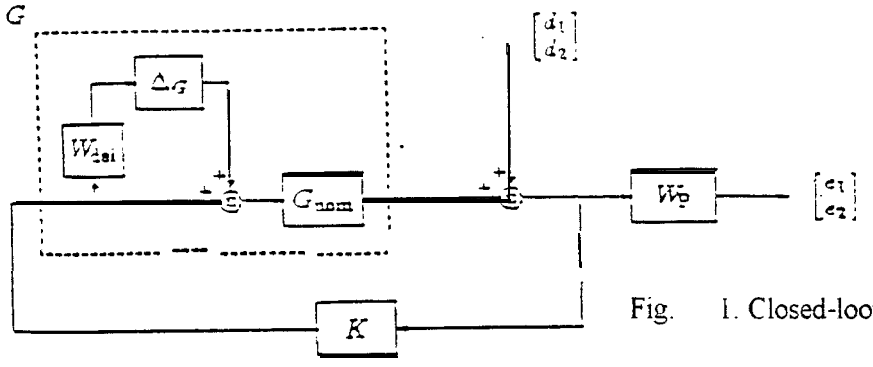


Fig. 1. Closed-loop System

The control design objective is to design a stabilizing controller K such that for all stable perturbations $\Delta_G(s)$, with $\|\Delta_G\| < 1$, the perturbed closed-loop system remains stable, and the perturbed weighted sensitivity transfer function,

$$S(\Delta_G) := W_p(I + G_{norm}(I + \Delta_G W_{del})K)^{-1} \qquad (12)$$

has $\|S(\Delta_G)\| \leq 1$ for all such perturbations. These mathematical objectives, which is called ***Robust Performance*** exactly fit in the *structured singular value* framework.

IV. POWER SYSTEM MODEL AND TEST RESULT

The 3-area, 6-machine power system [5] is presented in Fig.2. In this system, Machines 3 and 4 are exporting a total of 640 MW and Machines 5 and 6 are exporting a total of 610 MW. The exported power is consumed by the load L3. The analysis of this system with its models is performed using the Power System Tool Box [6].

There are five electromechanical modes of oscillations present in the system. Two Iow-frequency modes are the inter-area modes. Based on the mode shapes of the inter-area modes, we determine that

there are three coherent groups, Machines 1 and 2 form the first coherent group C, Machines 3 and 4 form the second coherent group B, and Machines 5 and 6 form the third coherent group A. The first inter-area mode (model) consists of the machines of Area B oscillating against the machines of Area A and C, and the second inter-area mode (mode2) consists of machines of Area A oscillating against the machines of Area C. The other three modes are the local modes of machine oscillations within the areas. Based on the system configuration, we insert a TCSC model in one of the tie lines between Buses 13 and 14 to enhance the damping of Mode 2.

The inter-area modes and their damping requirements are sensitive to system operating conditions. For robust design, the controller must be able to provide damping to all credible system conditions. The operating conditions are analyzed for variation of Z_{eq9-10} in the TCSC transmission path. The operating conditions represent the stiff, median and weak transmission strength configurations, with the power transfer between the areas remaining about the same.

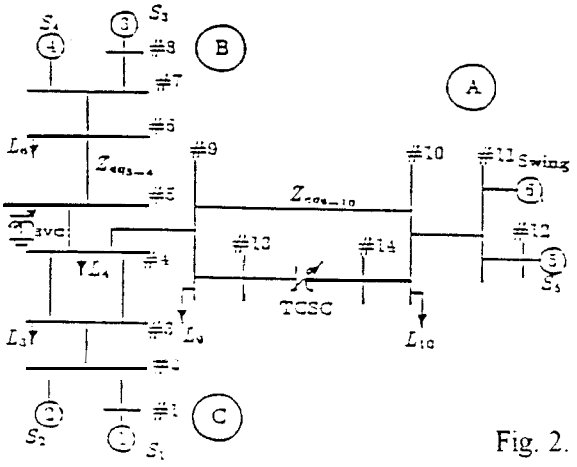


Fig. 2. Power System Model

in the following tests the Prony system identification method is used to identify a linear model for the power system. A rectangular pulse is applied to the reference input of the TCSC model (TCSC model consists of washout and lag blocks). The output signal is the relative speed signal between area A and C.

Figures 3-5 show the power system and 8'" order identified (linear) plant model outputs for three different operating conditions. Fig. 6 compares the frequency responses of the identified models for three different operating conditions. By considering the median model as the nominal model and comparing the frequency response of this model with the other models, we can find a weighting function for the uncertainty in the nominal model.

V. CONCLUSIONS

In this paper a new approach to robust controller design for TCSC's is presented. This approach attempts to dampen low frequency oscillations in multimachines power system at different operating conditions.

Prony Transfer Function Identification method is used to obtain a linear model of power system for use in controller design. The simulation results illustrate the good fit between the power system output and the linear models outputs for three different operating conditions.

The weighting fictions, which reflect the amount of uncertainty in the model, can be determined by comparing the frequency responses of these linear identified model. Fuzzy rules can be used to determine a weighting function which best reflects the uncertainty at each frequency in the nominal model. Future work will focus on weighting function identification and μ-*synthesis* design.

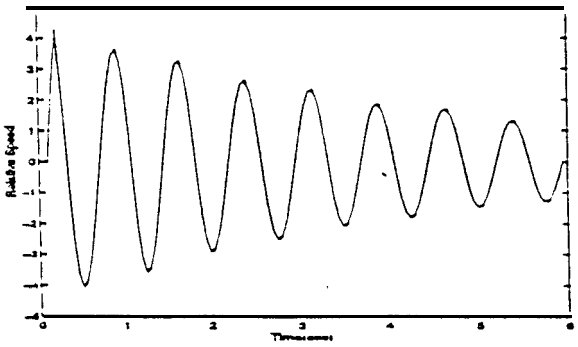


Fig. 3. Power system output (solid line) and identified model output (dash-dot) for median condition

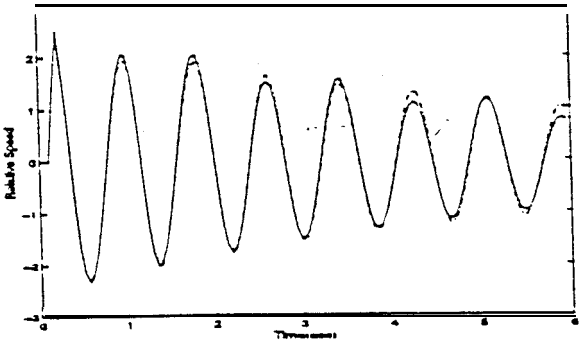


Fig. 4. Power system output (solid line) and identified model output (dash-dot) for stiff condition.

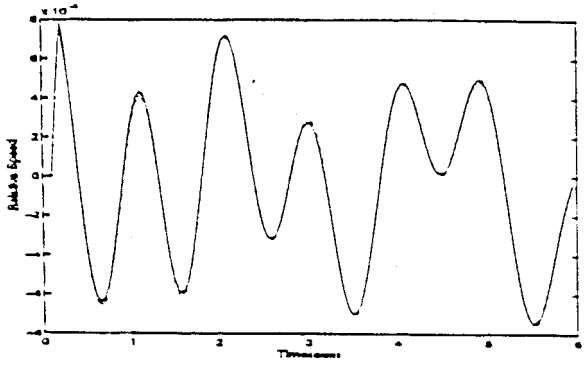


Fig. 5. Power system output (solid line) and identified model output (dash-dot) for weak condition.

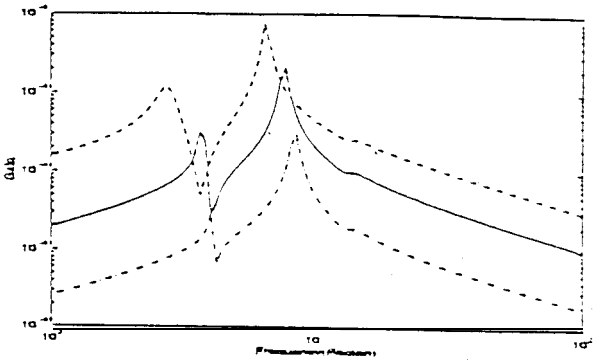


Fig. 6. Frequency responses for three different operating conditions.

Acknowledgment

The authors express their appreciation to Dr. G.N. Taranto permission for using his power system model. This work is partially funded by a grant from Autonomous Control Engineering Center ACE-48146. The authors wish to thank them for their financial support.

References

[1] R.J. Piwko, C.A. Wegner, S.J. Kinney, and J.D. Eden. "Subsynchronous Resonance Performance Test of The Slatt Thyristor-Controlled Series Capacitor," IEEE paper no. 95 SM 402-8 PWRD.

[2] Gary J. Balas, John C. Doyle, Keith Glover, Andy Packard, and Roy Smith. "μ-analysis and μ-Synthesis Tool Box," Mathworks, Inc.

[3] J.J. Paserba, N. W. Miller, E. Larsen, and R. Piwko. "A Thyristor Controlled Series Compensation Model For Power System Stability Analysis." IEEE paper no. 94 SM 476-2 PWRD.

[4] J. R. Smith, F. Fatehi, C. S. Woods, J. R. Hauer and D. T. Trudnowski. "Transfer Function Identification in Power System Applications." *IEEE Transaction on Power Systems*, Vol. T-PWRS-8, No. 3, pp. 1282-1290, Aug. 1993.

[5] G.N. Taranto, J.H. Chow, and H.A. Othman, "Robust Decentralized Control Design for Damping Power System Oscillations," Proc. 33rd IEEE Conf. on Decision and Control, Orlando, Florida, pp. 4080-4085, 1994.

[6] Joe H. Chow and Kwok W. Cheung, "A Toolbox for Power System Dynamics and Control." IEEE paper no. 92 WM 082-8 PWRS.

Preparation and Surface Analysis of a Fluorinated Amorphous Silicon for Photo-voltaic Device Application.

Hylton G. McWhinney, Dawn Burton and Thomas N. Fogarty
Prairie View A&M University

Abstract

Amorphous silicon films (a-Si:H) have been routinely deposited on a variety of substrates. Surface and interfacial studies were carried out with a PHI 5600 X-ray photo electron spectrometer. Co-deposition with fluorine yielded films having oxygen present as bulk oxide. The higher the fluorine content, the greater the amount of bulk oxygen observed. The presence of oxygen may be a contributing factor to inconsistent film properties of fluorine doped silicon materials, reportedelsewhere. A definite chemical interface between a layer containing fluorine and a layer made from pure silane has been delineated.

Introduction

Amorphous silicon alloy (a-Si:H) materials present several advantages in device applications. These include: (a) low cost methods to deposit large area, quality thin-film materials at low substrate temperatures on a wide variety of substrate shapes and composition; (b) adjustment of the optical band gap and electrical properties can be easily carried out to suit the need of the application; (c) potential use in stacked photo receptor devices. Poor materials properties include low charge carrier mobility and lifetime low deposition rates and photo degradation 1]. The latter is especially important in its application as solar cell energy sources.

Critical to the performance of these silicon devices is the high compressive stresses measured in these films (as high as 400 MPa) due to substrate mis-match. Evidence of ordering on crystalline substrate has been presented. Annealing the material after prolonged illumination appear to rejuvenate the material. Mechanical stress is greatest at the interface and decreases with distance toward the surface. The dark conductivity activation energy decreases with film thickness. The Urbach energy also follows this trend [2-4]. This means that the thinner the film the greater the average compressive stress and hence the greater the activation energy and the Urbach energy. The density of stress related defects should therefore be greater toward the substrate-film interface. Because of the above observations, there have been considerable ungoing efforts to develop new and compatible substrate materials in order to improve the performance of amorphous silicon.

Photo-induced degradation of amorphous silicon (a-Si:H) used.. in photo receptor devices has seriously limited the development and application of this class of materials. The general interpretation of this photo degradation phenomena is attributed to the creation of metastable silicon dangling bonds[5]. The practice of alloying the material with fluorine to stabilize the dangling bonds and improve performance characteristics has been an ungoing one over the years [6-8]. The bennefits of this approach have not always been positive. Koinuma et al. reported that films made from Si_2F_6 gave good properties when they were prepared in a system that was baked at 3.0 X 10^{-6} torr, whilst those prepared in the unbaked system (8 X 10-6 torr), show inferior electronic properties [6]. Evidence of oxide formation in those films with poor properties were also seen.

Consequently, there is a need to optimize the parametric conditions for the addition of fluorine to the matrix of amorphous silicon, in order to improve the desirable electronic properties that may be derived from this material modification process. Proper chemical characterization of the silicon containing fluorine is critical to the explanation of observed material properties. Understanding the material chemistry, may lead to improved material engineering, thus give rise to improved material performance. Modification of the substrate surface prior to deposition of the film is also considered critical to the improvement of performance of these materials. In this

study, X-ray photoelectron spectroscopy is applied to delineate the surface and interracial chemistry of amorphous silicon deposited from silane, silicon tetra fluoride and argon carrier gas.

Experimentation

Amorphous silicon (a-Si:H) material was deposited from silane (SiH_4) and argon, utilizing a 35 kHz Technic PE, PD1 I A plasma deposition system. Films were deposited at approximately 300 ^oC, 0.5 Torr and 100 watts. Prior to deposition the system was bled with argon to achieve a base pressure of approximately 0.08 Torr. An argon plasma was then lit and maintained for approximately 10 minutes. Argon/silane and argon/Silicon tetra fluoride ratios (seem) dilution typically range from 1 8:1 to 18:3 for the samples made in this study. Substrates utilized in this study were single crystal silicon, aluminum oxide (sapphire) aluminum strips and float glass. A Siemens GADDS (general area detector diffraction system) X-ray refractometer was employed in the determination of the extent of the amorphous nature of the deposited material. Deposition rates were established using a Dektak 11 profilometer. Surface and interracial characterization were carried out with a PHI 5600ci X-ray Photoelectron spectrometer. Valence band structure spectra were determined with a Torroidal monochromator with an aluminum source. Routine spectra were gathered with a standard dual anode X-ray source.

Results and Discussion

Results of film thickness measurements indicated that films ranging from 1.5 microns to 2.5 microns of a-Si:H made from Ar/SiH_4 gases, can be easily deposited. Those films made from Ar/SiF_4 were generally much thinner for the same deposition parameters (-0.5 to 0.8 microns). X-ray diffraction studies show that the deposited films are amorphous in nature. One striking observation is seen in the comparison of the elemental composition of films made from pure silane (SiH_4), films made si I icon tetra fluoride (SiF_4) and those made from a mixture of the two gases (SiH_4/SiF_4). The results are summarized in figure 1, where the relative elemental atomic percentage is plotted against the fraction of silicon tetra fluoride used. As the concentration of fluorine increases, the oxygen concentration rises dramatically, while there is a significant decrease in the silicon content. This phenomenon is typical of all the films regardless of the substrates used. The fluorine content itself remains relatively low, not exceeding 10% in the films. The binding energy of fluorine is found at 687.5 eV. This region is typical of silicon fluoride compounds.

Figure 2 shows XPS depth profile of the silicon 2p region for films deposited from pure silane and pure silicon tetra fluoride. A very thin layer of silicon dioxide is present on the surface of the a-Si:H film (less than 20 Angstroms) while the silicon., oxide species are present as surface/bulk matrix species in films deposited from SiF_4 and. iH_4/SiF_4. It would appear that the presence of fluorine in amorphous silicon promotes the scavenging of oxygen from the system, producing a "silicon oxy-fluoride" compound having silicon-oxygen bonds. This supports the findings of Koinumo et al., who studied films deposited from Si_2F_6. The presence of oxygen in a-Si:H:F compound is observed as a bulk component, while that in a-Si:H is of a very thin surface oxide overlayer. Figure 3 shows the Si 2p XPS region for several plasma compositions of argon, silane and silicon tetrafluoride. In one case the films were characterized for their surface species. In another case the films were sputtered to expose bulk properties. It is obvious that as the SiF_4 concentration is increased, the bulk formation of oxide increases.

Preliminary studies are being conducted using an amorphous silicon layer modified with fluorine, as an electronically inactive buffer layer (rich in fluorine content). The intent of this layer is to absorb the mechanical stresses near the substrate sample interface, by providing a good match and reducing the potential of ordering at the interface. There is a distinctive chemical interface between the pre-deposited layer of a-Si:H:F:O and a-Si:H deposited from pure silane. The stability of this interface layer under illumination and thermal conditions along with the electronic properties of a-Si:H on the buffer layer will be further investigated. It is expected that the performance of a-S i :H could be improved considerable y with this modification.

Conclusions

Results from these experiments indicate that oxide formation may occur in silicon material when deposited from silane and silicon tetrafluoride. In order to deposit bulk oxide free silicon, it may require that the plasma system be totally oxygen free during the deposition process, It was further observed that formation of bulk oxide species is heightened in the presence of fluorine. The fluorine content in films are relatively low when compared to the silicon and oxygen content. Future studies **on** the electrical properties of these films as a function of fluorine content are planned along with the investigation of substrate modification,

Acknowledgement

The Authors wish to acknowledge the support of the Office of Naval Research (Solid State Division), which has provided the necessary infrastructure and made it possible for these studies. The National Science Foundation through its RIMI program was instrumental in the development of the Surface Science Facility at Prairie View A&M University. The NASA Center for Applied Radiation Research (CARR) has also provided support for this project. The Department of Energy has also provided support for this project.

References

[1] Luft, W. and Tsu, S., Hydrogenated Amorphous Silicon Alloy Deposition Processes, Marcel Dekker, Inc., N.Y. (1993).

[2] Kurtz, S. R.. Tsu, S. Y. and Tsu, R., Correlation of Stress with Light Induced Defects in Hydrogenated Amorphous Silicon Films, Appl. Phys. Lett., #15, 49 (1986'),

[3] Stuzmann, M., The Role of Mechanical stress in the light Induced Degradation of Hydrogenated Amorphous Silicon, Appl. Phys. Lett., #1, 47, (1985).

[4] Nakatani, K. Ogasawara, M. Suzuki, K. and Okaniwa, H., Electrical properties of Hydrogenated Amorphous Silicon Layers on a Polymer Substrate Under Stress, Appl. Phys. Lett., #17,54 (1989).

[5] Stuzmann, M. Jackson, W. B., Tsai, C. C., Light Induced Metastable Defects in Amorphous Silicon: A Systematic Study, Phys. Rev., #1, 32 (1985).

[6] Nakayama, Y., Wakimura, K., Takahashi, S., Kita, H. and Kawamura, T., Plasma Deposition of a-Si:H:F films from SiH_2F_2 and SiF_4-SiH_4. , J. of Non-Crystalline Solids, #77 & 78, 797 (1985).

[7] Matsumura, Y., Ihara, H., Tachibana, H. and Tanaka, H., Highly Photo-Conductive Films Amorphous Silicon Produced by Thermal CVD Method Using Intermediate Species SiF_2, J. of Non-Crystalline Solids, #77 & 78,793 (1985).

[8] Langford, A. A., Fleet, M. L., Nelson, A. J., Asher, S. E., Goral, J. P. and Mason A., Determination of Fluorine Content in a-Si:H:F by Infrared Spectroscopy, Electron Probe "Microanalysis, X-ray Photoelectron Spectroscopy and Second Ion Mass Spectroscopy., J. Appl. Phys. #65 (12) 5154 (1989)

[9] Koinuma, H. , Manako, T. , Natsuaki, H., Fujiora, H. Fueki, K., "Photosensitive Amorphous Silicon Films Deposited from Hexafluorodisilane", J. Non-Crystalline Solids, #77 & 78, 801 (1985).

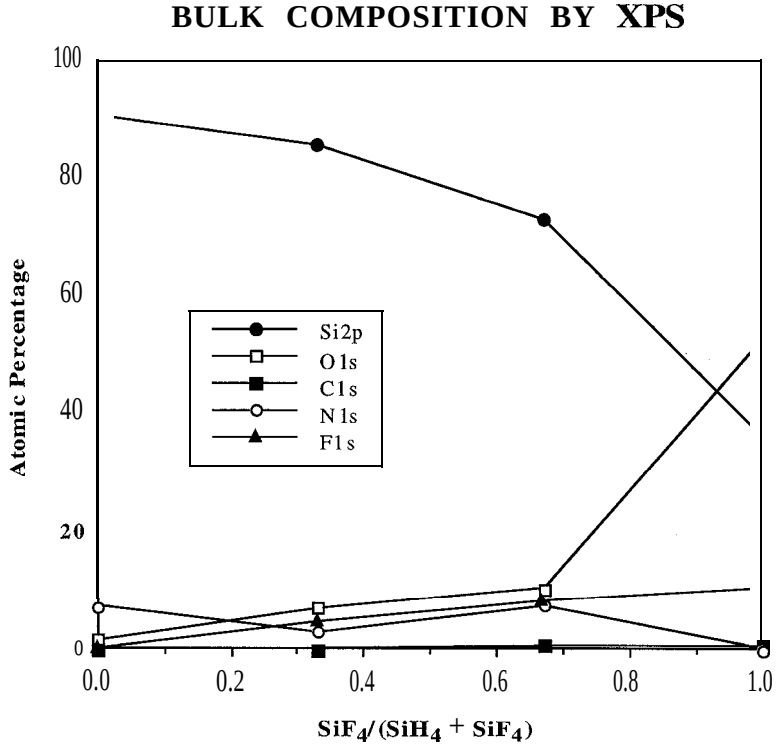


Figure 2. XPS relative atomic percentage of components in amorphous silicon films as a function of SiF_4 concentration.

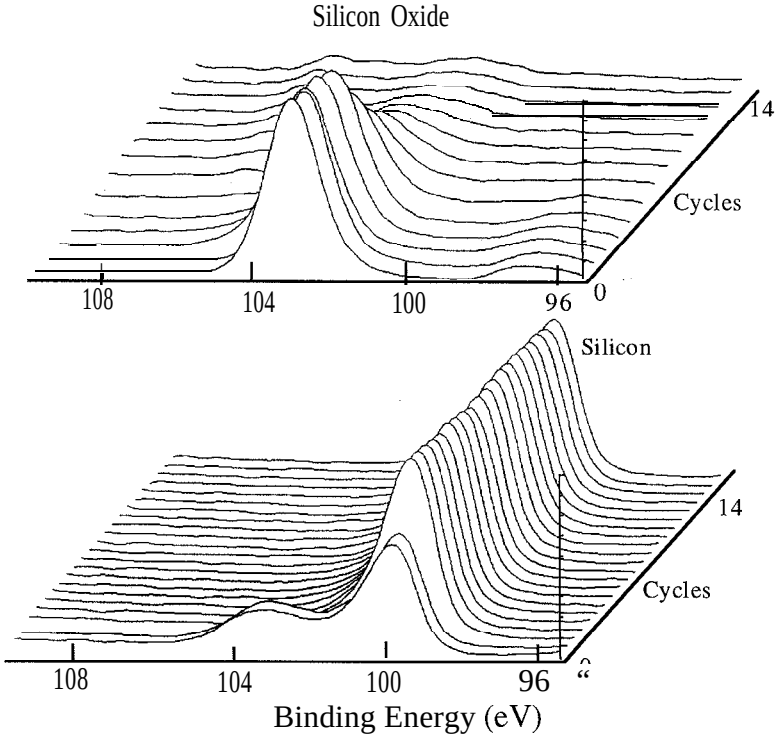


Figure 2. XPS depth profile of silicon 2p region for silicon film made from (a) Ar/SiH_4 and (b) Ar/SiH_4.

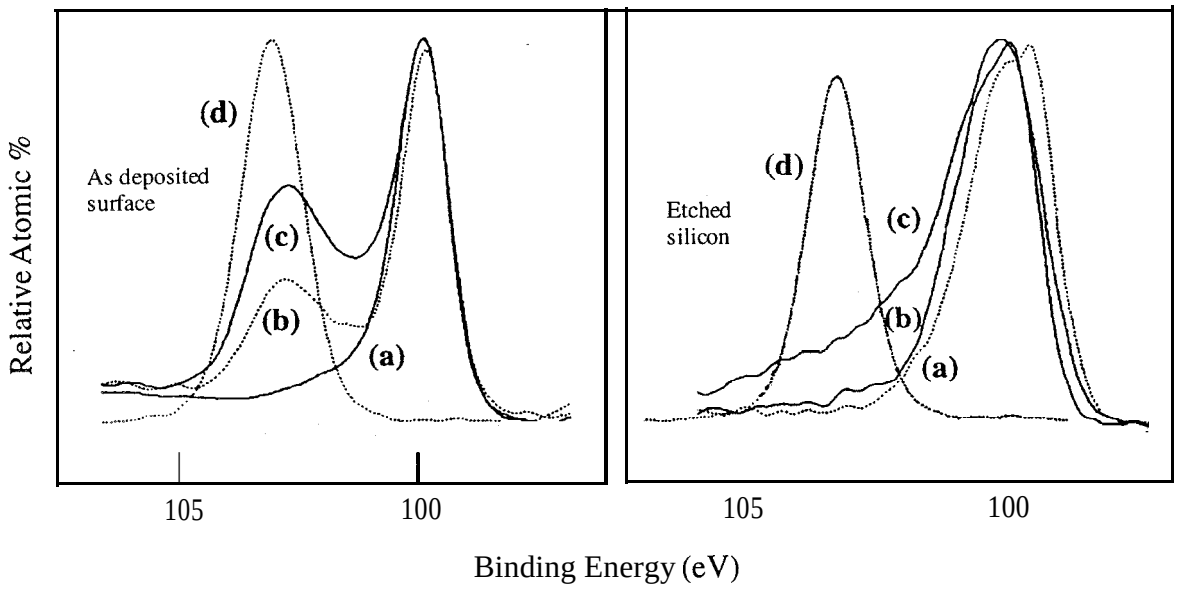


Figure 3. XPS spectra of Si 2p region for films made from various Ar:SiH4:SiF4 gas mixture ratios(a) 18:3:0 (b)18:2:1(c) 18:1:2 (d) 18:0:3. Peaks at ~100eV are for silicon metal. Peaks between 103 and 105 eV are silicon associated with oxide.

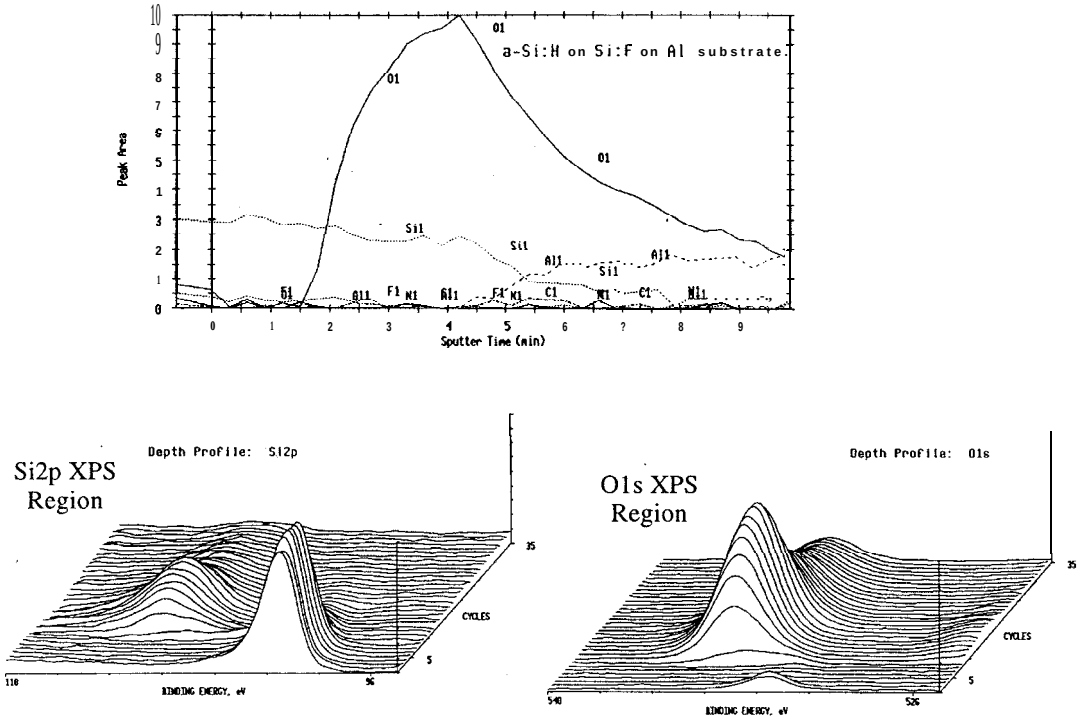


Figure 4. XPS sputter depth profile of a-Si:H film from silane (18:3:0) deposited on a layer previously deposited from SiF4 (18:0:3) on aluminum substrate.

Luminescence of Eu^{3+} doped SiO_2 thin films and glass prepared by sol-gel technology

Lymari Castro, Weiyi Jia, Yanyun Wang, Miguel Santiago and Huimin Liu
Department of Physics, University of Puerto Rico, Mayaguez, PR 00681-9016

I. Introduction

Trivalent europium ions are an important luminophore for lighting and display. [1 ,2] The emission of $^5D_0$ to $^7F_2$ transition exhibits a red color at about 610 nm, which is very attractive and fulfills the requirement for most red-emitting phosphors including lamp and cathode ray phosphorescence materials. [3,4] Various Eu^{3+} doped phosphors have been developed, and luminescence properties have been extensively studied. [1-6]

On the other hand, sol-gel technology has been well developed by chemists, [7] In recent years, applications of this technology to optical materials have drawn a great attention. [8,9] Sol-gel technology provides a unique way to obtain homogeneous composition distribution and uniform doping, and the processing temperature can be very low. In this work, Eu^{3+} doped SiO_2 thin films and glasses were prepared by sol-gel technology and their spectroscopic properties were investigated.

II. Sample preparation

Eu^{3+}-doped SiO_2 thin films and glasses were prepared by sol-gel technology. The flow chart of the preparation is shown in Fig. 1. Starting materials were tetraethoxysilane (TEOS = $Si(OEt)_4$), ethanol and water with the molar ratio 2:1:1. Drops of HCl acid was added as catalysis into the mixture to adjust its pH value to ~3.5. The mixture was stirred using an ultrasonic stirring machine. Meantime, adding certain amount of the aqueous solution of $EuCl_3$ dropwise to a required doping level (2 wt% Eu based on SiO_2). Wet gels were obtained by hydrolysis process of the alkoxide. For bulk samples, the obtained wet gel was dried slowly at 45 °C in air in a plastic cuvet sealed with parafilm having a pin hole on it for about two weeks. Clear transparent SiO_2 dry gels were obtained.

Thin films were prepared by spinning technique. The fresh wet gel was dropped onto a glass or quartz substrate which was spinning on a spin coating machine. The thickness of the thin films can be controlled by adjusting spin speed. The thickness of the thin films of Eu:SiO_2 prepared in this work is estimated to be about 700 nm.

The dry glasses and thin films were annealed at temperature 400 to 800 “C for 4 hours. The raising rate of the annealing temperature was less than 30 “C per hour in order to get crack-free samples. Both bulk samples and thin films were amorphous. X-ray diffraction spectrum of the sol-gel glass SiO_2 is a structureless broad peak. Similar result was obtained from the thin film samples.

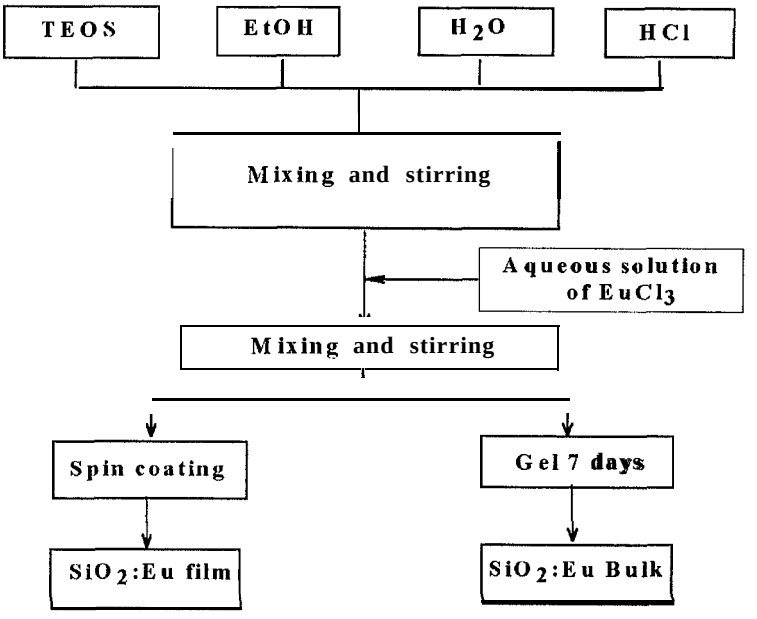


Fig 1. The flow chart for preparing SiO_2:Eu thin films and bulk SiO_2:Eu glass

111. Spectroscopic results

In luminescence measurement, the 457.9 nm emission of an argon laser was used to excite the $^7F_0$ - $^5D_2$ transition of Eu^{3+}. A SPEX double spectrometer and a SRS-400 photon-counter were used to analyze the emission signals. The photoluminescent spectra of Eu^{3+} doped SiO_2 glasses and thin films on quartz substrate, annealed at different temperatures, are shown in Fig. 2 and Fig. 3 respectively.

It can be seen that, with increasing the annealing temperature, the luminescence intensity were increased due to the reduction of the concentration of the OH quenching centers. in addition, the relative intensities of two peaks at 611 nm and 590 nm changed with annealing temperatures. The ratio of 611 nm intensity to that of 590 nm peak increases with annealing temperatures as shown in Fig. 4.

IV. Discussion

The ground state manifold of Eu^{3+} is ^7F_J with J=0, 1, 2, 3,4, 5 and 6. Due to strong spin-orbit coupling, the manifold split into 7 states with $^7F_0$ as ground state. The J degeneracy of each state can be further split due to Stark effect of the crystal field about the ion, $^7F_0$ is an orbit singlet, no Stark split occurs. The same situation happens to the excited state manifold of ^5D_J with J=0, 1, 2, 3 and 4. $^5D_0$ is an orbit singlet and serves as the lowest excited state. The intra-f configuration transitions can occur via electric and magnetic dipole radiation or vibronic transitions. These transition are restricted to the flowing selection rules [10]:

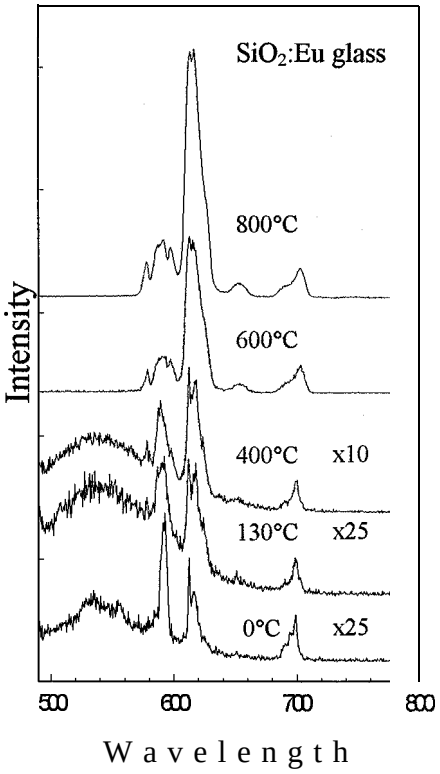


Fig.2 Fluorescent spectra of SiO_2:Eu glasses annealed at different temperatures.

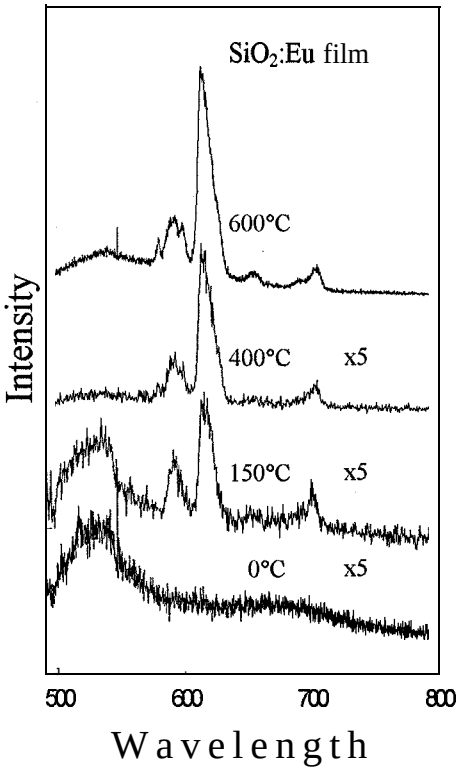


Fig. 3 Fluorescent spectra of SiO_2:Eu thin films on quartz substrate, annealed at different temperatures.

Electric dipole:

$$\Delta S=0,\ l \pm 1,\ |\Delta J| \leq 6 \text{ unless J or J' = O when } |\Delta J| = 2,4,\ 6;$$

Magnetic dipole:

$$\Delta S=\ \Delta L=0;\ \Delta J = 0,\ \pm 1 \text{ (but J=0 to J'=0 forbidden).}$$

Theoretically, transitions between f^n levels involve no change in parity and hence forbidden by the LaPorte rule. [10] However, the ligand field at the rare earth ions in the host distorts the free ion energy levels and causes higher states of opposite parity admix. The $4f^n$ states are actually a linear combination with $4f^{n-1}5d^1$ and the transitions become partially electric-dipole allowed. On the other hand, however, the transitions from ^5D_J to $^7F_{J'}$ are actually spin forbidden. State admixing of ^5D_J with $^7F_{J'}$ may relax the spin selection rule, but the transition probability is relatively low in spite of parity allowance, and the lifetime of the luminescence is quite long in the order of milliseconds.

The luminescence peak at 590 nm shown in Fig. 2 and 3 is the transition from $^5D_0$ to the ground state $^7F_1$. This transition is magnetic-dipole allowed, does not depend on the lattice environment and almost remains constant in different host materials. This transition is dominating in the samples not annealed or annealed at lower temperature. All other peaks in these samples are vibronic transitions. They are weak and broad. From the features of the luminescent spectra, it can be concluded that europium ions in the un-annealed samples stay at an environment of near inversion symmetry. In the samples annealed at higher temperatures, the local environment of europium ions might be distorted and deviated from inversion symmetry. As a result, the transition $^5D_0$ to $^7F_2$ becomes dominating. This transition is a so-called forced electric-dipole transition, governed by the selection rules AS = O, $\Delta J = 2$, 4 and 6 with the initial J=0, and $|\Delta L| \leq 2$. Transitions to the levels with uneven J are forbidden. $^5D_0$ to $^7F_{4,6}$ transitions are weak forced electric-dipole emission.

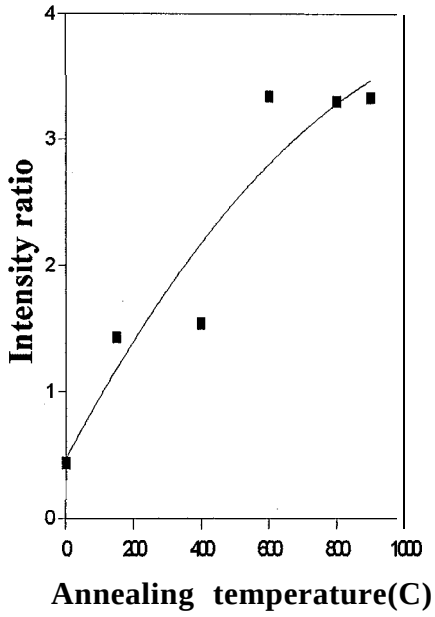


Fig.4 Intensity ratio of 611 nm peak to 590 nm peak of luminescence of Eu:SiO_2 glass versus annealing temperatures.

Transition from $^5D_0$ to $^7F_2$ is hypersensitive to the environment. Even for small deviations from inversion symmetry, it appears dominantly in the spectrum. Therefore, the $^5D_0 \rightarrow {}^7F_2$ / $^5D_0 \rightarrow {}^7F_1$ transition intensity ratio is a good measure of the site symmetry of rare earth ions. The intensity ratio between the transition $^5D_0 \rightarrow {}^7F_2$ at 611 nm and the transition $^5D_0 + {}^7F_1$ at 590 nm is rapidly increased with annealing temperatures and reached the stable value at 800 ‘C. In the case of thin films, the Eu sites deviated from inversion symmetry more obvious even at lower annealing temperatures. This may be due to the surface effects which distorts the site symmetry more effectively. As mentioned in the introduction, intense red emission at 611 nm is an important factor for phosphor application. The annealed SiO_2 glass and thin films provide such promising color.

V. **Conclusion**

Sol-gel technology provided a useful method to produce uniform thin films and bulk samples of Eu^{3+} doped SiO_2 glass. With increasing annealing temperature, the intensity of Eu luminescence increased due to reduction of concentration of OH quenching centers, and the emission peak at 610 nm from $^5D_0$ to $^7F_2$ electric dipole transition became dominant over the magnetic dipole transition from $^5D_0$ to $^7F_1$ because the local symmetry of Eu sites deviated from inversion symmetry owing to annealing process.

Work is supported by NASA under the grant MURC-NCCW-0088, DOE under grant DE-FG02-94ER75764, NSF under the grant EHR-9 108775 and ARO under grant DAAH04-96-10416.

References:

1. G. Blasse and B. C. Grabmaier, *"Luminescent Materials"*, Springer-Verlag, New York, 1994.
2. M. Dejneka, E. Snitzer and R. E. Riman, J. Lumin., 65 (1995) 227.
3. A.K. Levine and F. C. Palilla, Appl. Phys. Letters, 5 (1964) 118.
4. T. Kane, K. Kinameri and S. Seki, J. Electrochem. Sot., 129 (1982) 2296.
5. M. Buijs, A. Meijerink and G. Blasse, J. Lumin., 37 (1987)9.
6. G. A. West and K. W. Beeson, J. Mater. Res. 5 (1990) 1573.
7. C.J. Brinker and G.W. Scherer, "The *Physics and Chemistry of Sol-Gel Processing"*, the third edition, Academic Press, San Diego, 1990.
8. I. M. Thomas, S.A. Payne and G.D. Wilke, J. NonCryst. Solids, 151 (1992) 183.
9. D. Moutonnet, R. Chaplain, M. Gauneau, Y. Pelous and J.L. Rehspringer, Mater. Sci. Eng. B9 (1991) 455.
10. R. Reifeld, in: *"Structure and Bonding"*, Vol.22, Springer-Verlag, New York 1975, p123.

Fluorescence of Pentavalent Chromium in SiO_2 Sol-Gel Glasses

Weiyi Jia, Lymari Castro, Yanyun Wang, and Huimin Liu
Department of Physics, University of Puerto Rico, Mayaguez, PR 00681-9016

[. Introduction

Chromium ions are very attractive to optical spectroscopy and laser physics. It is well known that the first laser in the history is a ruby laser activated with Cr^{3+}.[1] It was found in early nineties that Cr^{4+} was also an interesting lasing ion in the near infrared, [2] and various Cr^{4+} lasers have been developed. [3] Very recently, it was reported that Cr^{2+} doped in CdSe crystals showed lasing action in the infrared. [4] The above achievement have stimulated an interest in searching for Cr^{5+} and investigating its optical properties.

Cr^{5+} is isoelectronic with Ti^{3+} and V^{4+}, having electron configuration $3d^1$. Ti^{3+} is the active center of commercial cw and femtosecond sapphire lasers, tunable in the range 680-1100 nm. V^{4+} doped in $YAlO_3$ and Al_2O_3 showed broad band emission near 635 nm. [5] Although EPR results of Cr^{5+} were reported, [6,7] the optical properties were less studied. Herren et al. reported an observation of luminescence from Cr doped in SiO_2 sol-gel glass. [8,9] The luminescence spectrum was assigned to pentavalent ions in their first paper, [8] and later it was identified to be the emission from the charge transfer transition of Cr^{6+}.[9] The first observation of photoluminescence from octahedrally coordinated Cr^{5+} in BaCaMg aluminate glasses was reported very recently. [10]

[n this work, we report luminescence results of Cr doped SiO_2 sol-gel glasses. The fluorescence spectra are very different from Herrens' results, and we believe it originates from pentavalent Cr.

IL Sample preparation

Cr-doped SiO_2 glasses were prepared by sol-gel technology.[11] Starting material was tetraethoxysilane (TEOS = $Si(OEt)_4$) doped with 0.2wt% Cr (based on SiO_2) in the isopropanol solution of $Cr(OOC_8H_{15})(OC_3H_7)_2$. Wet gels were obtained by hydrolysis of the alkoxides. The molar ratio of TEOS, ethanol and water (pH - 3.5) for hydrolysis was 2:1:1. The obtained wet gel was dried slowly at 45 'C in air in a plastic cuvete sealed with parafilm having a pin hole on it for about two weeks. Clear transparent SiO_2 dry gels were obtained. Cr doped samples show blue-green color indicating the existence of Cr^{3+}. The dry glasses were annealed at temperature 400 to 800 'C for 4 hours. It seems that Cr ions in SiO_2 dry gels are easy to be oxidized. Cr doped samples turned to yellow after annealing above 400 'C in air.

111. Experimental results

The absorption of Cr doped SiO_2 dry gel has two bands located at 410 and 590 nm (Fig. la), This is a typical result corresponding to the transitions from $^4A_2$ to $^4T_2$ and $^4T_1$ of Cr^{3+}. A weak fluorescence near 700 nm from $^4T_2$ state was also observed. After annealing in air, optical properties of the samples were dramatically changed due to oxidation of Cr ions. The

color of the samples turned to yellow from blue green. The absorption spectrum of Cr-doped SiO_2 glass annealed at 800 °C. is shown in Fig. lb. Two absorption bands were observed at 457 nm and 352 nm. The fluorescence spectra of the samples annealed at different temperatures are shown in Fig. 2.

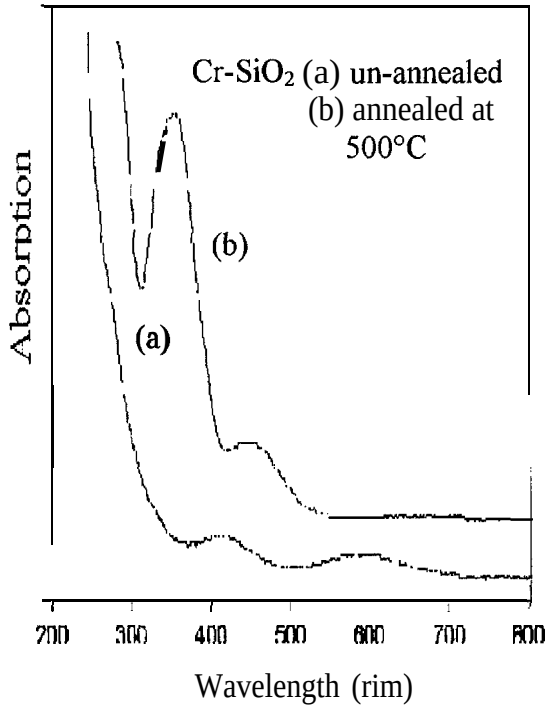


Fig 1. Absorption of Cr-doped SiO_2 glass (dry gel) (a) and annealed Cr doped SiO_2 glass (b).

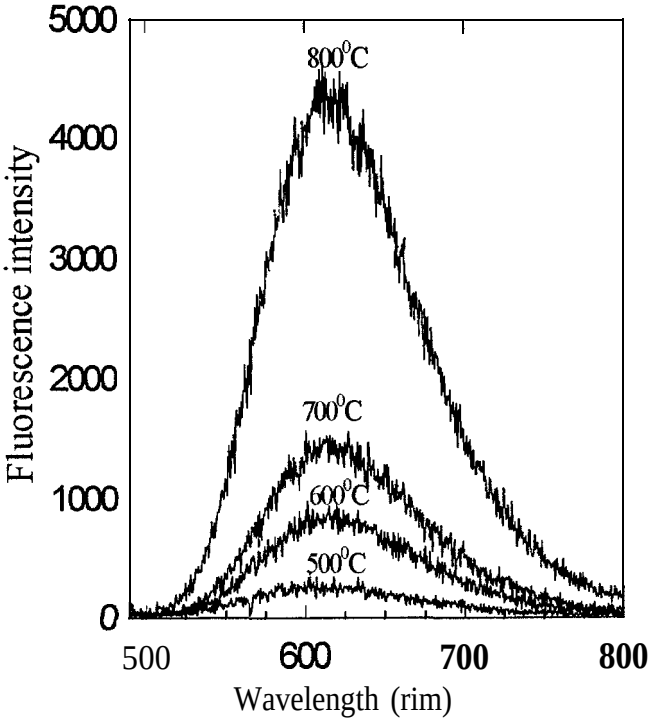


Fig 2. Fluorescent spectra of Cr-doped SiO_2 glasses annealed at different temperatures.

With increasing the annealing temperature, the fluorescent intensity was increased. The fluorescence band of the annealed samples is peaked at 630 nm and remains unchanged with annealing temperatures. The band width is about 2870 cm-'. The fluorescence spectrum at 10 K are structureless and similar to the result at room temperature.

Excitation spectrum, monitoring at 630 nm at room temperature, is shown in Fig.3. It can be seen that the broad band is an overlapping result of two absorption bands.

Decay process was studied by using second harmonic output of 532 nm of a Nd:YAG laser as a pump source and a multichannel scaler was used to record the transient signals. In the wide temperature region from 10 to 300 K, the decay process of the fluorescence are approximately exponential. Temperature dependence of the lifetime is shown in Fig.4. The lifetime drops from 35 μs at 10 K to 3.8 μs at 300 K.

IV. Discussion

Herren at al. have reported spectroscopic results of Cr doped sol-gel SiO_2 glass. [8,9] The absorption spectrum of their sample is peaked at 445 nm and 346 nm, and the fluorescence band is located at 672 nm with bandwidth of 4200 cm-l, They assigned the emission to the charge

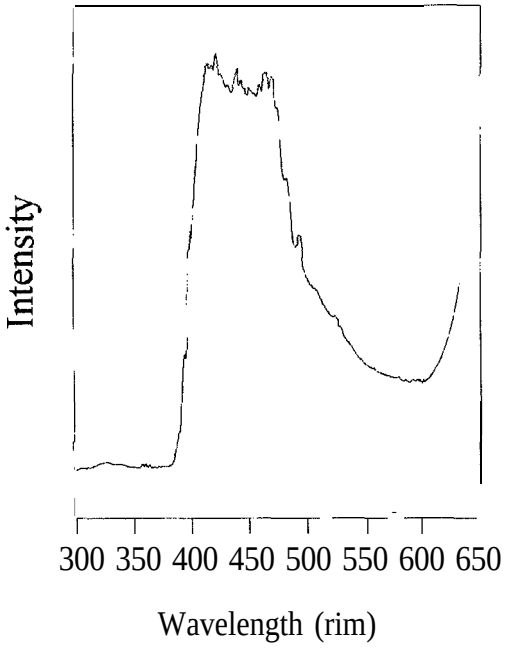


Fig 3. Excitation spectrum of Cr-doped SiO_2 glass annealed at 800°C, monitoring at 650nm.

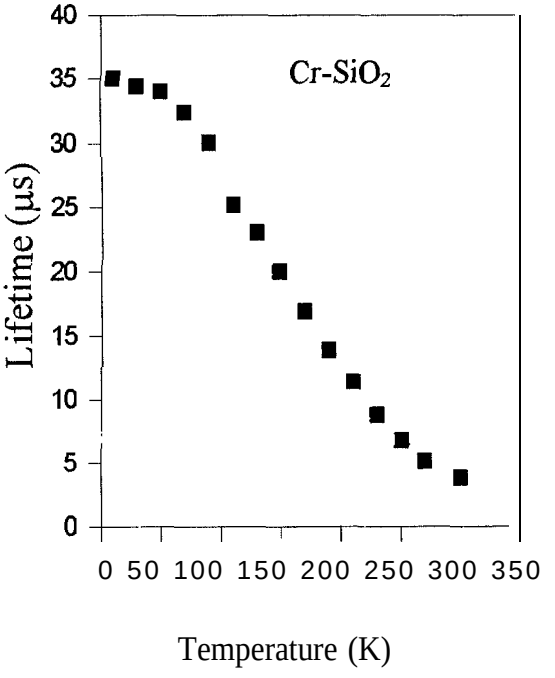


Fig 4. Temperature dependence of lifetime of fluorescence of Cr-doped SiO_2 glass annealed at 800°C.

transfer transition from the excited states $^3T_1$ and $^3T_2$ to the ground state $^1A_1$ of $Cr^{6+}O_4$ groups. This transition is spin forbidden and has a long lifetime, as observed in their experiment (7.7 ms at 12 K). This is the most important evidence to support the assignment. Compared with Herren's work, the spectroscopic results of our samples are very different. The absorption bands of our samples are located at 457 nm and 352 nm. The fluorescence spectrum is at 630 nm with bandwidth of 2870 cm-l, 50% narrower than that of Herren's sample. The lifetime of our samples is 35 μs at 10 K, 200 times shorter than that of Herren's sample. In addition, the fluorescent band of Herren's sample at low temperature exhibited a typical vibronic structure with phonon frequency 970 cm-l from the $Cr^{6+}O_4$ group. In contrast, the low temperature fluorescence of our sample is structureless. Based on the above observations, we believe that the fluorescence of our samples originates from octahedrally coordinated Cr^{5+}. In addition, the possibility of existence of Cr^{3+} and Cr^{4+} in detectable amount can be ruled out due to absence of emission near 700 nm (Cr^{3+}) and in the near infrared (Cr^{4+}).

As mentioned before, Cr^{5+} is one of $3d^1$ ions. in an octahedral crystal field, the energy level of a $3d^1$ ion, 'D, splits into two energy levels, the ground state $^2T_2$ and the excited state ^2E. Due to a low symmetry component of a distorted octahedral crystal field and Jahn-Teller effect, both $^2T_2$ and ^2E states may further split. A splitting of ^2E up to 2,200 cm-l was observed in Ti^{3+}

doped Al_2O_3 crystals due to the static Jahn-Teller effect. [12] The excitation spectrum shown in Fig.3 is corresponding to the transition $^2T_2$ to ^2E. The spectrum consists of two overlapped bands, which are ascribed to the splitting of ^2E state. The absorption peak at 357 nm shown in Fig. 1 does not appear in the excitation spectrum. This indicates that the absorption band does not belong to Cr^{5+}. It must be related with color center or other defect from the host SiO_2.

In SiO_2, Si stays at tetrahedral sites. In tetrahedral crystal field, the energy order of the splitting of ^2D of Cr^{5+} is reversed and ^2E serves as the ground state. The crystal field strength, 10 Dq, is four ninths of that of an octahedral field. So the energy separation between ^2E and $^2T_2$ states is much smaller and falls in the near infrared. This is not our case. Therefore, the possibility of Cr^{5+} to replace Si at tetrahedral sites in the SiO_2 glass can be ruled out.

In Table 1 are listed some spectroscopic results of various $3d^1$ ions. It can be seen that our spectroscopic observation is in consistence with the fluorescent results of octahedrally coordinated $3\,d^1$ ions observed in other host materials. The lifetime of our sample at 10 K is one order larger than that of other examples of $3\,d^1$. We do not fully understand the physical reason. It maybe due to the approximate inversion symmetry of local environment existing in SiO_2 sol-gel glass.[16]

Table 1. Spectroscopic Data of $3d^1$ ions at low temperature (- 10 K)

Ion	Coordination	Emission max.	FWHM	Lifetime	Reference
Tr^{3+}	octahedral	700nm	$1800cm^{-1}$	3.9μs	[12,13]
V^{4+}	Octahedral	635nm	$3100cm^{-1}$	1.27μs	[5]
Cr^{5+}	Octahedra	745nm	$2800cm^{-1}$	1.1μs	[10]
Cr^{5+}	Octahedral	630nm	$2870cm^{-1}$	34μs	[this work]
Cr^{5+}	Tetrahedral	1.1μm	$1900cm^{-1}$	100ns	[14]
Mn^{6+}	Tetrahedral	1.1μm	$1800cm^{-1}$	2μs	[15]

Finally, one question still remains: why our sample is different from Herren's. To answer this question we should check the process of sample preparation. Herren's samples were prepared using aqueous solution of chromium nitrate (or chloride) to replace water for hydrolysis. Our samples were prepared by using solution of 10wt% $Cr(OOC_8H_{15})_2(OC_3H_7)$ in isopropanol. The different starting materials may be an important factor to determine the final valence state of Cr during annealing process. In addition, the annealing period of Herrens' samples was 24 Hours, much longer than ours (4 Hours). Annealing temperature may not be an important factor to affect the valence state. In the range of annealing temperature from 400 to 800 "C used in this work a stable pentavalent state of Cr can be obtained (Fig. 2).

V. Conclusion

Uniform and transparent SiO_2 glass doped with chromium ions was successfully prepared by sol-gel technology. It was found that Chromium ions were easily to be oxidized to the pentavalent state through an annealing process. A broad band emission from pentavalent Cr was observed at 630 nm with bandwidth 2870 cm-l. The lifetime of the emission at room temperature was measured to be 3.8 μs. Pentavalent chromium ions, Cr^{5+}, are attractive and may hold promising of applications to laser or phosphor materials.

Work was supported by NASA under grant MURC-NCCW-0088, ARO under grant DAAH04-96-1 0416, and DOE under grant DE-FR02-94ER757640.

References:

1. T.H. Maiman, Nature, 187 (1960) 493.
2. V. Petricevic, S.K. Gayen and R.R. Alfano, Appl. Phys. Lett., 53 (1988) 2590.
3. H. Ellers, W.M. Dennis, W.M. Yen, S. Kück, K. Peterman, G. Huber and W. Jia, IEEE J. Quant. Electron., 29 (1993) 2508.
4. L.D. DeLoach, R.H. Page, G.D. Wilke, S.A. Payne, and W.F. Krupe, OSA Proceedings on Advanced Solid State Lasers, 24 (1995) 127.
5. J.P. Meyn, T. Danger, K. Peterman and G. Huber, J. Lumin., 55 (1993) 55.
6. N.S. Garif' yanov, Soviet Physics-Solid State, 4 (1963) 1795.
7. K. Tanaka and K. Kamiya, J. Mat. Scien. Lett., 10 (1991) 1095.
8. M. Herren, H. Nishiuchi, and Morita, J. Chem. Phys., 101 (1994) 4461.
9. M. Herren, K. Yamanaka and M. Morita, Technical Reports of Seikei University, 32, No.2 (1995) 61.
10. Huabiao Yuan, Weiyi Jia, D. Cohen, W.M. Yen and B.G. Aiken, Mat. Res. Sot. Symp. Proc., 455 (1997) 483.
11 C.J. Brinker and G.W. Scherer, the Physics and Chemistry of Sol-Gel Processing, the third edition, Academic Press, San "Diego, 1990.
12 P.F. Moulton, J. Opt. Sot. Am. B, 3 (1) (1986) 125.
13 P. Albers, E. Stark and G. Huber, J. Opt. Sot. Am. B, 3 (1)(1986) 134.
14 M.F. Hanzenkamp and H.U. Güdel, Chem. Phys. Lett., 251 (1996) 301.
15. T.C. Brunold, M.F. Hazenkamp and H.U. Güdel, J. Am. Chem. Sot., 117 (1995) 5598.
16. L. Castro, W. Jia, Y. Wang, M. Santiago and H. Liu, seethe Proceedings of this conference.

PHOTOREFRACTIVE EFFECT IN CdMnTe:V CRYSTAL

K. M. Pour, K. Chattopadahyay, H. Chen, K.T.Chen, S.Morgan, and A. Burger

NASA/Fisk Center for Photonic Materials and Devices,
Department of Physics, Fisk University, Nashville TN 37208

Abstract:

We present two-wave mixing result obtained with a CdMnTe:V crystal. A photorefractive gain coefficient of 0.20 cm-l was observed at 633 nm with the signal -to-pump ratio being of the order of 10-3. This crystal was grown by vertical Bridgman Technique and doped with Vanadium during the growth. The crystal were of good optical quality and showed high resistivity. Room temperature absorption and low temperature photoluminescence studies comparing the band edge and defect center at the doped and undoped CdMnTe crystal will also be discussed.

Introduction:

Photorefractive materials are nonlinear optical materials which experience a significant change in there refractive index when exposed to inhomogeneous illumination. The space charge electric fields formed inside photorefractive materials modulate the refractive index of the host materials via the electro-optic effect, which in turn causes the diffraction of incident optical fields and therefore provides the foundation of photorefractive applications. In general, semiconductor photorefractive materials have the advantages of significantly faster response times and better sensitivity in both the visible and near-infrared wavelength regions compared to conventional oxide based materials. This region is particularly interesting for applications to optical telecommunications.

H-VI semiconductors are good candidate for photorefractive materials. Previous work by Brost et. al have shown photorefractive effect in vanadiun-doped $Cd_{0.55}Mn_{0.45}Te$. The band gap of $Cd_{1-x}Mn_xTe$ can be adjusted by changing the mole fraction x, over a broad energy range. [1] As Mn is substituted for Cd in the CdTe lattice, the energy gap increases, to the point that the originally opaque material becomes transparent to the visible light. The dopant Vanadium provides the deep level donor needed for photorefraction. In this paper we will present the photorefractive gain coefficient of this material obtained from two beam coupling experiment. The dependence of photorefractive gain coefficient on intensity ratio of signal (I_s) and pump (I_p) beam will also be presented. The measurements were performed at 633 nm using He-Ne laser. We will also present the low temperature photo luminescence and absorption spectra of this crystal.

Theory:

In two beam coupling experiment two beams of different intensity intersect inside the semi-insulating sample to create interference pattern as shown in Figure 1. Charge carriers which are in a constructive interference region get excited to the conduction band. These charge carriers then undergo diffusion and get trapped in destructive interference region. These effects result in charge redistribution, which will give rise to the space-charge field. As a result of this field acting through the linear electro optic effect, refractive index modulation is generated inside the material. If the grating is shifted with respect to incident optical interference pattern the beam coupling results in energy transfer from one beam to the other.

By chopping the pump beam and monitoring the signal beam, the photorefractive gain can be measured. The ratio of the modulation of the signal-beam ΔI_s to the signal beam I_s

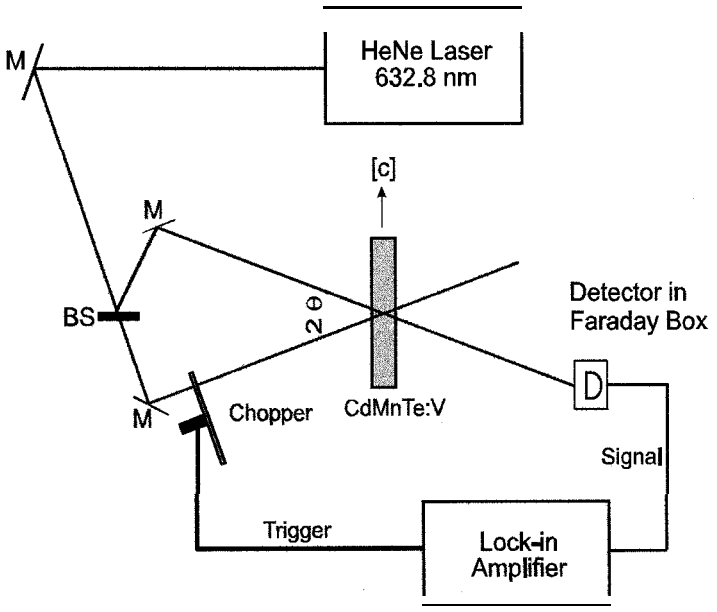


Figure 1: Experimental setup for photorefractive two-beam coupling.

without pump beam is the called the photorefractive Gain. However photorefractive gain coefficient Γ could be found using Equation 1 given by[2]

$$\frac{\Delta I_s}{I_s} = \frac{\exp(\Gamma L) - 1}{1 + \beta \exp(\Gamma L)} \qquad \textbf{(1)}$$

In this equation L is the overlap length of the beams and β is the intensity ratio of the signal beam to pump beam. Equation 1 could be solved for Γ to give the relation

$$\Gamma = \frac{1}{L}\ln\left[\frac{\frac{\Delta I_s}{I_s}+1}{1-\beta\frac{\Delta I_s}{I_s}}\right] \qquad (2)$$

The gain coefficient Γ could also be estimated by using simplified model given by Kukhtarev and others[3], and Equation 3 shows this relation.

$$\Gamma = \frac{2\pi\left(n_0^3 r_{eff}\right) k_B T \qquad kg}{\lambda\cos\theta \quad (\quad e \quad) 1+\left(k_g^2/k_D^2\right)}\xi_0 \qquad (3)$$

where ξ_0 is the electron-hole competition factor, k_D is the inverse of the Debye screening length, r_{eff} is the effective electro-optic coefficient and n_0 is the linear refractive index. Choice of component of r_{eff} depends on crystal point group, sample orientation, and beam polarization. For one level model, $k_d^2=(e^2/\varepsilon k_B T)N_{eff}$ where N_{eff} is the effective trap density and ε is the dielectric constant. The value of grating wave vector is given by $k_g=2k\sin\theta$ where 2θ is the angle between the two crossed beams.

Figure 2 shows the interaction of the two beams inside the crystal. From simple geometry we can calculate the interaction length, once we know the beam diameter and the angle between the interacting beams. This relation is given in Equation 4.

$$L = \frac{a+b}{2\sin 0} \qquad (4)$$

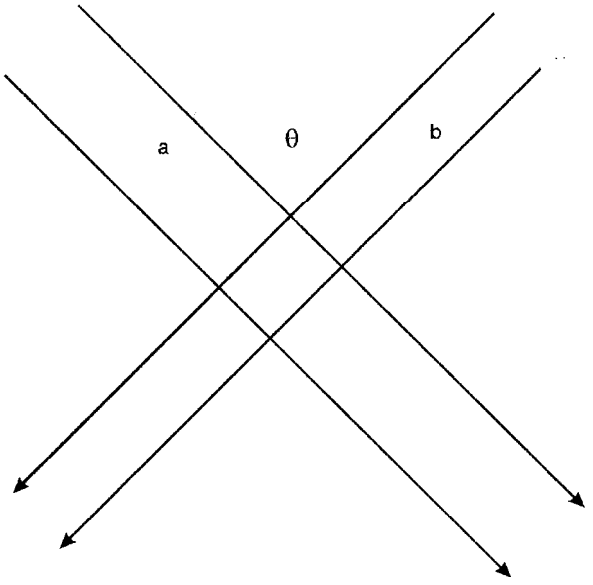


Figure 2. Dependence of L on θ and beam thickness

Results and Discussion :

A CdMnTe:V crystal with stoichometry of $Cd_{0.55}Mn_{0.45}Te$ doped with Vanadium at the level of 10^{19}atom cm^{-3} was grown by vertical Bridgman Techique[4]. Figure 3 shows room temperature absorption spectra of our material. The band edge is near 2.1 eV.

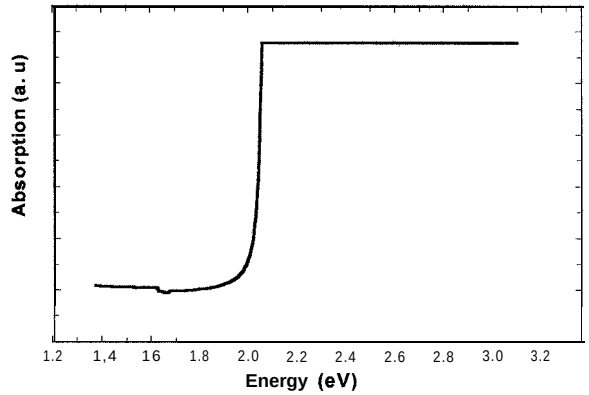


Figure 3. Room temperature absorption spectra of $Cd_{.55}Mn_{.45}Te$.

For the low temperature PL measurements, the samples were cooled to 11 K using an APD Cryogenic, Inc., system equipped with dual HC-4MK1 helium compressors. The sample was illuminated with the 488 nm laser line. Figure 4 shows the PL spectra. A dominant band is seen for each sample, which corresponds to transitions from the first excited ^4G to the ground state of ^6S of the localized Mn^{2+} ions.[5] This luminescence peak is essentially at the same

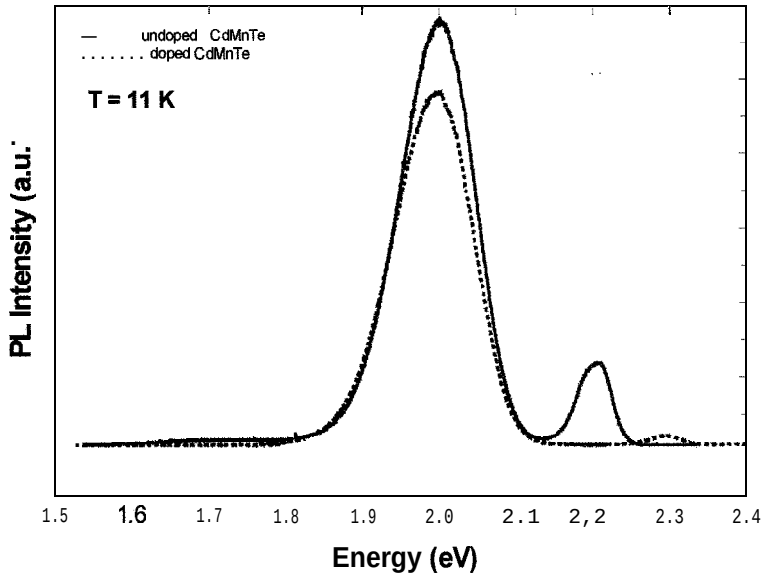


Figure 4. PL spectra of vanadium doped and undoped CdMnTe

energy level of 1.97 + 0.01 eV for all compositions of x >0.30.6 The higher energy peaks at 2.2 eV for the undoped sample and 2.3 eV for the doped sample is associated with recombination across the fundamental gap. This peak has a linear dependence on the mole fraction x of Mn.

As shown in Figure 1, the two beam I_s and I_p which were vertically polarized were made to overlap inside the crystal. Then by using a chopper we chopped the I_p and measured I_s and also its modulation AI, which had exact frequency of the chopper. To make sure that the I_p was not finding it's way to the detector by means of external reflection and diffraction we blocked the I_s to see if we would still observe some modulation with chopper frequency. Our detector was mounted inside a Faraday box and we placed a pinhole in front of the detector to prevent the effects of scattered and diffracted light. Then we measured the photorefractive gain and also studied its dependence on the intensity ratio β. The pump beam was modulated at 84 Hz by a chopper. The results obtained with a wavelength of 633 nm, a beam crossing angle of 2θ of 35°, and beam diameter of 1 mm are shown in Figure 5. To our knowledge this is the first report of such study on vanadium doped CdMnTe crystal. Brost et. al. had shown the same dependence of Γ on intensity ratio, but they had applied an external field to their CdMnTe crystal to enhance the gain. The maximum measured gain coefficient for wavelength of 633 nm at a 1.06 μm grating period was 0.20 cm^{-1}. This value is slightly better than earlier reported values[4] on the same material.

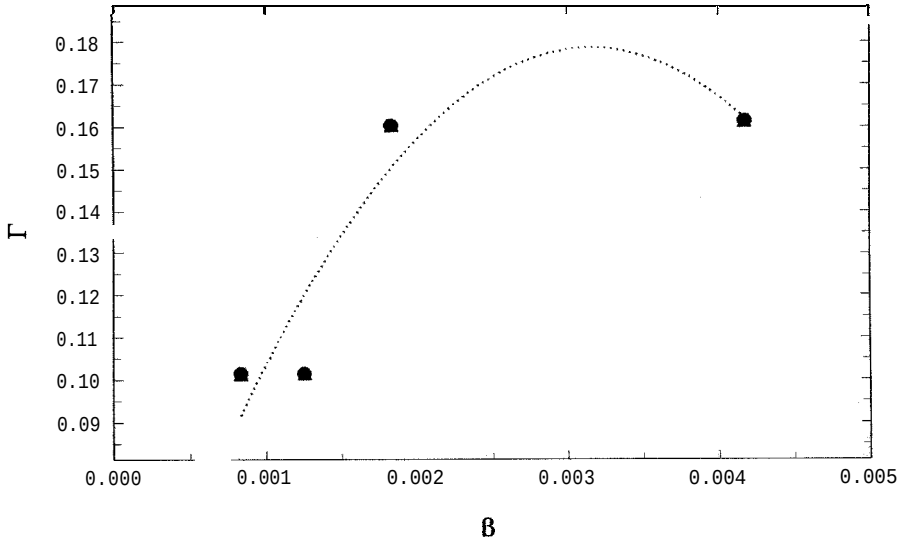


Figure 5. Two beam coupling gain coefficient Γ as a function of incident beam ratio β.

Since overlap length is a function of 0 and beams thickness as evident from Equation 4, it is obvious that one can not change L by changing θ without affecting the value of kg By calculating the overlap length of our two beams with thickness a and b we realized that the thickness of the crystal is smaller than the overlap length for most of our incident angles. Therefore we were unable to study the effect of incident angle on the photorefractive gain of the material.

Conclusion :

Here in this article we have reported the photorefractive effect in a vanadium doped CdMnTe crystal. We have shown the dependence of the photorefractive gain coefficient with the intensity ratio of the signal to pump beam, without the application of any external electric field. The low temperature PL spectra show the transitions between the excited and ground state of Mn^{2+} ions and excitonic recombination. Further work needs to be done to understand the role of vanadium in the properties of CdMnTe.

Acknowledgments :

This research is supported by the NSF Grant HRD-9550605, US Air Force Grant GYF1457096-03 120, and NASA Grant NAGW-2925.

Biography :

K. M. Pour is a graduate student in the Masters degree program in Physics at Fisk University. He is currently working on photorefractive measurements of II-VI semiconductors at the NASA/Fisk Center for Photonic Materials and Devices. His research mainly involves optical and electrical characterization of semiconductor and optical materials. He is also interested in solid state physics, device physics, theoretical physics and astronomy. He is the recipient of the United States Academic Achievement (USAA) Award. He plans to pursue a Ph. D. degree after graduation from Fisk.

References :

1. P. Lautenshlager, Slogthetidis, L. Vina, and M. Cardona, Phys. Rev. 1332 (1 985) 3811.

2. R. B. Bylsma, P. M. Bridenbaugh, D. H. Olson, and A. M. Glass, Appl. Phys. Lett., 51 (1987) 51.

3. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. I. Vinetskii, Ferroelectrics 22, (1979) 949.

4. George A. Brost, Kevin M. Magde, Sudhir Trivedi, Optical Materials 4 (1995) 224.

5. M. Blume and R. Orbach, Phys. Rev. 127, (1962) 1587.

6. M. P. Vecchi, W. Giriat, and L. Videla, Appl. Phys. Lett. 38 (1 991) 99.

Surface Passivation of CdZnTe Detector by Hydrogen Peroxide Solution Etching

M. Hayes', H. Chen[1], K. Chattopadhyay[1], A. Burger' and R. B. James[2]
[1]NASA/Fisk Center for Photonic Materials and Devices,
Department of Physics,
Fisk University, Nashville, TN 37208
'Advanced Electronics Manufacturing Technologies Department,
Sandia National Laboratories, Livermore, CA 94550

Abstract

The spectral resolution of room temperature nuclear radiation detectors such as CdZnTe is usually limited by the presence of conducting surface species that increase the surface leakage current. Studies have shown that the leakage current can be reduced by proper surface preparation. In this study, we try to optimize the performance of CdZnTe detector by etching the detector with hydrogen peroxide solution as function of concentration and etching time. The passivation effect that hydrogen peroxide introduces have been investigated by current-voltage (I-V) measurement on both parallel strips and metal-semiconductor-metal configurations. The improvements on the spectral response of ^{55}Fe and ^{241}Am due to hydrogen peroxide treatment are presented and discussed.

Introduction

Recently, a new generation of x-ray and gamma-ray imaging detectors and spectrometers based on CdZnTe (CZT) has been realized[1-4]. CZT detectors have rapidly captured a significant interest due to a compact size, good energy resolution and room temperature operation capability. Although the physical properties of the bulk crystal set a fundamental limit to the performance of the detector, it is the fabrication process that determines the actual performance of the device. Of great importance is surface passivation. This is because unpassivated surfaces usually lead to high surface leakage current and l/f noise current which significantly affect the overall performance of the detector via spectral broadening.

Due to a very limited number of reports on CZT surface passivation especially with regard to detector performance, an in-dept investigation on this issue is needed. in this work, we present a study of surface passivation of CZT detectors by hydrogen peroxide (H_2O_2) solution etching. I-V curves as well as x-ray and gamma-ray spectral response of CZT detectors treated with different H_2O_2 concentrations and different etching times were compared with that of the untreated sample to reveal the passivating effect of H_2O_2 cm CZT detector performance.

Experimental

The samples used in this investigation were grown by the High Pressure Bridgman method and purchased from eV-Products, Inc. They were first polished on a mechanical polisher with 0.05 μm particle size alumina suspension and then rinsed with methanol. The samples were then chemically etched in 5°/0 Bromine in methanol for 2 minutes and washed in methanol. This chemical etching process is necessary to ensure room temperature response of CZT detectors[5,6]. Au contacts were then deposited by the thermal evaporation method immediately after the chemical treatment. Pt leads with a diameter of 0.01 in. were next attached to the contacts using Aquadag (graphite suspension in water). The contact areas were then covered with "Humiseal" to provide mechanical stability. In addition to the Metal-Semiconductor-Metal detector configuration, we also fabricate a one-size parallel strips configuration to measure the interstrip surface leakage current before and after the oxidation. The devices were finally treated with hydrogen peroxide aqueous solution at different concentrations and etching times. The passivated areas are those regions on the CZT surface just outside the metal contacts.

For I-V characterization, each sample was mounted on a Teflon holder and then placed in a closed aluminum testing box. The measurements were carried out in room temperature using a Keithley programmable electrometer (Model 617) in conjunction with a Bertran high voltage power supply (Series 225). The data acquisition was computer controlled via a GPIB interface.

For the detector performance study, ^{55}Fe and ^{241}Am sources were used. The source was placed 1 mm away from the detector surface. The detector was mounted on a specially designed boron nitride holder and inserted in an aluminum testing box from which the electrodes were connected to an EG&G preamplifier (Model 142A). The charge pulse produced in the detector volume by absorption of gamma radiation was converted to a voltage pulse. The signal was further amplified by a Tennelec TC 242 shaping amplifier before it was fed into a multichannel analyzer for spectral analysis.

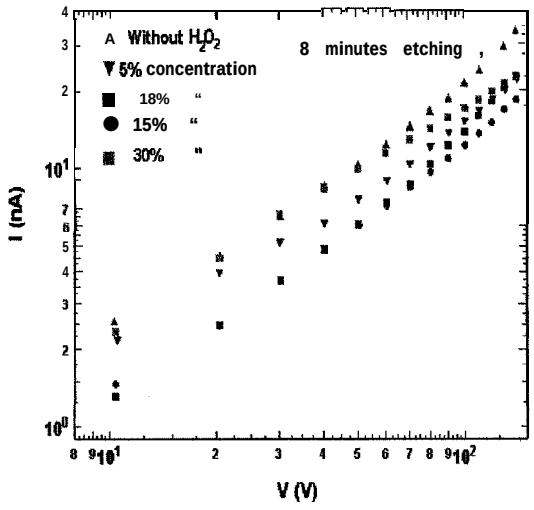
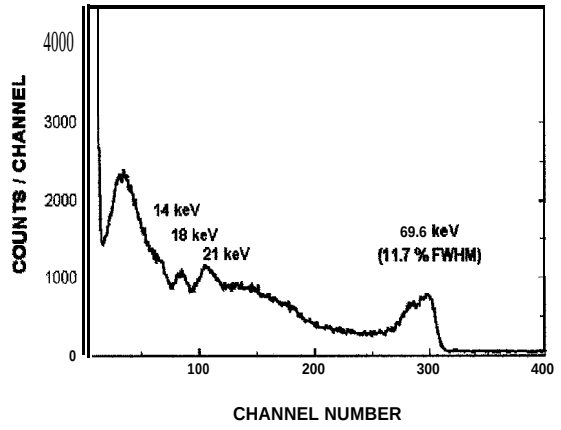


Fig. 1 Comparison of I-V curves of CZT detector treated by different concentrations of H_2O_2 aqueous solution for 8 minutes.

Results and discussion

Shown in Fig. 1 are the I-V curves of CZT detectors treated with different concentrations of H_2O_2 for 8 minutes. Obviously, the treatment leads to a lower level of the leakage current, as much as 43% decreasing for 15% H_2O_2 solution. Consequently, the energy resolution in spectral response is significantly enhanced as shown in Fig. 2 where the x-ray photopeaks and the 59.6 keV gamma peak of ^{241}Am source are well resolved after H_2O_2 passivation (FWHM at 59.6 keV

(a)



(b)

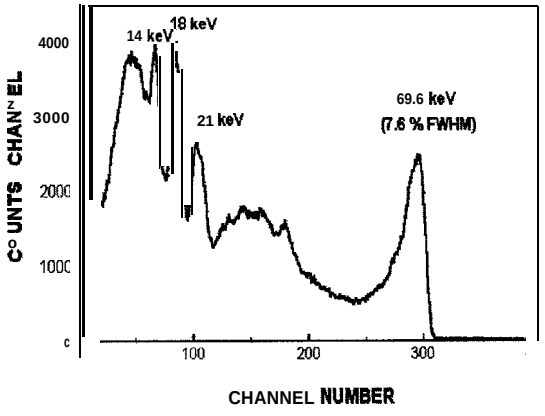


Fig.2 ^{241}Am spectrum of CZT detector treated by concentration of H_2O_2 in 8 minutes (room To, A=0.07 cm^2, L=0.08 cm, 100V bias, 100s accumulation time, 2 μs peaking time) : **(a)** untreated, **(b)** 15%.

Table 1. Comparison of I-V and ^{241}Am gamma spectrum results of CdZnTe detector treated by different concentrations of H_2O_2 (for 8 minutes).

H_2O_2 concentration (% vol.)	Current density at E = 1000 V/cm (nA/cm^2)	%FWHM at 59.6keV ^{241}Am photopeak (100V bias)
without H_2O_2	303	11.7%
5%	214	10.45%
10%	194	8.8%
15%	172	7.6%
30%	239	10.2%

improves from 11.7% to 7.6%). A summary of I-V measurements and ^{241}Am spectral responses of CZT detectors as a function of H_2O_2 concentration is shown in Table 1.

An interest phenomena can be observed is that the beneficial effect of surface oxidation of CZT detector via H_2O_2 seems to have a threshold level, beyond that the" result can be worsen. This is evidenced by a higher level of leakage current and a worse energy resolution of the sample treated with 30% concentration compared to that corresponding to 15°/0 concentration as shown in Fig. 1 and Table 1.That is to say, at higher concentration of H_2O_2, the passivating effect seems to be less effective.

This phenomena can be even more easily seen when we compare the I-V curves and spectral responses of CZT detector treated with the same concentration (1 5°/0) of H_2O_2 but at different etching times. The passivating effect seems to be very time dependent as shown in Fig. 3 and Fig. 4 where the detector leakage current and the interstrip surface leakage current are measured as a

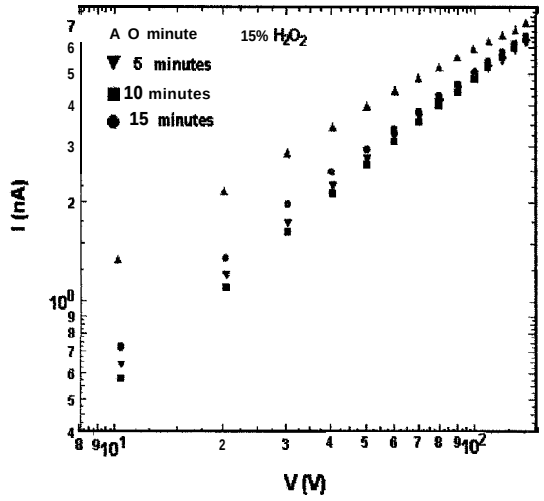


Fig.3: Comparison of I-V results of CZT detector passivated by 15% H_2O_2 aqueous solution in different times.

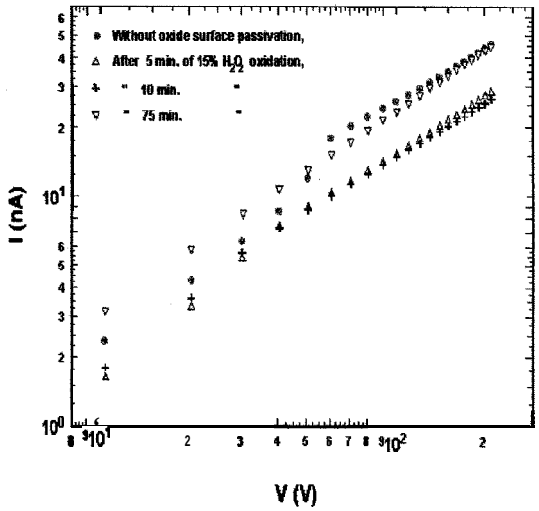


Fig.4 : Inter-strip (2mm) surface leakage current measurements of CZT passivated by 15% H_2O_2 aqueous solution in different times.

Table 2. Comparison of I-V, ^{55}Fe and ^{241}Am spectral results of CdZnTe detector passivated by 15% H_2O_2 in different times.

Oxidation time (minutes)	Current density at 100V bibs (nA/cm^2)	%FWHM at 5.9keV ^{55}Fe photopeak (at 100V bias)	%FWHM at 59.6keV ^{241}Am photopeak (at 100V bias)
0	85	36%	8.2%
5	74	34%	6.7%
10	69	33%	6.7%
15	72	3594	7.9%

function of H_2O_2 etching time respectively. After a period of about 10 minutes, the leakage current can no longer be reduced. In fact, as can be seen in Fig. 3, the result is worsen after 15 minutes and it is much worse for very long etching time as shown in Fig. 4 for the 75 minutes etching sample. The improvement in leakage current due to passivating effect and the time dependence phenomena is also consistent with detector performance as shown in Table 2 where %FWHM at 5.9 keV ^{55}Fe photopeak and 59.6 keV ^{241}Am photopeak of CZT detectors at different etching times are compared. This

threshold effect of surface passivation has also been observed for CZT detectors passivated by a different method of oxidation, via low energy atomic oxygen bombardment[7].

The reduction of the leakage current can be attributed to the insulating oxide layer generated. The oxide species has been identified as mainly TeO_2[8]. Other possibilities are $CdTeO_3$ and $ZnTeO_3$. However, the amorphous oxide generated can be any of a variety of ternary compound of the form $[CdTe]_{1-x}O_x$ with variable bandgap depending on the amount of oxygen presented in the generated film[9,10]. At higher H_2O_2 concentration, especially at longer etching time, some nonstoichiometric and more conductive oxide species may have been formed and consequently caused a degrading effect on the oxidized CZT detectors and this may explains the worsening effect observed in this investigation.

Summary

In conclusion, the results reported in this investigation have clearly demonstrated the important role of surface passivation on CZT detector performance. The study shows that the beneficial passivating effect of surface oxidation of CZT detector via H_2O_2 aqueous solution etching is strongly etching-time and concentration dependent. Finally, the study suggests that a 15% H_2O_2 treatment for 8 minutes will results in optimum passivation for room temperature CZT radiation detectors.

Acknowledgment

The authors acknowledge the support of NASA through the Fisk Center for Photonic Materials and Devices, by Grant NCC8-133.

References

1) R. B. James, T. E. Schlesinger, J. Lund and M. Schieber, in *Semiconductors for Room Temperature Nuclear Detector Applications,* edited by T. E. Schlesinger and R. B. James, Vol.43, (Academic Press, San Diego, 1995) p. 335.

2) L. A. Hamel, J. R. Marci, C. M. Stable, J. Odom, F. Birsa and P. Shu, IEEE Trans. Nucl. Sci., 43, (1996) 1422.

3) J. R. Marci, B. A. Apotovsky, J. F. Butler, M. L. Cherry, B. K. Ryan, A. Drake, F. P. Doty, T. G. Guzik, K. Larson, M. Mayer, M. L. McConnell and J. M. Ryan, IEEE Trans. Nucl. Sci. 43, (1996) 1458.

4) L. M. Barlett, C. M. Stable, D. Palmer, L. M. Barbier, S. D. Barthelmy, F. Birsa, N. Gehrels, J. F. Krizmanic, P. Kurczynski, J. Odom, A. M. Parsons, C. Sappington, P. Shu, B. J. Teegarden and J. Tueller, Proceedings of SPIE, *2806 (1996) 616.*

5) H. Chen, S. U. Egarievwe, Z. Hu, J. Tong, D. T. Shi, G. H. Wu, K.-T. Chen, M. A. George, W. E. Collins, A. Burger, R. B. James, C. M. Stable and L. M. Bartlett, Proc. of SPIE Int.

Symp. on Opt. Sci., Eng. and Instrum., 2859 (1996) 254.

6) H. Chen, J. Tong, Z. Hu, D. T. Shi, G. H. Wu, K.-T. Chen, M. A. George, W. E. Collins, A. Burger, R. B. James, C. M. Stable and L. M. Bartlett, J. Appl. Phys. *80 (6) (1996) 3509.*

7) H. Chen, K. Chattopadhyay, K.-T. Chen, A. Burger, M.A. George, J. J. Weimer, P. K. Nag, J. C. Gregory and R B. James, Proc. 1997 AVS Nat Symp. (to be published).

8) K.-T. Chen, D.T. Shi, H. Chen, B. Granderson, M.A George, W.E. Collins, A. Burger and R.B. James,
J. Vat. Sci. Technol. A 15(3)(1997) 850.

9) F. J. Espinoza-Beltran, O. Zelaya, F. Sanchez-Sinencio, J. G. Mendoza-Alvarez, M. H. Farias and L. Banos,
J. Vat. Sci. Technol. A 11(6), Nov/Dec (1993) 3062.

10) F. J. Espinoza-Beltran, F. Sanchez-Sinencio, O. Zelaya-Angel, J. G. Mendoza-Alvarez, C. Alejo-Armenta, C. Vazquez-Lopez, M. H. Farias, G. Soto, L. Cots-Araiza, J. L. Pena, J. A. Azamar-Barrios and L. Banes, J. Apply. Phys. No. **10A**, Oct. (1991) 1715.

Biography

Miguel "Hayes is a graduate student in the Master degree program in Physics at Fisk University. He has been involved in CZT detectors fabrication and optimization in the NASA/Fisk Center for Photonic Materials and Devices. He is interested in electronics, solid state physics, devices fabrication and evaluation. He plans to pursue a Ph.D. degree after graduation from Fisk.

Molecular Modeling of the Compounds with Nonlinear Optical Properties

Tatiana V. **Timofeeva, Beatriz** H. **Cardelino,** Ronald D. Clark
Department of Physical Sciences, New Mexico Highlands University, Las Vegas, New Mexico, 87701.

Abstract

The molecular **polarizability** characteristics for a large series of **naphthoquinone** and **quinoline** derivatives have been calculated. The dependence of calculated **hyperpolarizability on** the positions and the number of donor and **acceptc r substituents** is discussed.

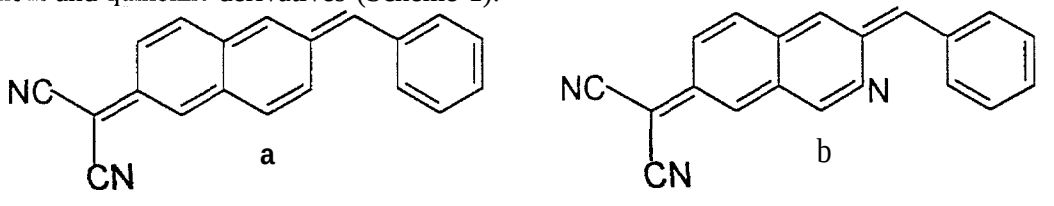
Introduction

Many general requirements for the molecular structure of organic molecules which form materials with specific properties are known [1-3]. These requirements are the main guidelines to the synthetic chemists in their search for compounds with specific optical properties. For instance, in order to find an efficient compound with **nonlinear** optical properties one needs to obtain a molecule with a conjugated π-system, and donor and acceptor groups on the most distant parts of this system. Such molecules demonstrate high **polarizability,** and **NLO** efficiency of the compound related to the **polarizability,** i.e. efficiency of the electron donor and acceptor **substituents** and z-system length. There are some elaborate methods to calculate molecular **polarizability** characteristics on the basis of known **molecular** geometry [4-6]. As a rule, calculated **polarizabilities** correlate with the experimentally found **NLO** properties in solution where the relative positions of **polararizable** molecules are not fixed. On the contrary, not every crystal which consists of highly **polarizable** molecules can reveal the desirable optical properties. For instance, a **centrosymmetric** molecular pattern and some other specific **molecular** relative positions in a crystal extinguish the desirable properties.

Unfortunately, there are no reliable ways to predict compounds with efficient **NLO** characteristics in the crystalline state, which are obviously the most important for practical applications. We proposed a simple **model** which can help to select molecules which are not inclined to form **centrosymmetric** crystal structures. In the literature and in our work it was shown [7-9] that molecules which are inclined to form **centrosymmetric** isolated **dimers** most probably will form **centrosymmetric** crystals, and, on the contrary, molecules which form acentric isolated **dimers** will form acentric crystals. We can model the formation of the **dimers** of different types, compare the formation energy, and select molecules with desirable crystal structures. The method of **dimer** structure and energy calculation was described in details in our previous work [9].

Even if general rides to design molecules with **NLO** properties are known there are a lot of possibilities in organic and **organometallic** chemistry to construct such **molecules** and there are many details which can change desirable properties, For instance, the number and the relative positions of donor and acceptor **substituents** attached to a "core" molecule with conjugated n-system can dramatically change the **polarizability** of the particular molecule.

In the present paper we analyzed the influence of the number and relative positions of the donor and acceptor **substituents (cyano-** and **methoxy-groups)** on the **polarizalility** of the large series of **naphthoquinone** and **quinoline** derivatives (Scheme 1).



Scheme 1

We also calculated dimer formation for some of these molecules. To our knowledge, the synthesis and properties of the above mentioned compounds have not been described yet. So our results should be useful for justifying the synthesis of compounds of this type for the fabrication of NLO materials.

Methods of calculations

Calculations of the conformation and geometry parameters of all the molecules under investigation in the free state were performed using the MM3 program package [10,11]. For computation of the dipole moments, second- and third-order polarizabilities, a modification of the static field method [12] developed in [4,5] was used. In each case, 252 static-field calculations were obtained (MOPAC program, AM1) which were analyzed by the HYPER program [6'],

The energy of the molecular dimers was calculated with the MM3 program package using a stochastic search procedure [13]. The total energy was calculated as the sum of the intra- and intermolecular interactions. The stochastic procedure has some advantages in comparison with mapping of the intermolecular energy as a function of the interplanar distance and in-plane rotation angle described by Itoh et al. [7,8] because molecular conformations and relative molecular orientation in dimers are not restricted during optimization, Two types of dimers were taken into account - centrosymmetric C_i and C_2 symmetry, and their energy was compared in order find the preferable relative molecular orientation.

Results and discussion

For our modeling we chose compounds with a z-conjugated system consisting of a naphthoquinone "core" with dicyanomethylidene and phenylmethylidene substituents in the 2,6-positions (Scheme 1,a). We varied the number and positions of the cyano and methoxy substituents attached to the z-conjugated system in order to find positions corresponding to the highest hyperpolarizabilities values. Methoxy and cyano (donor and acceptor) substituens are not the most efficient ones [14] but we chose them for the synthetic reasons and because of the relative stability of the proposed compounds. We started from the molecule with the minimal number of substituents (molecule I, Table 1) and investigated the series until the compounds core almost completely substituted (molecule XX). We also included in this series several molecules with a longer chain of conjugation and quinoline derivatives.

From the molecular mechanics calculations we see a definite alteration of the bond lengths (and so bond orders) in the naphthoquinone fragment. The range of single bonds is 1.35-1.38 Å, and range of double bonds is 1.46-1.48 A. Such alteration is considered to be a good condition for obtaining a high polarizability in organic molecules [15].

According to molecular mechanics, calculations all molecules with a and b cores (Scheme 1) consist of two planar fragments (naphthoquinone and benzene). The dihedral angle between these fragments for most of the naphthoquinone derivatives is equal to 50±2°. There are some exceptions however. For instance, for molecules XII, XVII and XVIII with close positions of CN- and OMe-groups this angle is equal to 54-55°, and for molecules XVI and XX this angle is even larger (61°) because of the stericaly unfavorable positions of OMe-groups in Ph-substituents.

In quinoline derivatives (XXIV, XXV) the dihedral angles between the two planar fragments are significantly smaller than in naphthoquinone derivatives (36-37°). It is understandable since short steric interaction in naphthoquinone derivatives exists between H atoms (for instance, C-H,. .H-C 2.46 A, molecule XV, Figure 1,a) and in quinoline derivatives there is a similar interaction between N and H atoms (C-H.. .N 2.38& molecule XXIV, Figure 1,b). These results indicate that the polarizability values should be higher for the last group of compounds which are merely planar and where n-conjugation is larger.

Table 1

Dipole moments, β and γ polarizabilities for molecules - XXV

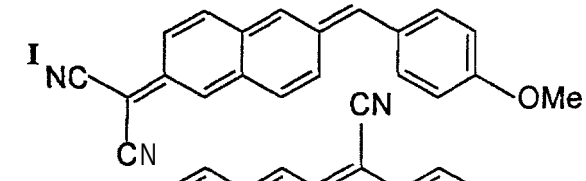
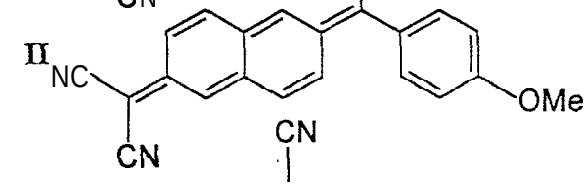
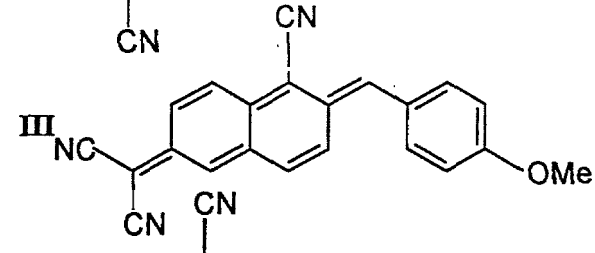
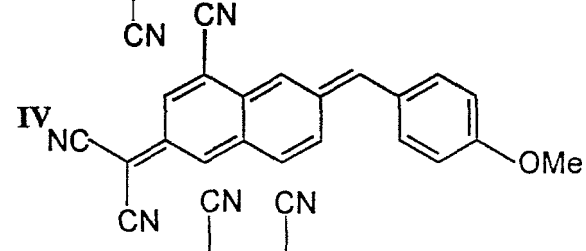
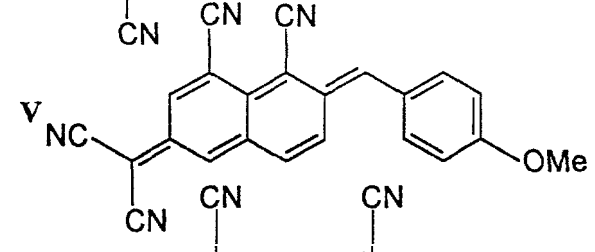
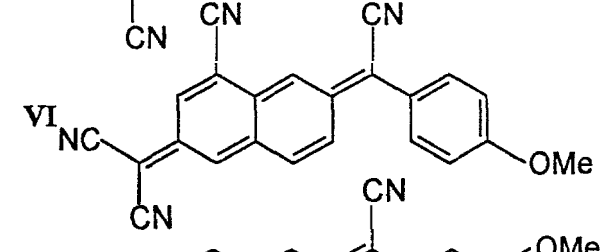
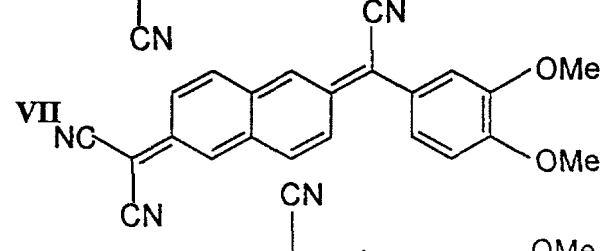
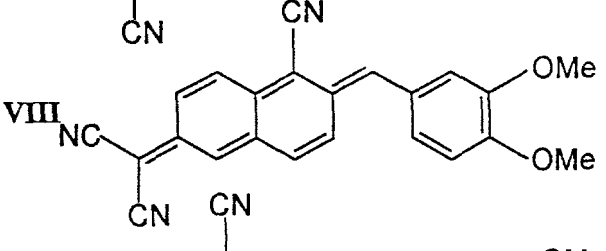
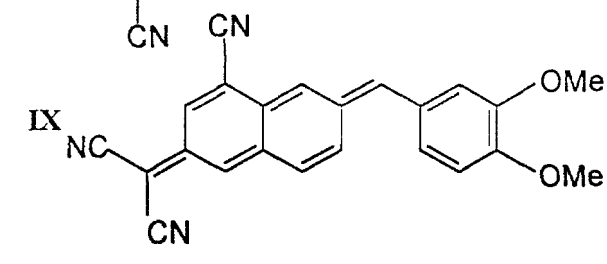
	μ	β	γ
I 	*9.45*	149.49 + 4.56	*59.57 + 1.62*
II 	*6.87*	149.30 + 6.46	*70.63 + 3.03*
III 	*7.64*	*198.53 + 7.91*	*68.20 + 3.42*
IV 	*8.93*	202.33 + 6.65	66.77 + 2.75
V 	*8.20*	239.14 +10.41	73.37 + 4.48
VI 	7.52	200.12 + 8.86	76.47 +4. 13
VII 	4.47	103.33 + 3.05	61.37 + 0.60
VIII 	7.00	183.17+8.72	65.40 + 2.50
IX 	7.36	215.31 +12.58	73.17+4.08

Table 1 (continued)

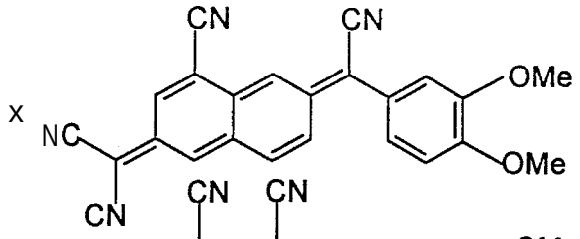
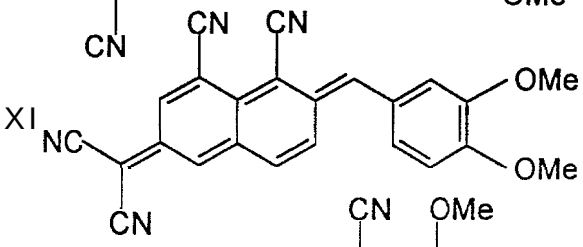
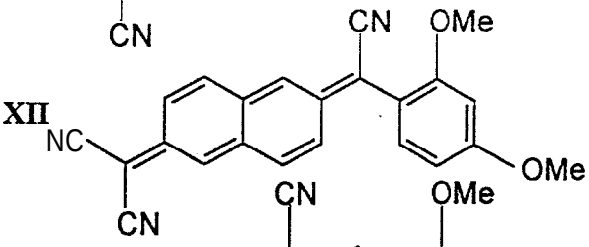
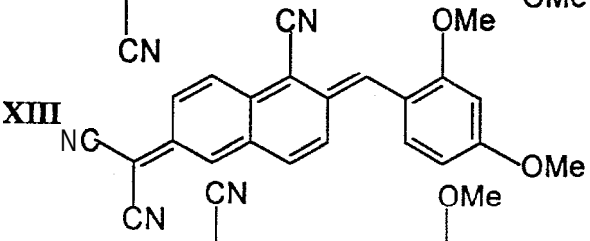
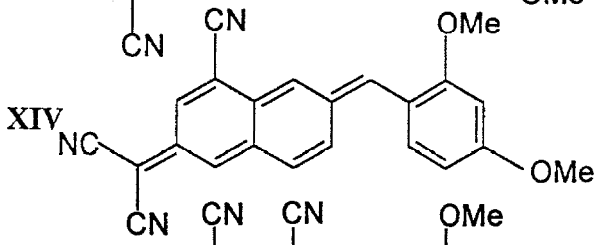
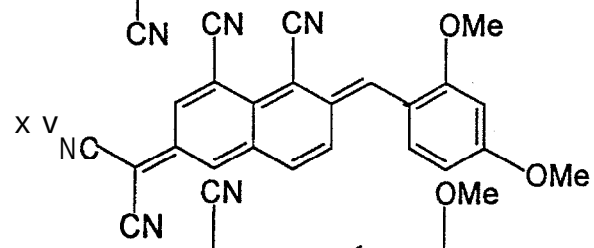
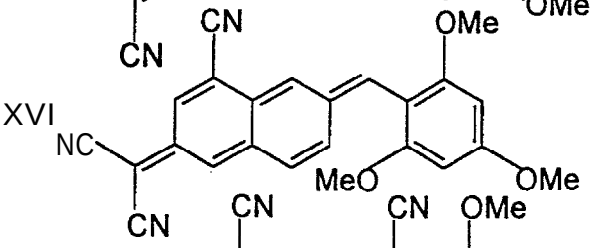
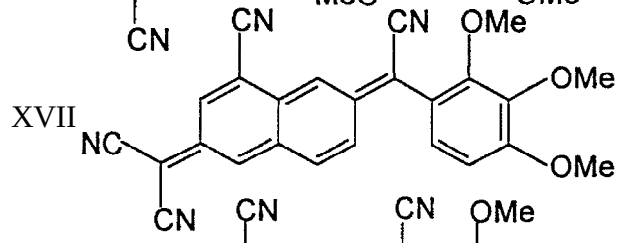
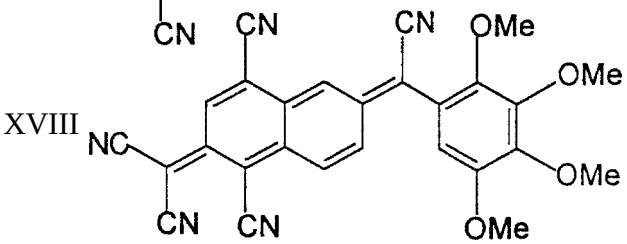
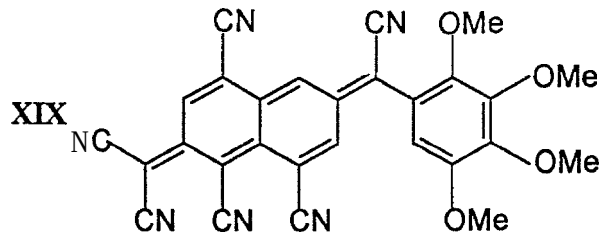
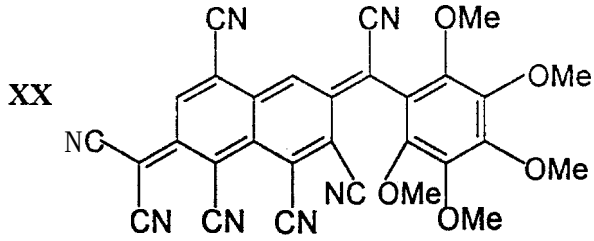
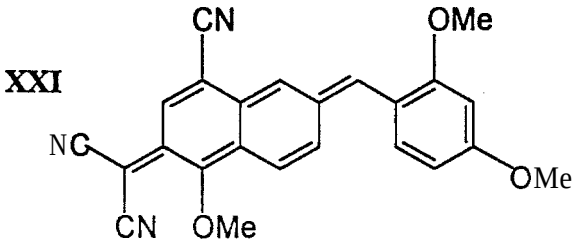
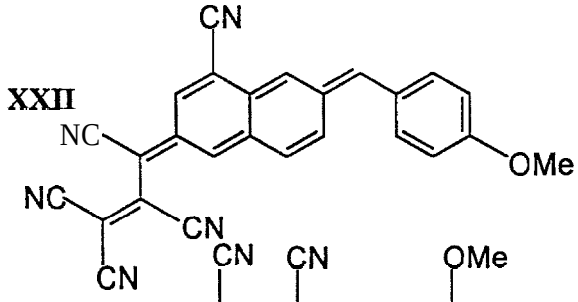
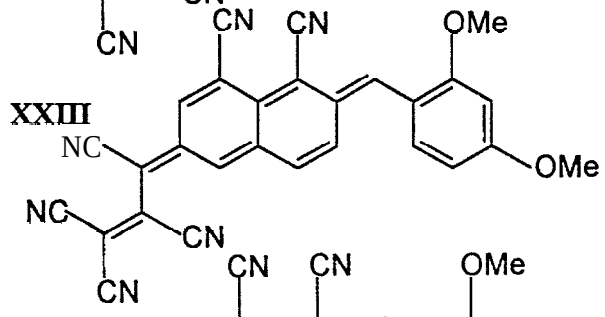
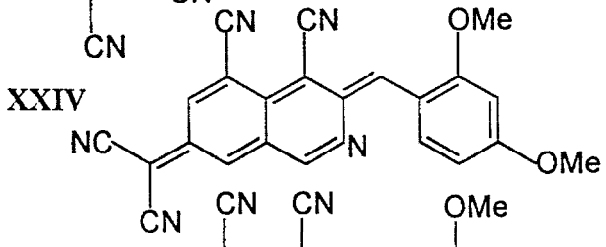
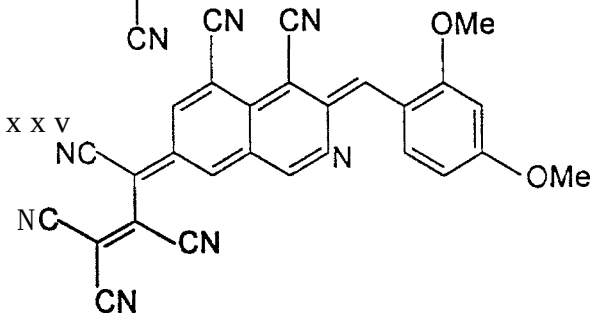
X 	5.63	152.33 + 4.73	66.46 + 1.08
XI 	7.77	227.62 +12.18	72.43 + 3.96
XII 	7.89	127.01 + 4.51	58.38 + 1.39
XIII 	8.72	195.04 + 5.26	64.58 + 2.76
XIV 	10.07	207.50 + 3.60	66.06 + 2.36
x v 	9.53	240.95 + 6.09	71.07+3.69
XVI 	11.11	157.32 + 3.08	43.49 + 0.44
XVII 	6.70	125.23 + 2.49	53.83 + 0.48
XVIII 	7.76	100.70 + 2.17	50.91 + 0.37

Table 1 (continued)

XIX 	7.28	84.62 + 3.72	58.39 + 0.56
XX 	6.50	61.15+ 1.25	44.16+ 1.74
XXI 	10.07	217.69 + 2.99	68.52 + 2.70
XXII 	10.55	360.94 + 6.57	128.59 + 9,51
XXIII 	11.41	.416.71 +13.47	144.51 +15.55
XXIV 	7.79	409.55 +18.71	122.86 + 8.79
x x v 	9.62	661.32 +61.72	230.65 +20.92

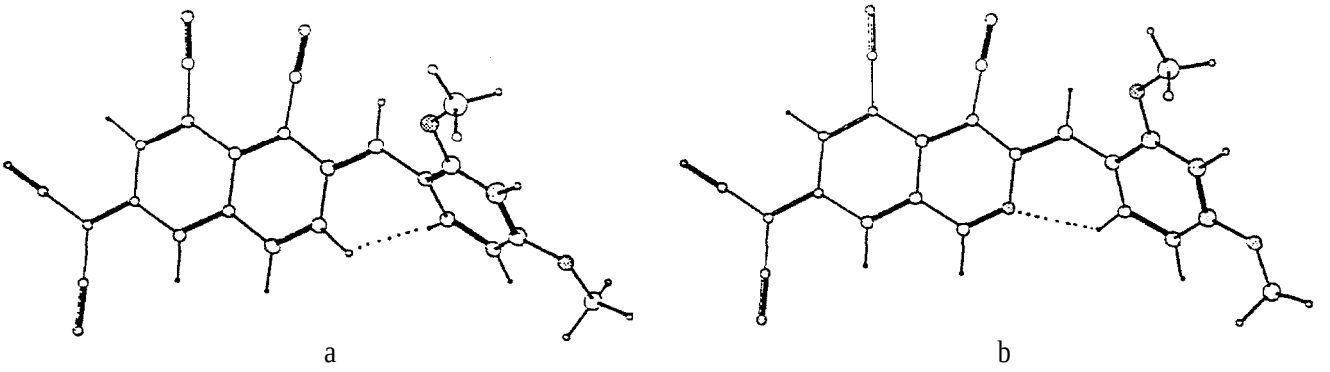


Figure 1. Molecular mechanics conformations of molecules XV (a) and XXIV (b)

The dipole moment values (m, D), second order polarizability β (10^{-51} $Cm^3 V^{-2}$) and third order polarizability y (10^{-61} $Cm^4 V^{-3}$) for molecules I-XXV are listed in Table I. According to quantum chemical calculations most of the investigated compounds have high polarizability values. The β value for molecule I is, for instance, more then ten times larger than for *o*-methoxydicyanovinylbenzene, DIVA [9]. Nevertheless β values vary with the number and position of donor and acceptor substituents. According to the calculations the β value grows at least 30% when an additional acceptor substituent is introduced at the 4- or 6-positions (molecules III, IV, VIII, IX, XIII, **XIV**), and significantly more when both these positions are substituted with acceptor groups (molecules V, XI, XV). On the contrary, introduction of a CN-group in methylidene bridge (molecules 11, VII, XII), do not influence or even decrease the β values. Number and positions of donor substituents appear to be not so significant for β values (compare, for instance, molecules III, VIII, and **XIII**). Molecules with large number of donor and acceptor substituents (XVII-XX) do not have high polarizabilities values, and we suppose-that that is partly due to the less planar conformations of these molecules, and partly to the unfavorable substituents positions.

It is remarkable, that introduction of donor and acceptor substituents in the 1,4-positions of naphthoquinone fragment (molecule XXI) also increase the β value. We found several substituent positions which corresponds to high β values for naphthoquinone derivatives. It is clear, that more effective donor and acceptor substituents for the same "core" molecule might significantly enlarge polarizability values. For instance if OMe- and CN-groups in the 1,4-positions (molecule XXI) were exchanged for NH_2 and NO_2 groups, the β value will be equal to 295. 13±8 .4910-51 $Cm^3 V^{-2}$. More significant effect was found for molecules XXII and XXIII, where a dicyanomethylidene fragment was substituted with a longer cyanopolyene fragment. The other increase of the polarizability was found for the quinoline derivatives (molecules XXIV, XXV) where the substituent positions are the same as in efficient naphthoquinone derivatives.

It have been pointed out that molecules with high values of dipole moments are not inclined to form acentric crystal structures [16]. So for dimer modeling we chose molecule IX, with a medium value of dipole moment. It was found that a dimer with C_2-symmetry has lower energy than a dimer with C_i-symmetry. This result seems to us to be promising and we arc planning to calculate dimer energies for other molecules with high polarizabilities values.

Acknowledgments

This project was supported by NASA Alliance for Nonlinear Optics, Grant NASW-4078, and NASA Cooperative Agreement NCC8- 144.

References

1. Zyss, J. Molecular Nonlinear Optics, Acad. Press. Inc. **1994**, 479 p.
2. Kanis, D. R.; Ratner, M.A.; Marks, T.J. *Chem. Rev.* 1994,94, 195.
3. Long, N.J. *Angew. Chem. Int. Ed. 1995,34,21.*
4. Dewar, M.J.S.; Stewart, J.J.P. *Chem. Phys. Lett. 1984,* 111,416.
5. Kurtz, H. A.; Stewart, J. J. P.; Dieter, K.M. *J. Comput. Chem. 1990,* 11, *82.*
6. Cardelino, B.H.; Moore, C. E.; Stickel, R.E. *J. Phys. Chem. 1991,95, 8645.*
7. Itoh, Y.; Oono, K.; Isogai, M.; Kakuta, A. *Mol. Cryst. Liq. Cryst.* 1989, 170,259.
8. Itoh, Y.; Oono, K.; Isogai, M.; Kakuta, A. *SPIE, Nonlinear Optical Materials. 1988, 1017, 127.*
9. Antipin, M.Yu., Barr T. A., Cardelino, B,, Clark, R. D., Moore C. E., Myers, T., Penn, B., Romero, M., Sangadasa, M., Timofeeva, T.V. *J. Phys. Chem. B, 1997,* 101,2770.
10. Allinger, N. L., Yuh, Y, H., Lii, J.-H. *J. Amer. Chem.* Sot. 1989, 111, 8551.
11. Lii, J.-H.; Allinger, N.L. *J. Amer. Chem. Sot. 1989,* 111, 8566.
12. Cardelino, B.H., Moore; C. E., Frazier, D.O. *J. Phys. Chem. A, 1997, 101,2207.*
13. Saunders, M. *J. Amer. Chem. Sot, 1987, 109,3150.*
14. Zyss, J, J. *Chem. Phys., 1979,70,3341.*
15. Meyers, F., Marder, S. R., Pierce, B. M., Bredas, J.L. *J. Amer. Chem. Sot., 1994, 116, 10703.*
16. Dalton, L. R., Harper, A.W., Robinson, B.H. Proc. *Nat. Acad. Sci. USA, 1997,94, 4842.*

EFFECT OF HALOGENATION ON THE NONLINEAR OPTICAL PROPERTIES OF PORPHYRIN AND SUBSTITUTED PORPHYRINS

Beatriz H. Cardelino[a], Craig E. Moore[b], Angela Benloss[a], Albert N. Thompson, Jr.[a], Rosalie A. Richards[a], Celeste A. Roney[a], Mohan Sanghadasa[c]

a) Chemistry Department, Spelman College
b) Space Science Laboratory, NASA Marshall Space Flight Center
c) Physics Department, University of Alabama in Huntsville

Abstract

The effect that fluorine and chlorine substitution has on the nonlinear optical properties of porphyrin, tetramethylporphyrin and tetraphenylporphyrin has been theoretically studied. The calculations of nolinear optical properties have been obtained by performing finite-field calculations on structures determined by semiempirical methods. In addition, tetra(p-chloropheny l)porphyrin and tetra(p-bromophcnyl)porphyrin were synthesized by the condensation of pyrrol and the appropriate aldehyde. Thin films of polymethylmethacrylate were obtained containing these materials, by spin coating onto glass substrates. The films were characterized by third-harmonic generation. It was determined that the experimental conditions enhance the third-order polarizability of the tetraphenylporphyrins by a factor of about 1.6.

Introduction

Porphyrins are an important class of cyclic conjugated tetrapyrrole molecules, essential constituents of a number of biological systems. Similarly to other tetrapyrrole molecules (e.g., phthalocyanines) porphyrins have potential use as organic nonlinear optical materials, particularly in films and polymers [1]. Of particular importance for NASA's mission has been the experiments by Debe et al. [2] that have shown that these heterocyclic polynuclear are excellent candidates for film formation by vacuum deposition under microgravity conditions.

The literature contains abundant information regarding porphyrins. Recent investigation analyzed the nature of the porphyrin phenyl electronic interaction in ortho, meta, and para substituted tetraphenylporphyrins [3]. The possible construction of large assemblages of porphyrins forming large channels has been discussed by Abrahams et al [4]. Densely packed monolayer of porphyrin have been synthesized and characterized by Li et al. [5] showing relatively large second-harmonic generation. Brigaud et al. [6] have found that substituents in tetraphenylporphyrin may cause distortions of the porphyrin molecule arising in four different conformations. Metalloporphyrins have been theoretically [7-8] and experimentally [9-11] studied.

In this investigation, the effect of halogen substitution in porphyrin, tetramethyl- and tetraphenylporphyrin has been investigated. Tetra(p-chloropheny l)porphyrin and tetra(p-bromophenyl)porphyrin were synthesized by the condensation of pyrrol and the appropriate aldehyde. Thin films of polymethylmethacrylate were obtained containing these materials, by spin coating onto glass substrates. The films were characterized by third-harmonic generation.

Method

A) Theoretical calculations. Fully optimized structures were obtained using the AM1 Hamiltonian implemented in MOPAC [12]. A modified version of the MOPAC program was used to obtain information on polarization versus static electric fields. As has been described in references [13] and [14], the above data has been fitted into polynomial expansions of variable degrees to obtain the second- and third-order polarizabilities (β and γ, respectively) and their associated numerical uncertainties. It has been previously shown that trends may be accurately obtained irrespective of the method for obtaining the molecular structure [15].

B) Synthesis. Pyrrole, 4-chlorobenzaldehyde and 4-bromobenzaldehyde were purchased from Aldrich Chemical Company. Pyrrole was distilled under reduced pressure before use. Tetra(p-chlorophenyl)porphyrin and tetra(p-bromopheny l)porphyrin were prepared by the Adler-Longo method in the following manner: Equimolar quantities of pyrrole (0.0 15 mol) and the appropriate aldehyde were refluxed in propionic acid for 1h under ambient conditions [16]. The resulting black solution was cooled to room temperature, then further cooled on an ice bath. The insoluble product was collected by suction filtration and washed with small aliquots of propionic acid followed by diethyl ether. The crystals were air dried.

C) Measurements. The third-order nonlinearity of tetra(p-chlorophenyl)porphyrin and tetra(p-bromo)phenylporphyrin were investigated using third-harmonic generation (THG) in thin films [17]. Solutions of each compound in HPLC grade chloroform (Fisher Scientific) were prepared with the concentrations in the range of 0.1 M to 1.0M. Any undissolved particles were removed by filtering each solution using Acrodisc CR PTFE syringe filters with pore sizes of 0.45 μm and 0.2 μm, and at each stage the filtering was repeated. Such a thorough filtering was required to obtain thin films with good optical quality. Each solution was mixed with 100 g/L solutions of polymethylmethacrylate (PMMA; Kodak) with molecular weight of 350,000 in HPLC grade chloroform at 1:1 to 1:2.5 volume ratios. The thin films containing the porphyrins in PMMA were obtained by spin coating onto a 0.25 mm thick 25x25 mm polished CFQ ribbon quartz plates (Heraeus-Amersil, Inc.) which had been vacuum annealed for about 8 hours. Spin coating was carried out using various rates of rotation and duration using a spin coating system (Headway Research Inc., Model 1-EC101D-R435), and the films made at the rates of rotation of 4000 rpm for 30 s had the best optical quality and therefore they were used in the THG experiment.

The primary source of the THG experiment was a Nd:YAG laser (Spectra Physics DCR-11 with near Gaussian optics) operating in the Q-switch mode at a wavelength of 1064 nm at the pulse repetition rates between 3-10 Hz. The pulse duration of a single laser pulse was about 10 ns and the maximum energy was about 400 mJ. The 1064 nm wavelength was converted to 1907 nm wavelength by a 50 cm long Raman shifter filled with hydrogen gas at a pressure of about 7 MPa. The Raman antistokes at 1907 nm and any backscattering at 1064 nm due to Brillouin scattering were separated by using a dichroic mirror (MBP Technologies, Inc.) placed at 45 degree angle of incidence. The 1907 nm laser beam was split into two beams by a laser line beam splitter. The ratio of the intensities of these two beams was adjusted by rotating a half-wave plate. The beams were focused on a reference material (vapor deposited polydiacetylene) and the porphyrin sample which were located at an equal distance from the beam splitter. The sample was mounted on a stepper motor driven, computer controlled rotational stage (Ealing, Model 35-2500). The third harmonic light generated by the reference material and the porphyrin sample was isolated by a set of filters and detected by a high gain photomultiplier tube

(Hamamatsu, Model R4632) and amplified by a fast pre-amplifier (ORTEC Model 9301) and an amplifier (ORTEC Model 485) before being processed by gated integrators and boxcar averagers (Stanford Research, Model SR 250). The measurements were then repeated several times at the normal incidence and at each run about 2000 laser pulses were fired. The measurements were also repeated with various power levels of fundamental laser light. The thickness of the films were measured using Form-Talysurf profilometer (Rank-Pneumo). The absorption of porphyrin at 636 nm was also taken into account and appropriate corrections were made.

Results and Discussion

A) Theoretical calculations. Figure 1 depicts the placement of halogens in the porphyrin molecules studied. Positions E, F, G, and H correspond to the location of a hydrogen atom (porphyrin), methyl groups (tetramethylporphyrin) or phenyl groups (tetraphenylporphyrin). The results are summarized in Table I. Fluorine substitution was included in porphyrin, tetramethyl- and tetraphenylporphyrin, while chlorine substitution was restricted to tetraphenylporphyrin. Halogenation of the phenyl groups has always been done at the para position, except for fluorination at only the phenyl groups, in which case also pentahalogenation of the phenyl group has been considered.

The symmetry group of porphyrin is C_{2v}, in semiempirical as well as ab-initio Hartree-Fock calculations, instead of D_{2h} in density function theory calculations. This explains the non zero β for this molecule obtained. Methylation and phenylation at the meso carbons tends to increase symmetry, reflected in the lower β values. Consistently, fluoro substitution at the meso carbons give the larger β values. The largest β value was obtained for tetrafluoroporphyrin, with the fluoro substitution at these positions. In general, no substantial change in the β values can be detected in comparing fluorine versus chlorine substitutions.

The most dramatic effect with respect toy is the presence of phenyl groups in tetraphenylporphyrin. Contrary to what is seen for β, in the case of γ, chlorine has an enhancement in the property compared to fluorine. The best results were obtained for chlorine substitution at the phenyl rings.

The results are summarized in figures 2 and 3 for second- and third-order polarizabilities, respectivel y.

B) Synthesis. Shiny purple crystals were obtained, with a 15 to 17°/0 yield. The UV-visible and proton NMR data were in agreement with published results [18].

C) Measurements. Using the THG experiment, the macroscopic third-order nonlinear susceptibility, $\chi^{(3)}(-3\omega;\omega,\omega,\omega)$, of thin films of tetra(p-chloro)tetraphenylporphyrin and tetra(p-bromo)tetraphenylporphyrin in PMMA matrix with various number densities and thicknesses were measured relative to a fused silica standard. The number densities of the material in thin films varied from 1.65 x 10^{22} to 2.46 x 1022 cm^{-3} and the thickness of the thin films varied from 0.50 μm to about 1.0 pm. Although third harmonic signal generated by the thin films was measured as a function of the angle of incidence, no Maker fringes were observed because the thickness of the films were much less than the coherence length. For the comparison of their nonlinearity independent of the number density N, the average value of $\gamma^{(3)}(-3\omega;\omega,\omega,\omega)$ of each compound at the molecular level was determined assuming:

$$\chi^{(3)}(-3\omega;\omega,\omega,\omega) = \mathbf{N}\, f_{3\omega}(f_{\omega})^3 \langle\gamma^{(3)}(-3\omega;\omega,\omega,\omega)\rangle$$

where $f_{3\omega}$ and f_{ω} are local field correction factors at the third-harmonic and fundamental frequencies, respectively. In the data analysis, Lorentz-Lorenz local-field factors were used. The values of average $\gamma^{(3)}(-3\omega;\omega,\omega,\omega)$ obtained for tetra(p-chlorophenyl)porphyrin and tetra(p-bromo)phenylporphyrin were 1.50 x 10-59 and 1.16 x 10-59 C m^4 V^{-3} (1.21 x 10-34 and 0.94 x 10-34 esu), respectively.

D) Comparison of experimental and calculated values. The value for tetra(p-chlorophenyl) porphyrin of 1.50 x 10-59 C m^4 V^{-3} maybe compared with the value of (67 ±16) x 10^{-61} C m^4 V^{-3} for compound 57 in Table I. There are three main reasons for the difference between these two numbers: core, environmental, and dispersion effects. Core corrections were obtained in reference [13] for carbon and nitrogen. Disregarding a core correction for chlorine, the y value for tetra(p-chlorophenyl)porphyrin including core effects is (84 ± 17) x 10^{-61} C m^4 V^{-3}. In addition, it was shown from a study on dicyanovinylbenzenes [15], Schiff's bases [19], and organic liquids [13] that adjustment factors may be obtained from comparison between calculations and measurements to account for the experimental conditions (environmental and dispersion effects) within a family of compounds. Thus, for tetraphenylporphyrins, a factor of 1.8 ± 0.4 is obtained from the chloro-substituted molecule. A calculation on tetra(p-bromophenyl)porphyrin results in a factor of 1.4+ 0.4. It seems valid to say that the experimental conditions of this study enhance the third-order polarizabilities of tetraphenylporphyrins by a factor of about 1.6. The corresponding value for the second-order polarizability (β) of dicyanovinylbenzenes is 2.0, for β of the Schiff's bases is 8.7, and for the third-order polarizability (γ) of organic solvents is 2.2.

Conclusions

The higher values of β were obtained for perfluorinated structures, while y values were higher for chlorinated structures. Comparison between experimental and calculated values suggests that the solvent and dispersion effects of the present measurements enhance the third-order polarizabilities (y) by a factor of about 1.6. Push-pull substituents would enhance the hyperpolarizabilities of porphyrin, tetramethylporphyrin and tetraphenylporphyrin more substantially than halogenation.

References

[l] J. L. Bredas, C. Adant, P. Tackx, A. Persoons; Chem.Rev. 94,243 (1 994), and references therein.

[2] M.K.Debe, R.J.Poirier, D.D.Erikson, T.N.Tommet, D.R.Field, K.M.White; Thin Solid Films 186,257 (1990).

[3] A. Gosh; J.Molec.Strut.(Theochem) 388,359 (1996).

[4] B.F.Abrahams, D.F.Hoskins, D.M.Michail, R.Robson; Nature 369,727 (1994).

[5] D. Q.Li, B.I.Swanson, J.M.Robinson, M. A.Hoffbauer; J. Am.Chem.Soc.1 15,6975 (1993).

[6] O.Brigaud, P. Battioni, D.Mansuy, C. Giessner-Pretre; New.J.Chem.16, 1031 (1992).

[7] M.Matsuzawa, M.Ata, D.A.Dixon; J.Phys.Chem.99, 7698 (1 995).

[8] W.Jentzen, M. C. Simpson, J.D.Hobbs, X. Song, T.Ems, N. Y.Nelson, C.J.Medforth, K. M. Smith, M. Veryat, M.Mazzanti, R. Ramasseul, J. C. Marchon, T. Takeuchi, W. A. Goddard 111, J. A. Shelnutt; J.Am.Chem.Soc.1 17, 11085 (1995).
[9] A.Sen, P. C.Ray, P.K.Das, V.Krishnan; J.Phys.Chem.100, 19611 (1 996).
[10] D. Beljonne, G.E.O'Keefe, P.J.Hamer, R.H.Friend, H.L.Anderson, J.L.Bredas; J. Chem.Phys. 106,9439 (1997).
[11] M. Hosoda, T. Wada, H. Sasabe; Mol. Cryst. & Liq.Cryst.Sci.Technol.Sect.B: Nonlinear Opt.7, 199 (1994).
[12] QCPE Program 455 version 6.1 (1990).
[13] B.H.Cardelino, C.E.Moore, D. O.Frazier; J.Phys.Chem.A 101,2207 (1997).
[14] B.H.Cardelino, C.E.Moore, R.E.Stickel; J.Phys.Chem.95, 8645 (1991).
[15] M. Y.Antipin, T.A.Barr, B. H. Cardelino, R.D.Clark, C.E.Moore, T.Myers, B. Penn, M. Romero, M. Sanghadasa, T.V. Timofeeva; J. Phys.Chem.B 101,2770 (1997).
[16] A. Adler, F. R. Longo, J. D. Fenarelli, J. Goldmacher, J. Assour, L. Korsakoff; J. Org. Chem. 32,476 (1 967).
[17] M. Sanghadasa, B. Wu, R. D. Clark, H. Guo, B. G. Penn; SPIE Proc. Vol. 3147,185 (1997).
[18] H. Scheer, J. Katz; in Porphyrins and Metalloporphyrins; K. M. Smith, cd.; Elsevier: NY, p. 399 (1975).
[19] K. Bhat, J. Choi, S.D. McCall, M.D. Aggarwal, B.H. Cardelino, C.E. Moore, B.G. Penn, D. O. Frazier, M. Sanghadasa, T. A. Barr, N, B. Laxmeshwar; Comp.Mat.Sci. 8,309 (1 997),

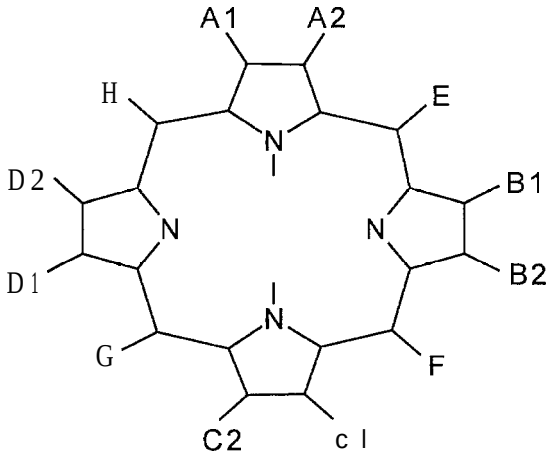


Figure 1- Substitution code for porphyrin, tetramethylporphyrin and tetraphenylporphyrin.

Table I - Properties of halogenated porphyrin (P), tetramethylporphyrin (TMP) and tetraphenylporphyrin (TPP). Please refer to Figure 1 for the substitution code. Substitution codes E, F, G, H, imply that all H atoms at the corresponding location have been substituted (three methyl hydrogens in TMP, and five phenyl hydrogens in TPP) while pE, pF, pG and pH imply a para substitution in the corresponding phenyl ring of TPP.

Number	μ (D)	α (Å3)	β (10^{-51} C m^2 V^{-3})	<μ•β>	γ (10^{-61} C m^3 V^{-4})	Substitution code
PORPHYRIN (P)						
1	**1.06**	41.50	23.60 ± 0.01	-0.28	19.74 ± 0.03	**None**
	Fluorine substitution					
2	**2.55**	**43.92**	**86.21 ± 0.13**	-0.18	31.00 ± 0.30	E,F,G,H
3	**0.43**	**43.78**	13.81 ± **0.00**	0.00	21.55 ± 0.04	B1,B2,D1,D2
4	**1.07**	**43.05**	**22.44 ± 0.01**	-0.13	21.87 ± 0.06	A1,A2,C1,C2
5	**0.40**	**45.37**	**10.76 ± 0.00**	-0.03	23.74 ± 0.08	A1,A2,B1,B2, C1,C2,D1,D2
6	**1.35**	**47.48**	**30.57 ± 0.01**	-18.31	**27.84 ± 0.21**	A1,A2,B1,B2, C1,C2,D1,D2, E,F,G,H
TETRAMETHYLPORPHYRIN (TMP)						
7	0.30	48.47	8.66 ± 0.01	-4.85	**22.28 ± 0.13**	**None**
	Fluorine substitution					
8	**2.26**	**48.13**	**41.69 ± 0.04**	-17.20	**22.42 ± 0.09**	E,G
9	2.81	47.91	**48.15 ± 0.06**	-11.36	**22.11 ± 0.03**	E,F,G,H
10	**2.07**	**51.07**	**28.18 ± 0.02**	-19.17	**23.88 ± 0.07**	A1,A2,B1,B2, C1,C2,D1,D2, E,F,G,H
11	1.26	50.71	13.47 ± 0.01	-9.52	**27.34 ± 0.25**	A1,A2,B1,B2, C1,C2,D1,D2
12	6.86	48.15	30.39 ± 0.02	-76.21	**21.58 ± 0.06**	E,F
13	5.09	47.31	44.72 ± 0.02	-49.46	**21.37 ± 0.01**	E,G
TETRAPHENYLPORPHYRIN (TPP)						
14	0.31	84.37	15.59 ± 0.78	-6.79	**53.94 ± 5.00**	**None**
15	0.20	82.43	2.90 ± 0.03	-13.07	**40.32 ± 1.06**	**None**
16	0.21	82.44	2.18 ± 0.03	-1.66	**40.30 ± 1.06**	**None**
17	0.20	82.43	2.14 ± 0.02	-20.99	**40.31 ± 1.06**	**None**
	Fluorine substitution on pyrrol rings					
18	**1.71**	**84.43**	10.05 ± 0.21	**71.30**	**50.88 ± 2.36**	B1
19	2.01	84.49	15.90 ± 0.06	**14.75**	**47.65 ± 2.17**	A1, B1
20	0.14	84.83	5.21 ± 0.18	**-86.75**	**49.64 ± 2.00**	B1 , D1
21	1.57	84.73	11.32 ± 0.11	**77.01**	**47.00 ± 1.20**	A1,B1,D1
22	0.84	83.95	19.73 ± 0.25	**-31.52**	**48.82 ± 1.38**	A1,B1,C1,D1
23	1.17	84.37	18.23 ± 0.13	**6.08**	**50.29 ± 2.21**	A1
24	0.89	84.15	19.79 ± 0.31	**-27.87**	**49.27 ± 1.89**	A1, **Cl**
25	2.25	84.15	15.46 ± 0.22	**-40.18**	**48.94 ± 1.64**	A1,B1,C1
	Fluorine para substitution on phenyl rings					
26	**2.19**	**84.93**	**15.13 ± 1.50**	**-36.92**	**55.05 ± 3.43**	pE
27	**2.85**	**85.55**	**16.56 ± 0.63**	**-71.73**	**55.62 ± 6.36**	pE,pF

28	0.41	85.56	16.23 ± 1.57	-15.03	56.08 ± 3.97	pE , pG
29	1.65	86.18	16.46 ± 1.78	54.61	57.36 ± 4.40	pE,pF,pG
30	**0.42**	**86.77**	**17.58 ± 0.75**	***0.00**	**58,46 ± 8.71**	pE,pF,pG,pH
31	**1.86**	**85.06**	**16.40 ± 1.42**	**64.88**	56.38 ± 3.70	pE
32	**0.41**	**85.56**	**16.16 ± 1.57**	**-15.78**	**56.25 ± 4.02**	pE,pG
33	**1.68**	**86.29**	**17.44 ± 2.15**	**55.33**	**59.50** ± 4.99	pE,pF,pG
	Fluorine substitution on pyrrols and para substitution on phenyl rings					
34	3.54	85.03	10.66 ± 0.79	85.70	**51.87 ± 2.17**	B1,pE
35	4.67	85.50	14.25 ± 0.12	38.48	**48.93 ± 1.96**	A1,pE,B1,pF
36	0.21	85.54	8.99 ± 0.46	-17.26	**50.62 ± 1.98**	B1,pE,D1,pG
37	3.19	86.64	12.80 ± 0.37	51.14	**48.56 ± 1.22**	A1,p,B1,pF, D1,pH
38	0.80	86.43	22.87 ± 0.29	-33.19	**53.24 ± 2.68**	A1,pE,B1,pF, C1,pG,D1,pH
39	3.04	85.02	18.22 ± 0.66	33.81	**51.64 ± 2.38**	A1, pE
40	0.73	85.29	18.71 ± 0.33	-15.16	**51.12 ± 2.03**	A1,pE,C1,pF
41	3.74	85.97	18.06 ± 0.80	-57.57	**52.27 ± 2.38**	A1,pE,B1,pF, **Cl**, pG
	Fluorine substitution on phenyl rings					
42	**3.52**	**85.89**	**34.06 ± 0.07**	70.09	**49.40 ± 1.98**	**E**
43	5.21	87.89	28.22 ± 0.02	75.31	**49.48 ± 2.27**	E,F
44	4.63	89.71	19.18 ± 0.56	-18.57	**48.46 ± 0.97**	E,F,G
45	2.49	87.86	65.37 ± 0.04	-76.21	**41.73 ± 0.28**	E,F,G,H
	Chlorine substitution on pyrrol rings					
46	**1.28**	**85.32**	**15.22 ± 0.15**	**-87.98**	52.01 ± 2.36	B1
47	**1.39**	**86.37**	**17.56 ± 0.04**	**55.35**	**50.58 ± 1.57**	A1,B1
48	0.21	86.53	6.50 ± 0.44	25.77	53.96 ± 2.50	B1,D1
49	1.12	87.54	8.22 ± 0.27	24.14	**52.50 ± 1.77**	A1,B1,D1
50	0.75	87.30	15.56 ± 0.14	-46.86	**53.94 ± 1.74**	A1,B1,C1,D1
51	0.61	85.34	18.10 ± 0.11	24.07	**51.68 ± 2.66**	**Al**
52	1.51	86.71	14.67 ± 0.33	-83.55	**56.89 ± 3.09**	A1,B1,C1
	Chlorine para substitution on phenyl rings					
53	**1.88**	**86.23**	**16.69 ± 1.94**	**-33.15**	**58.78 ± 4.53**	pE
54	**2.47**	**88.15**	**18.27 ± 0.76**	**-66.95**	**61.08 ± 9.54**	pE , pF
55	0.41	88.03	16.48 ± 2.27	-21.58	59,53 ± 5.46	pE , pG
56	1.40	90.06	17.33 ± 3.48	59.70	64.53 ± 7.14	pE,pF,pG
57	0.40	91.87	19.70 ± 1.51	-12.79	67.14 ±15.82	pE,pF,pG,pH
58	1.62	86.33	16.13 ± 1.75	68.69	58.06 ± 4.27	pE
59	0.43	88.15	17.39 ± 2.74	-30.75	61.77 ± 6.25	pE , pG
60	1.40	90.05	17.32 ± 3.55	61.01	64.44 ± 7.16	pE,pF,pG
	Chlorine substitution on pyrrol and phenyl rings					
61	**2.87**	**87.05**	**16.60 ± 0.96**	**-69.86**	**54.39 ± 2.29**	B1,pE
62	**3.75**	**89.92**	**17.52 ± 0.03**	**85.52**	**53.98*2.73**	A1,pE,B1,pF
63	**2.50**	**92.97**	**8.87 ± 0.57**	**38.64**	**58.16 ± 2.30**	A1,pE,B1,pF, D1,pH
64	**0.81**	**94.61**	**19.21 ± 0.41**	**-43.11**	**64.26 ± 5.58**	A1,pE,B1,pF, C1,pG,D1,pH
65	3.04	85.01	18.22 ± 0.66	33.81	51.64 ± 2.38	**Al**, pE
66	2.66	92.29	16.47 ± 0.92	76.77	65.09 ± 4.61	A1,pE,B1,pF, **Cl**, pG
67	1.03	89.34	21.78 ± 0.48	-25.17	58.79 ± 2.84	A1,pE,C1,pG
68	0.31	90.13	8.94 ± 1.16	-63.21	57.91 ± 2.90	A1,B1,pE,pF

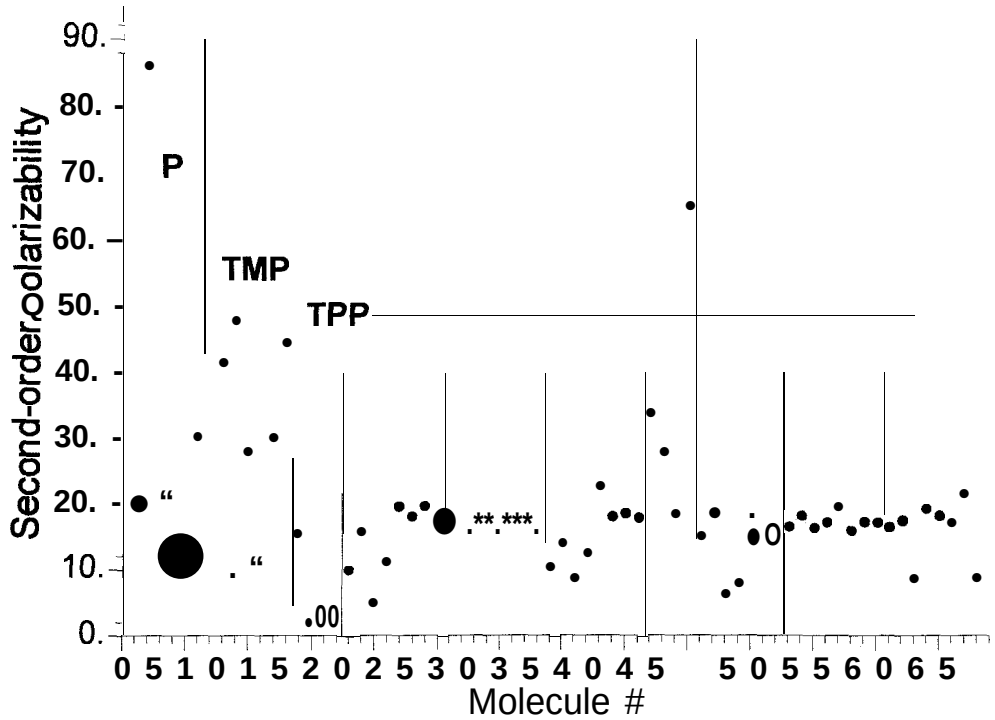


Figure 2- Static **molecular second-order** polarizabilities (β) **of selected** unsubstituted **and** halogenated **porphyrin (P),** tetramethylporphyrin (TMP) **and** tetraphenylporphyrin (TPP). **Please refer to** Table **I for the molecule numbers.**

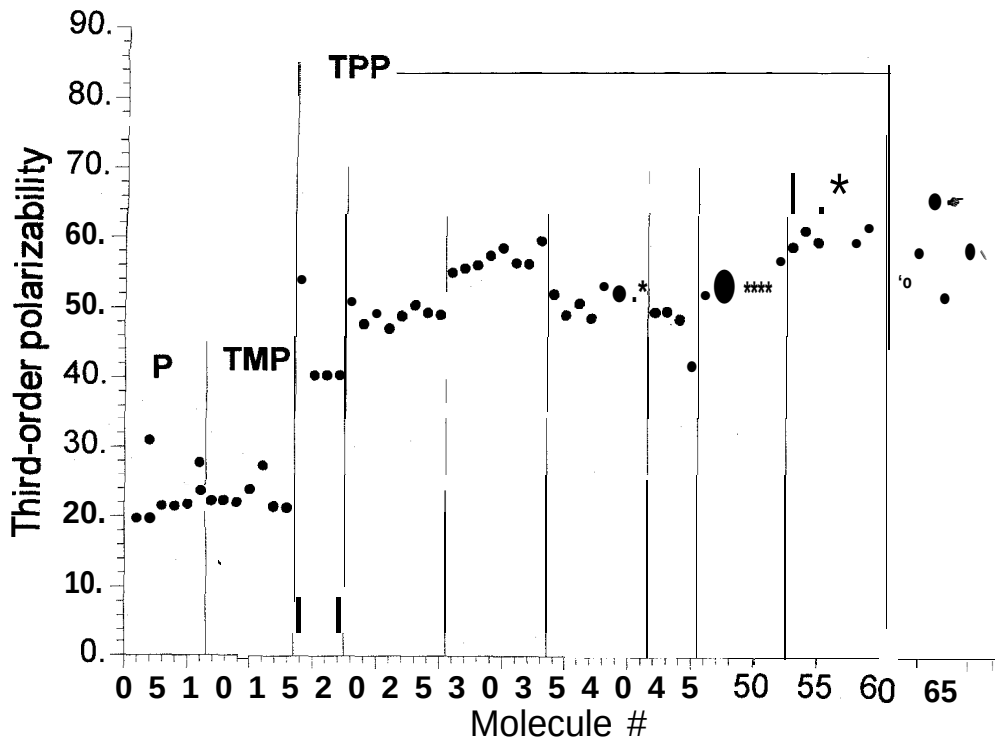


Figure 3- **Static molecular third-order** polarizabilities (γ) **of selected** unsubstituted **and** halogenated **porphyrin (P),** tetramethylporphyrin (TMP) **and** tetraphenylporphyrin (TPP) **Please refer to Table I for the molecule numbers.**

Synthesis of New Heterocyclic Imidazoles for Nonlinear Optical Chromophores

Haiying Li,[a] Javier Santos,[a] Donald Van Derveer,[b] Eric A. Mintz,[a] Xiu R. Bu[a]

a: Department of Chemistry & Center for High Performance Polymers and Composites, Clark Atlanta University, Atlanta, GA 30314

b: School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332

Abstract; A series of novel NLO imidazole chromophores containing tricyanovinyl acceptor coupled with thiophene have been developed.

The heterocyclic compounds such as five-membered ring heterocyclic azole derivatives have been used as nonlinear optical (NLO) chromophores since these compounds exhibited excellent thermal stability in guest-host systems,[1-5] excellent volubility in common organic solvents and outstanding miscibility with high performance polymers (up to 5 0°/0 without phase separation).[6] We have recently reported the incorporation of NLO imidazole chromophores into polyamides leading to NLO side chain polymers.[7] Previous effort to further enhance the molecular nonlinearity has only received limited success. Attempts to optimized the conjugation pathway by maximizing the coplanarity of aromatic rings at 4,5 position of the ring have resulted in low volubility due to high rigidity caused by highly fused aromatic rings.[8]

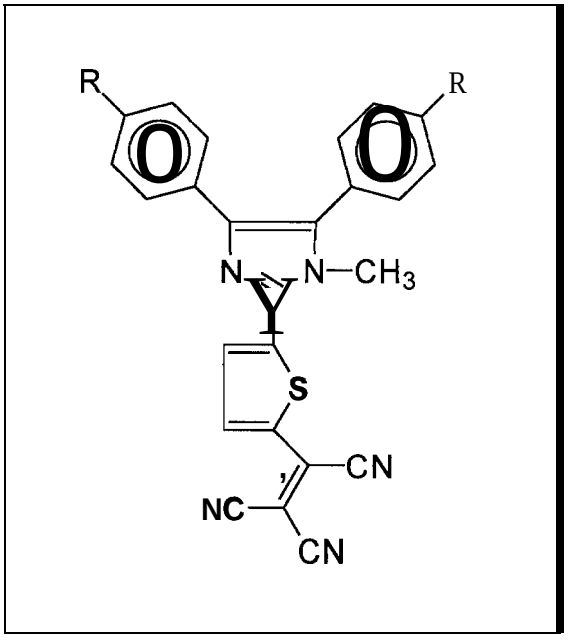


Figure 1. The target molecular structures which contain thermally stable heterocyclic imidazoles and efficient electron acceptors-tricy anovinylthiophenes

Our novel approach to the design of a new imidazole system **(Fig 1)** with enhanced molecular nonlinearity and excellent volubility is based upon three principles: (1) the two separate aromatic rings in 4,5 position of the imidazole ring should interact sterically leading to poor crystal packing, therefore, to the enhancement of volubility; (2) the steric effect from these two aromatic rings will also minimize the dipole-dipole interaction between chromophores, (3) the incorporation of highly efficient tricyanovinyl acceptors coupled with thiophenes should give the chromophores with superior molecular nonlinearity. Previous studies revealed that tricyanovinyl acceptors are far superior to dicyanovinyl acceptors by a factor of 2-3 in enhancing molecular nonlinearity,[9] and that tricyanovinyl acceptors incorporated with thiophenes further dramatically enhance molecular nonlinearity by a factor of 4-5 compared with dicyanovinyl acceptors, and by a factor of 10-11 compared with nitro acceptors.[10]

Two synthetic routes have been widely used in the preparation of tricyanovinyl derivatives: direct reaction of tetracyanoethylene with an activated aromatic ring or stepwise conversion starting with an aldehyde (from aldehyde first to dicyanovinyl by condensing with malonitrile, then to tricyanovinyl by nuclophilic addition to dicyanovinyl group by sodium cyanide, followed by oxidation with $Pb(OAc)_4$). However, these routes are inapplicable with imidazole-containing thiophenes. We present here a novel approach to tricyanovinylation of imidazole-containing thiophenes.

The total synthesis of novel chromophore 3 using the new methodology was achieved in four steps, as illustrated in Scheme 1, starting with commercially available and inexpensive 4,4'-dimethoxybenzoin. The benzoin was converted into 4,4' -dimethoxylbenzil using cupric acetate. The heterocyclic compound 2-(2' -thienyl)-4,5-bis(4'-methoxyphenyl)imidazole, **1,** was prepared in 94°/0 yield by treatment of 4,4' -dimethoxyl)benzil and 2-thiophenecarboxyaldehyde with ammonium acetate in glacial acetic acid. The alkylation of **1** with methyl iodide in the presence of potassium carbonate in DMAc led to the formation of 1-methyl-2-(2'-thienyl)-4,5-bis(4'-methoxyphenyl)) imidazole 2 in a yield of 71 O/O. This alkylation at the 1-position of the azole ring is necessary to ensure the formation of a monoanion species of thiophene (thienyllithium) in the following step. Treatment of 2 with n-butyl lithium in THF-hexane in-situ generated the red anion solution at low temperature, followed by the addition of tetracyanoethylene, leading to the formation of the desirable 1-methyl-2 -[2' -(5' -tricyanovinyl)thienyl] -4,5 -bis(4'-methoxyphenyl)-imidazole 3 in 65 % yield.

This synthetic route has been extended to other substituents in 4,5 aromatic rings that chromophores 7 (R=H), 8 ($R=CH_3$), 9 ($R=NMe_2$) have been prepared in the same fashion (Scheme 2). All of these new chromophores are very soluble in most common organic solvents such as THF, chloroform and methylene dichloride. The general molecular structure of this new series of chromophores has been determined by x-ray crystallographic study of **8(Fig** 2). The two aromatic rings at the 4 and 5 positions of the azole ring sterically interact leading to a dihedral angle of 60.32 °C where one aromatic ring twists more than the other relative to imidazole ring (56.5° vs. 11.69°).

In comparison with other imidazole based chromophores, these new chromophores exhibit large bathochromic shift **(Fig 3),** which seems to be dictated by strength of the acceptors. For example, compound 3 is red-shifted by nearly 100 nm compared to compound **10,** which has a nitro acceptor group on the thiophene, and by over 150 nm compared to compound **11,** which has a nitro acceptor group on phenyl. This trend suggests that the new chromophores will exhibit much larger molecular nonlinear response by an increase in first hyperpolarizability.

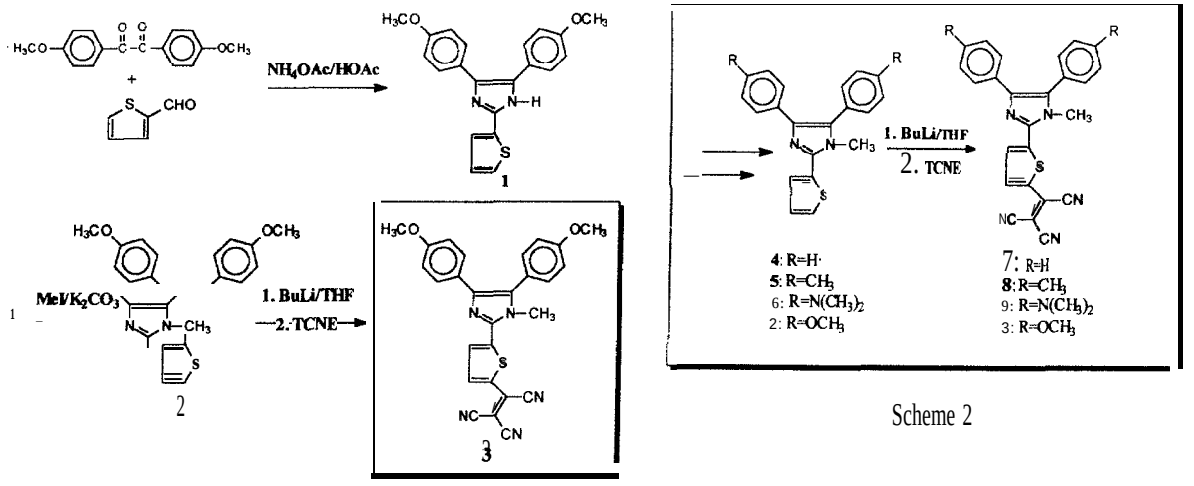
In conclusion, a novel tricyanovinylation has been developed that permits the direct synthesis of a series of novel heterocyclic imidazole NLO chromophores by reaction of tetracyanoethylene with thienyllithium of imidazoles. Effort are in progress to extend this new methodology to prepare other important NLO chromophore containing tricyanovinyl acceptors.

Acknowledgment

Financial support from NASA(NAGW-2939) is gratefully acknowledged.

REFERENCES

1. Carter, K. R.; Miller, R. D.; Hedrick, J. L. *Macromolecules, 1993,26,2209.*
2. Moylan, C. R.; Miller, R. D.; Twieg, R. J.; Betterton, K. M.; Lee, V. Y.; Matray, T. J.; Nguyen, C. *Chem. Mater.* 1993,5, 1499.
3. Twieg, R. J.; Betterton, K. M.; Burland, D. M.; Lee, V. Y.; Miller, R. D.; Moylan, C. R.; Volksen, W.; Walsh, C.A. *SPIE.* 1993, vol. *2025,94.*
4. Meinhardt, M. B.; Cahill, P. A.; Seager, C. H.; Beuhler, A. J.; Wargowski, D. A. *Mat. Res. Sot. Symp. Proc.* 1994, Vol. 328,467.
5. Reinhardt, B. A.; Kannan, R.; Bhatt, J. C. *SPIE.* 1994, *vol. 2229,24.*
6. Stähelin, M.; Burland, D. M.; Ebert, M,; Miller, R. D.; Smith, B. A,; Twieg, R. J.; Volksen, W.; Walsh, C. A. *Appl. Phys. Lett. 1992,61,* 1626,
7. Li, H. Y.; Bu, X. R.; Ahmed, M.; Mintz, E. A. *Polym. Prepr. (Am. Chem. Sot., Div. Polym. Sci.)* 1996, 37(1), 599. also see: (a) Kulig, J. B.; Moore, C. G.; Brittain, W. J. *Polym. Prepr. (Am. Chem. Sot., Div. Polym. Sci.)* 1994, 35(1), 492. (b) Carter, K. R.; Hedrick, J. L.; Twieg, R. J.; Matray, A. M.; Walsh, C. A. *Polym. Prepr. (Am. Chem. Sot., Div. Polym. Sci.) 1994,* 35(1), *529. (c)* Li, H.; Mintz, E. A.; Bu, X. R. *Polym. Prepr. (Am. Chem. Sot., Div. Polym. Sci.)* 1995, *36(2), 290.* (d) Wang, J. F.; Marks, T. J.; Lin, W. P.; Zhou, H. T.; Wong, G. K. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Sci.) 1995, 36(2), 308.*
8. see 2 and 3.
9. Katz, H, E.; Singer, K. D.; Sohn, J. E.; Dirk, C. W. King, L. A.; Gordon, H. M. *J. Am. Chem. Soc.* 1987, 109,6561.
10. Rae, V. P.; Jen, A. K-Y.; Wong, K. Y.; Drost, K. *J. Chem. Sot. Chem. Commun. 1993,* 1118.



Scheme 2

Scheme 1: A typical example for the preparation of novel chromophore containing tricyanovinylthiophene structures.

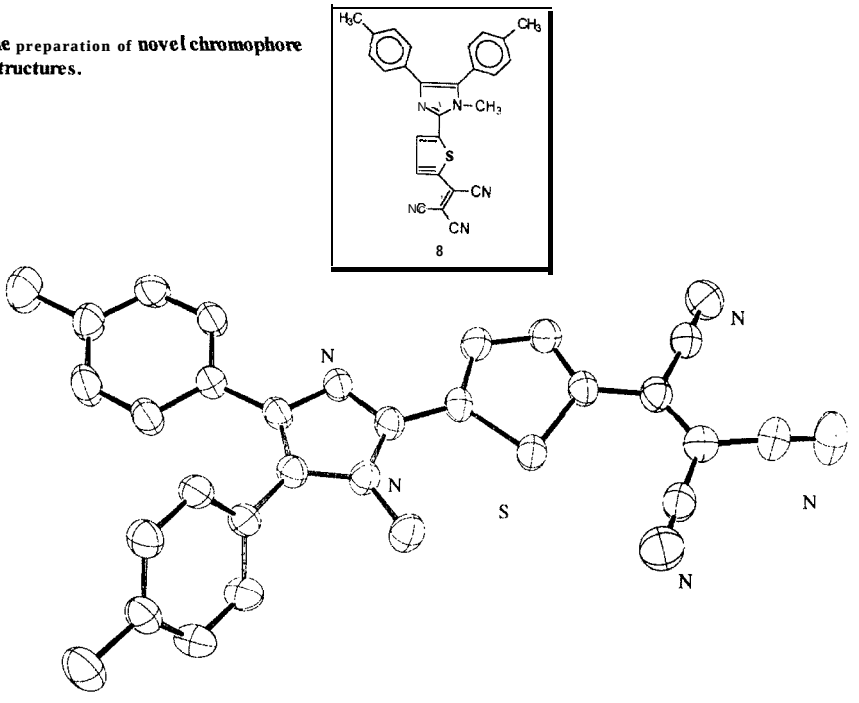
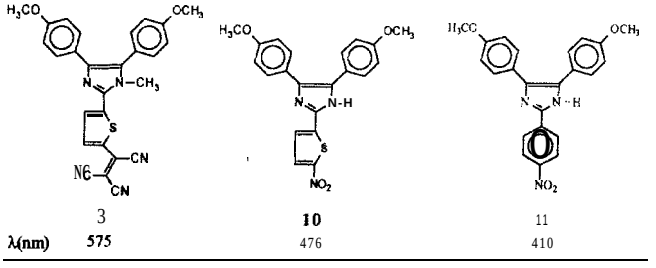


Figure 2. Molecular structure of new chromophore 8 containing tricyanovinylthiophene.



	3	10	11
λ(nm)	575	476	410

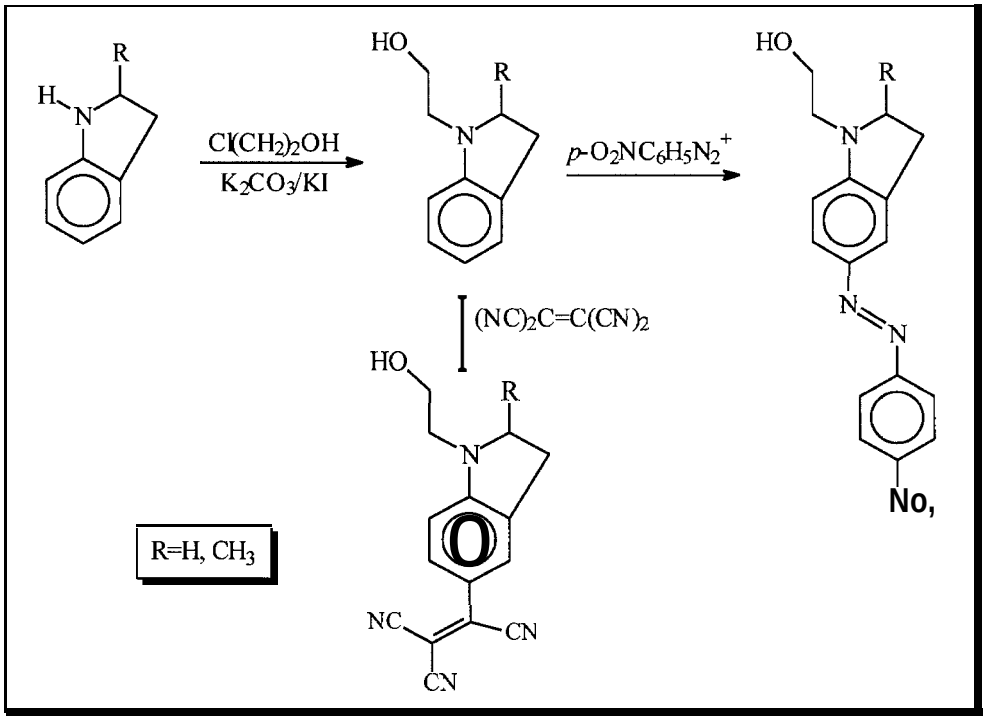
Figure 3. List of of. typical charge transfer energy of a new chromophore compared with those from other imidazole chromophores.

Development of Polyimides-Based NLO materials for Electrooptical Applications

Jacqueline Rutherford, Xiang Li, Eric A. Mintz, Xiu R. Bu*
Department of Chemistry and **Center** for High Performance polymers and Composites
Clark Atlanta University
Atlanta, GA 30314

Development of thermally stable optical materials for nonlinear optics have recently focused on the covalent incorporation of NLO chromophores into high performance polymers, especially thermally stable and processable polyimides.[1-3] One key aspect for the incorporation of robust NLO chromophores into high Tg polymers is to sustain poling induced order.[4] Other advantages include high loading level of chromophores, and elimination of possible phase separation as well as chromophore sublimation at processing or working temperature. We have prepared several polyimide based polymers which are covalently linked with thermally stable chromophores[5-7] that we have developed, since polyamides generally exhibit high Tg and good film transparency. Here, we report the development and subsequent incorporation of indoline based chromophores into polyamides, leading to thermally stable NLO polymers. Indolines are used as thermally stable donor groups because cycloamine structures are usually more stable than regular acyclic dialkylamines.

Three types of chromophores are derived from indoline by the addition of different acceptors: tricyanovinyl, nitrophenylazo, and nitro-stilbene. The reactions are illustrated in Scheme 1. The functionalized chromophores were synthesized by treating 2-chloroethanol with indoline in the presence of a base, followed by the addition of either tetracyanoethylene to give the tricyanovinyl acceptor based chromophore or 4-nitrophenyldiazonium salt to give the 4-nitrophenylazo acceptor based chromophore. The nitrostilbene based chromophores were prepared in four steps from indoline by reaction of indoline with 2-chroloethanol, followed by esterification with acetic anhydride, formulation with $POCl_3$/DMF, and condensation with 4-nitrophenylacetic acid (not shown). The polymer preparation is illustrated in Scheme 2. Treatment of hydroxyl functionalized chromophores with phenol-containing polyamides under coupling conditions gives NLO polyamides in which chromophores are covalently linked to polymer backbones.

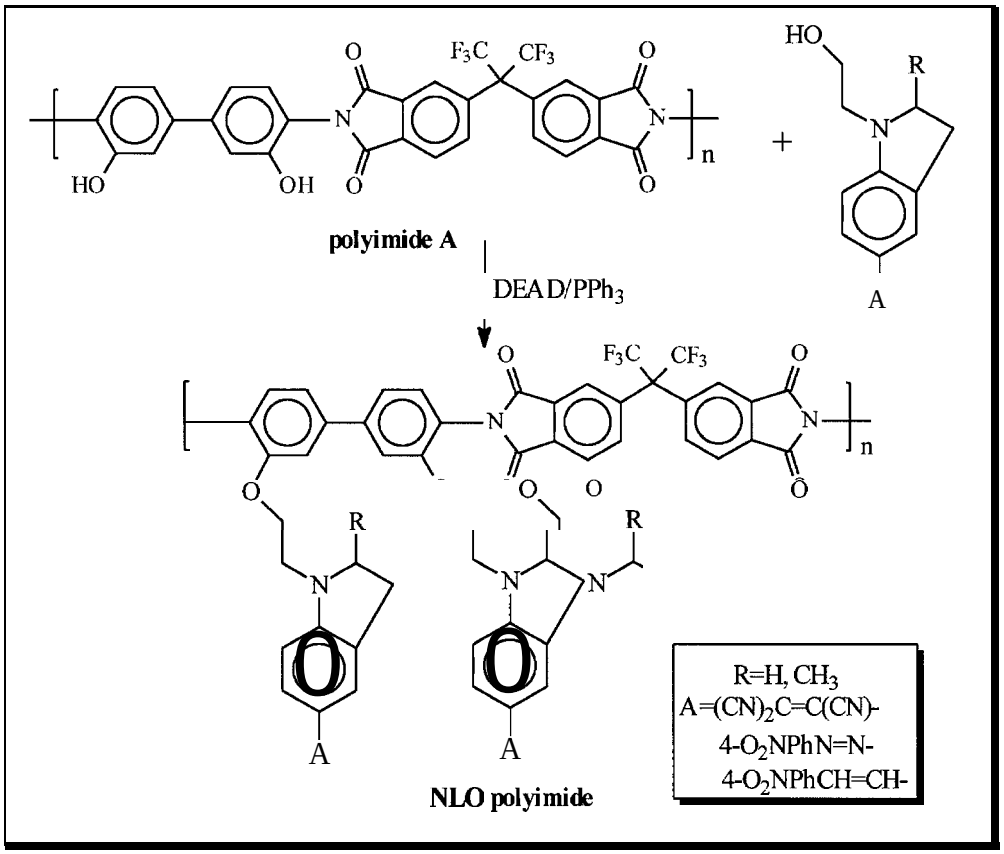


SCHEME 1 FOR SYNTHESIS OF FUNCTIONALIZED INDOLINE BASED CHROMOPHORES WHICH WILL BE ATTCHED TO POLYAMIDES

The characterization of polymers includes verification of chromophore attachment, and thermal stability analysis. Representative analyses for NLO **polyimide B** ($R=CH_3$, A=tricyanovinyl in Scheme 2) are shown by FT-IR (Fig. 1), electronic spectrum (Fig.2), 'H NMR (Fig.3), and DSC thermal analysis (Fig.4). The attachment of chromophores to the polymer backbone was confirmed by appearance of a CN band at approximately 2200 cm^{-1} in the FT-IR spectrum of the polyimide, also by a charge transfer band at 535 nm from visible spectrum of the polyimide. The loading level of chromophore was assessed by 'H NMR, in which free hydroxyl protons of phenol were monitored and compared with polyimide A (Fig. 3). More than 70% of phenol sites have been substituted with chromophores.

The thermal properties were evaluated by DSC to determine the glass transition temperature (Tg). The high Tg value of 294 ºC is very impressive considering the presence of the high loading level of chromophores. All other indoline based NLO polyamides exhibit similar high thermal stability with Tg values higher than 220 'C.

Preliminary evaluation of electrooptic coefficients of NLO polyimide B gave r_{33} values of 1 pm/v at 633 nm, 0.3 pm/v at 830 nm for the polymer film with a thickness of 0.2 pm, which was poled in a poling field of 100v/μm.



SCHEME 2 FOR SYNTHESIS OF NLO POLYAMIDES CONTAINING INDOLINE BASED CHROMOPHORES

Acknowledgment

We gratefully acknowledge the support of this research by NASA(NAGW-2939). We also thank Dr. Yue Zhang for helpful and fruitful discussion.

References

1. D. Yu, A. Gharavi, L. Yu, *Macromolecules,* **28, 784(1995).**
2. A. K.-Y. Jen, Y.-J. Liu, Y. Cai, V. P. Rae, L. R. Dalton, *Chem. Commun.* 2711(1994).
3. M. W. Becker, L. S. Sapochak, R. Ghosen, C. Xu, L. R. Dalton, Y. Shi, W. H. Steier, A. K.-Y. Jen, *Chem. Mater,* **6, 104(1994).**
4, T. Verbiest, D. M.. Burland, M. C. Jurich, V. Y. Lee, R. D. Miller, W. Volksen, Science, 268, 1604(1995).
5. X. R. Bu, G. Lai, M. Ahmed, J. Rutherford, E. A. Mintz, *Polym. Prepr., 37(2), 254(1996).*
6. H. Y. Li, X. R. Bu, M. Ahmed, E. A. Mintz, *Polym. Prepr.,* 37(1), *599(1996).*
7. X. R. Bu, H. Y. Li, D. VanDerveer, E. A. Mintz, *Tetrahedron Lett., 37(41), 733 1(1996).*

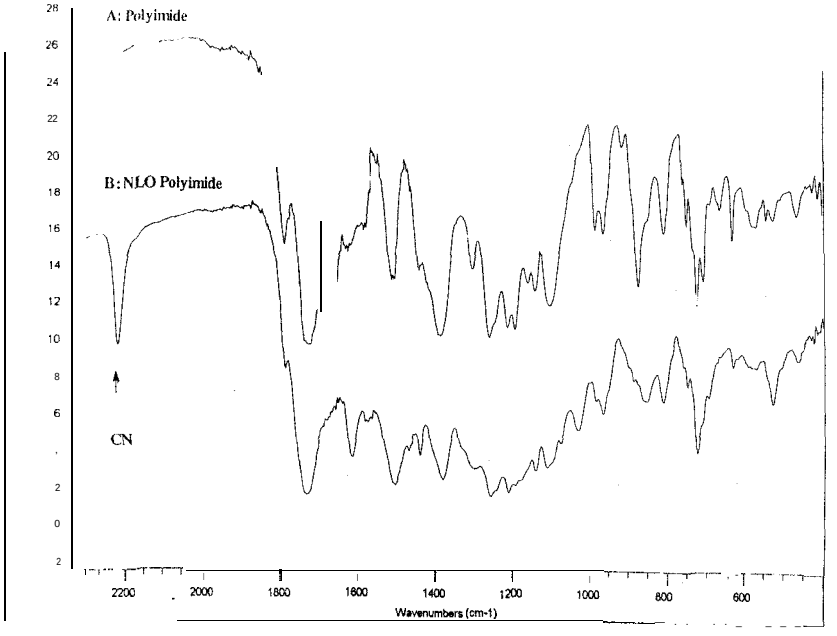


Fig. 1. FT-IR spectrum of NLO Polyimide B and starting polyimide A.

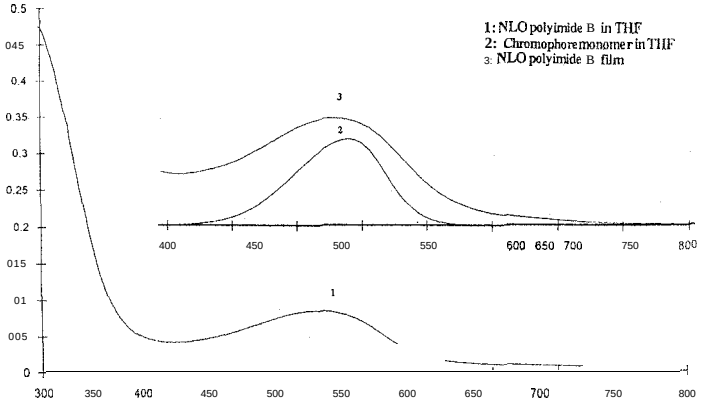


Fig.2. CHARGE TRANSFER BANDS OF NLO POLYIMIDE B IN BOTH THE SOLUTION AND SOLID STATES NOTE THE CHROMOPHORE MONOMER IS USED FOR COMPARISON

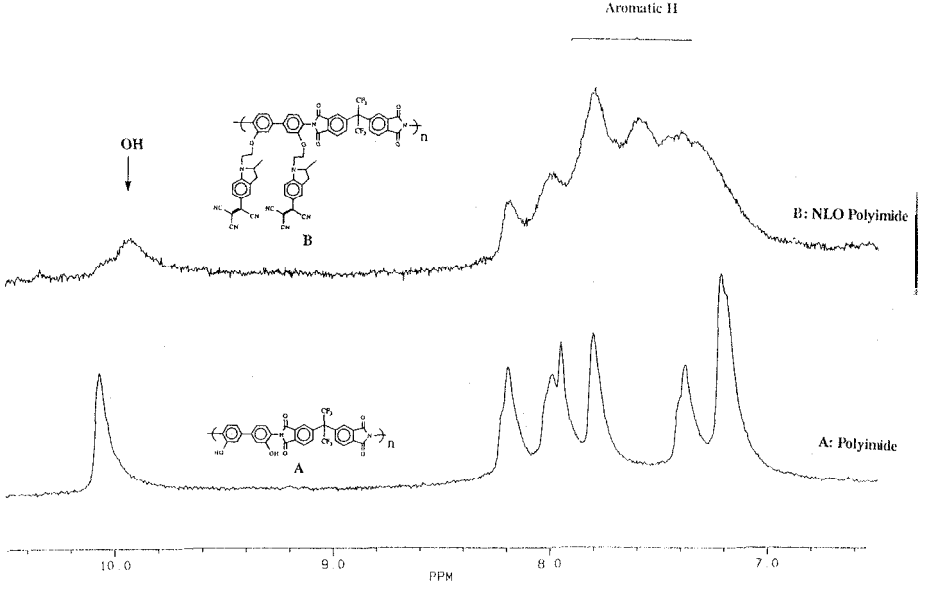


Fig. 3. ^{1}H NMR (250 MHz) SPECTRA OF PHENOL FUNCTIONALIZED POLYIMIDE (A) AND NLO POLYIMIDE (B) IN THE REGION OF 6-10.5 PPM (in d_6 DMSO)

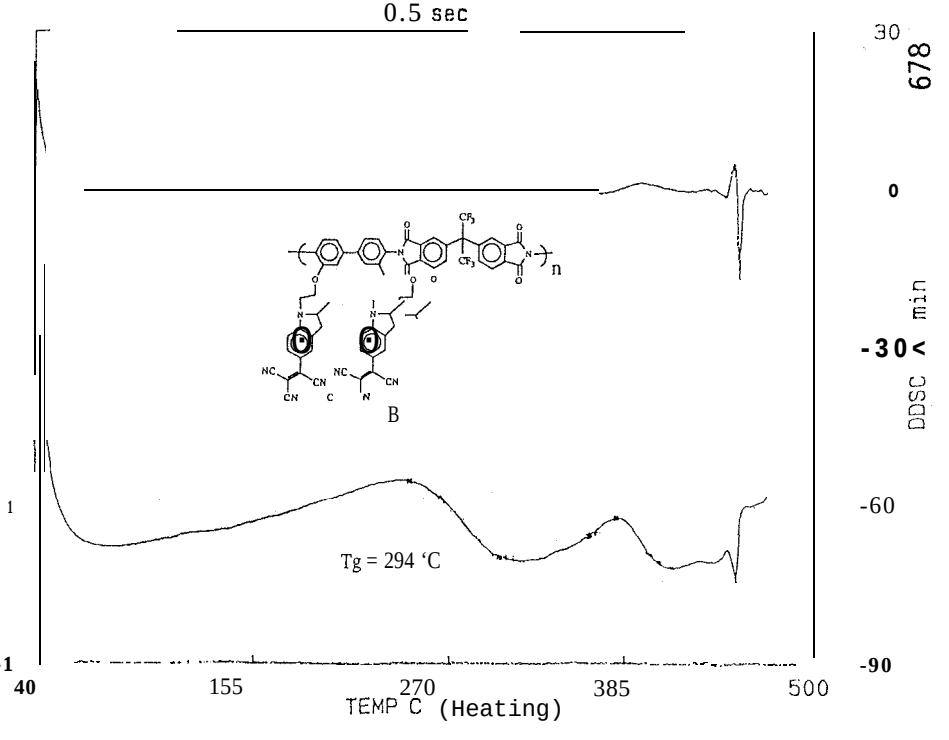


Fig. 4 DSC THERMAL ANALYSIS OF NLO POLYIMIDE (B)

AUTOMATIC TELESCOPE SEARCH FOR EXTRASOLAR PLANETS

Gregory W. Henry

Center for Automated Space Science *and* Center of Excellence *in Information Systems, Tennessee State University, Nashville, TN 37203*

Abstract. We are using automatic photoelectric telescopes at the Tennessee State Universit y Center for Automated Space Science to search for planets around nearby stars in our galaxy. Over the past several years, wc have developed the capability to make extremely precise measurements of brightness changes in Sun-like stars with automatic telescopes. Extensive quality control and calibration measurements result in a precision of O. 1% for a single nightly observation and 0.02% for yearly means, far bet ter than previously t bought possible with ground-based observations. We are able, for the first time, to trace brightness changes in Sun-like stars that are of similar amplitude to brightness changes in the Sun, whose changes can be observed only with space-based radiometers. Recent ly, exciting discoveries of the first extrasolar planets have been announced, based on the detection of very small radial-velocity variations that imply the existence of planets in orbit around several Sun-like stars. Our precise brightness measurements have been crucial for the confirmation of these discoveries by helping to eliminate alternative explanations for the radial-velocity variations. With our automat ic telescopes, we are also searching for transits oft hese planets across the disks of their stars in order to conclusively verify their existence. The detection of transits would provide the first direct measurements of the sizes, masses, and densities of these planets and, hence, information on their compositions and origins.

Astronomers at Tennessee State University (TSU) are using automatic photoelectric telescopes (APTs) to measure brightness changes in a variety of stars. The telescopes are located at the Fairborn Observatory automatic-telescope site in southern Arizona (Figures 1 & 2). These telescopes, and their associated precision photometers, operate complet ely automatically wit bout the presence of a human operator. A site-cent rol computer, interfaced to an automated weather station: opens the observatory at the beginning of each night and signals the individual telescope-control computers when it is time to begin and cease operations. The telescopes receive their observing instruct ions over the Internet from TSU, and the data are returned there automatically each morning. The speed and efficiency of the automatic telescopes allow us to make extensive quality-control and calibration measurements each night in addition to the science observations, so that the precision of our observations is maximized (Henry 1995a, b). In a collaborative program with astronomers at the Harvard-Smithsonian Center for Astrophysics and the Mount Wilson Observatory (Baliunas et al. 1997a), two telescopes, with apertures of 0.75 **m** and 0.80 m, are dedicated to long-term monitoring of 150 Sun-like stars in order to det ect subtle brightness changes that accompany their decade-long magnet ic cycles. The detection and characterization of these luminosity cycles in a large sample of stars should help us to understand more fully long-term changes in the Sun and their effect on earth's climate (see Henry & Henry, this volume). Recently, the discovery that several of these Sun-like stars might host planetary systems has given added importance to the study of their brightness changes.

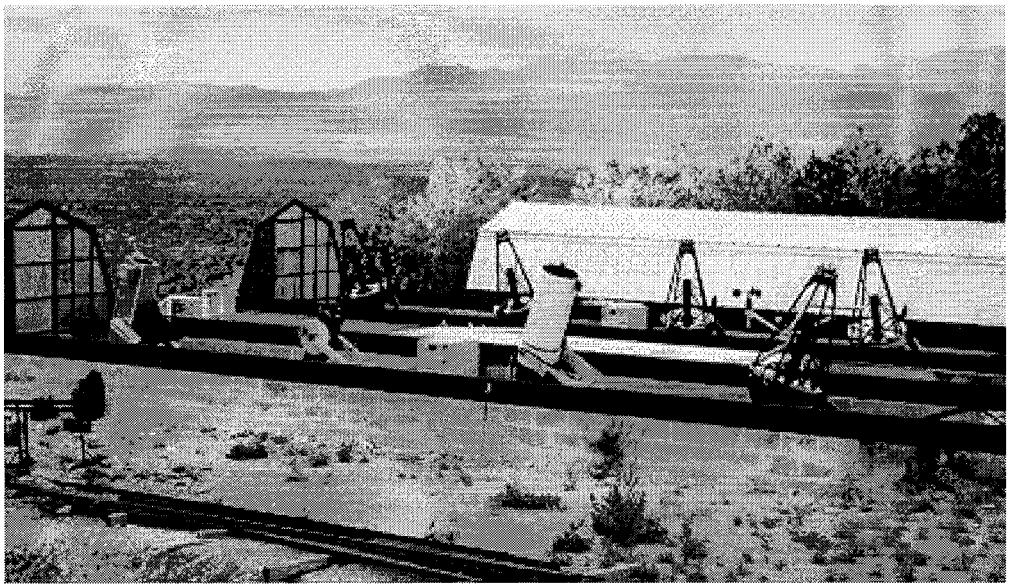


Figure 1. The automatic-telescope observing site in the Patagonia Mountains of southern Arizona. Currently, eight operating telescopes are housed in two roll-off-roof enclosures (shown in their open position). Four additional automatic telescopes are under construction and will be housed in the third (closed) shelter in the background. The 0.75 m and 0.80 m APTs used for precise measurements of Sun-like stars and the search for extrasolar planets are the rightmost telescopes in the picture. (Photo by F. Fekel)

At the ninth Cool Stars Workshop in 1995 October in Florence, Italy, Michel Mayor and Didier Queloz of Geneva Observatory in Switzerland announced one of the most exciting astronomical discoveries in recent years: a Jupiter-sized planet in orbit around the Sun-like star 51 Pegasi (Mayor & Queloz 1995). Although a small number of planetary-mass companions to pulsars had been found previously (e.g., Wolszczan 1994), 51 Peg's companion represented the first known extrasolar planet around a "normal" star, in this case one nearly" identical to the Sun. What was abnormal, however, was the distance of this planet from 51 Peg: 0.05 astronomical units (AU) – several times closer than Mercury is to the Sun. In the two years following this discovery, nine additional Sun-like stars were found to have at least one planet (see Naeye 1997 and Maran 1997 for recent popular-level reviews). Table 1 lists the known planets around Sun-like stars, in order of increasing orbit al period, along with their known properties. Efforts to understand the origins and unexpected characteristics of these planets are discussed by Marcy & Butler (1997).

Except for the case of the reported planets around HD 95735, which were found with astrometric methods, the extrasolar planets were discovered via highly refined and precise radial-velocity techniques (Baranne et al. 1996, Cochran & Hatzes 1994, Butler et al. 1996). In this technique, the existence of a planet around a star is *inferred* from the reflex motion imparted to the star by the orbit al motion of the planet. This motion results in a periodic Doppler shift in the light received from the star, which is measurable to a precision of about 3 ms^{-1} (Butler et al. 1996). Only the planet's minimum mass can be deduced from this information, because the inclination of the planet's orbit is unknown. Minimum masses for the new planets are given in Table 1, along with their orbital periods and the observed radial-velocity amplitudes imparted to their stars. Current technology does not allow the extrasolar planets to be imaged or their light to be analyzed spec-

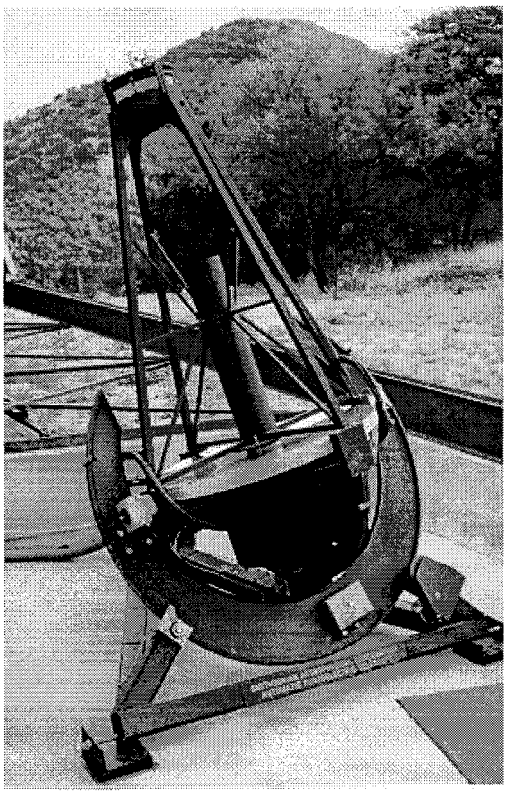


Figure 2. The 0.75 m APT at Fairborn Observatory. The black tub mounted behind the primary mirror is a single-channel, environmentally-controlled, precision photometer with an EMI 9124QB photomultiplier detector. A CCD camera in the photometer quickly and accurately centers each target star in the focal-plane diaphragm; a flip mirror then allows the light to fall on the PMT detector. This telescope can make differential photometric observations several times faster than a manual observer. (Photo by F. Fekel)

troscopically, so the minimum masses represent the only physical characteristic of the planets that can be measured at this time.

In the near term, radial-velocity techniques will continue to be the most productive method for the detection of additional extrasolar planets. However, a potential pitfall exists in this indirect technique: there are other possible causes of radial-velocity variations. Any star that undergoes radial pulsations, such as the well-known Cepheid- and Mira-type stars among others, will display Doppler shifts in their spectra that could be confused with reflex motion. In addition, a variety of stars exhibit non-radial pulsation modes that distort the shapes of spectroscopic line profiles and so mimic radial-velocity changes (Hatzes 1996). In still other stars, surface magnetic activity, such as starspots, plages, and convective inhomogeneities, can also cause variable line-profile distortions and apparent radial-velocity shifts as these surface features are carried across a stellar disk by rotation or as they evolve over the course of a star's magnetic cycle (Saar & Donahue 1997).

The time scales of these phenomena (days to decades) precisely match those of the radial-velocity variations being sought as evidence for planetary systems and are, therefore, potential sources of false detections of new planetary systems. All of these phenomena, however, should be accompanied, at some level, by light variations that would betray their presence. Therefore, our

Table 1. Properties of Extrasolar Planets

Planet	P_{orb} (days)	ccc.	Orbital Rad. (AU)	RV Amp. ($m\ \ s^{-1}$)	Mass (M_{Jup})	Discoverer
τ Boo b	3.31	0.018	0.05	*469*	≥ 3.87	Butler et al. (1997)
51 Peg b	4.23	0.012	0.05	*55.9*	≥ 0.45	Mayor & Queloz (1995)
v And b	4.62	0.11	0.06	*74*	≥ 0.68	Butler et al. (1997)
ρ^1 Cnc b	14.64	0.05	0.11	*77*	≥ 0.84	Butler et al. (1997)
ρ CrB b	39.645	0.028	0.23	*67.4*	≥ 1.1	Noyes et al. (1997)
HD 114762 b	83.9	0.35	0.4	*570*	≥ 9	Latham et al. (1989)
70 Vir b	116.6	0.40	0.43	318	≥ 6.6	Marcy & Butler (1996)
16 Cyg B b	800.8	0.63	1.6	*43.9*	≥ 1.5	Cochran et al. (1997)
47 UMa b	1090	0.03	2.1	*45.5*	≥ 2.39	Butler & Marcy (1996)
HD 95735 b	2100	0.0	2.2		~ 0.9	Gatewood (1996)
ρ^1 Cnc c	>3000				> 5	Butler et al. (1997)
HD 95735 C	11000	0.0	11		~ 1.1	Gatewood (1996)

precise photometric observations from the automatic telescopes are important for the elimination of these alternative explant ations for the radial-velocity variations. None of the systems listed in Table 1 have any variability that would discredit the new planetary discoveries (Henry et al. 1997, Baliunas et al. 1997b, Henry et al. 1998).

Figure 3, for example, presents our observations of 51 Peg plotted modulo the 4.23-day orbital period of the proposed companion. Fourier analysis of the data, assuming that period, gives an upper limit of $0.02\% \pm 0.02\%$ for any light variability. This eliminates radial pulsations and surface magnetic activity as possible causes of the radial-velocity variations and places tight constraints on non-radial pulsations as well. Gray (1997), however, reports that he *has* detected subtle line-profile variations in his spectra of 51 Peg that imply some form of non-radial pulsations, even though they are not expected in this type of star. He concludes that "the presence of a planet is not required to explain the data." His claim has stirred considerable controversy, especially since other measurements have not revealed the line-profile variations (see Marcy & Butler 1997). In a more detailed analysis, Gray & Hatzes (1997) conclude that the evidence for line-profile variations is "strong, but not conclusive"; they propose a mode of non-radial pulsation that is consistent with the line-profile variations and might be consistent with the lack of observed photometric variations. Further high-resolution spectroscopy and precision photometry are being obtained to resolve the controversy.

This serves to illustrate the problems with indirect detection methods for extrasolar planetary systems. Even astrometric methods, which seek to visibly resolve the reflex motion caused by planetary companions, are not immune to some of these problems. Surface magnetic activity, for instance, can cause the photocenter of a star to vary systematically on stellar-rotation and magnet it-cycle time scales. Therefore, high-precision photometric observations will be necessary for validating astrometic planetary discoveries as well. The photometry will also be useful for

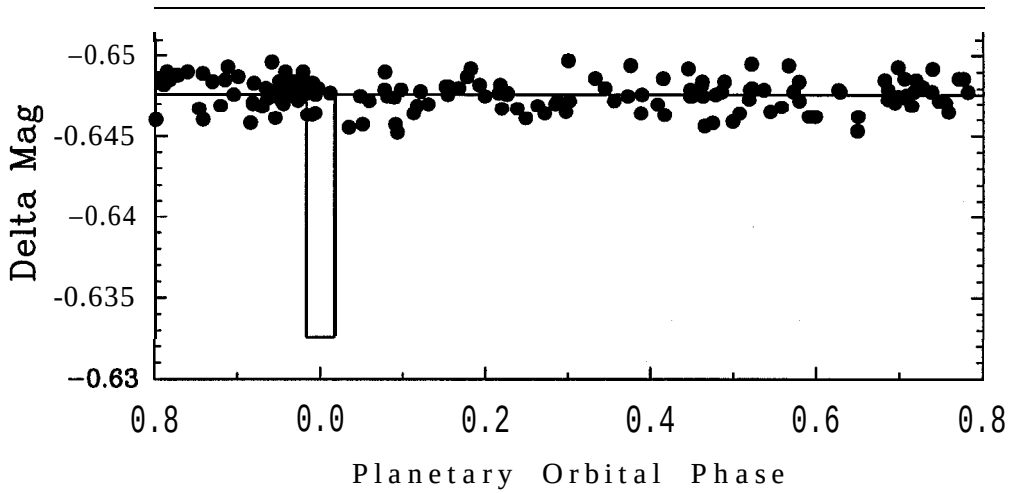


Figure 3. Photometric observations of 51 Peg plotted modulo the planetary orbital period of 4.23 d. Analysis of the data reveals no light variability that would cast doubt on the existence oft he planet. The line indicates the drop in light that would be observed if the orbit of the planet were aligned with our line of sight, resulting in a transit. Our observations clearly rule this out for 51 Peg.

determining which stars have the stable energy output necessary to sustain life that might exist on planets within the habitable zones around the stars.

Although direct detection of extrasolar planetary systems maybe decades away, perhaps requiring large, space-based interferometric telescopes, the observation of planetary *transits* via ground-based photometry may be able to provide crucial information much sooner. Our careful photometry of thes ystems listed in Table 1 has so far failed to reveal any transits during the times of conjunction predicted by the radial-velocity data, even though the short-period planets each have approximately a 10% probability for transits (see Figure 3). The detection of a photometric transit, in addition to providing indisputable proof of the planet's existence, would lead directly to the actual (as opposed to the minimum) mass, the size, and the density of the planet. These parameters would then result in improved theoretical models of the composition and origin of extrasolar planets (e.g., Saumon et al. 1996, Guillot et al. 1997, Ward 1997, Boss 1997). Our automatic telescopes are continuing to search for these transits. Several new radial-velocity searches examining larger samples of stars with larger telescopes are now underway, and we are constructing additional APTs to monitor these stars. When the next few stars with short-period planets are identified, the probability that our APTs can find one of them exhibiting photometric transits will approach unity.

Acknowledgments. Astronomy with automated telescopes at Tennessee State University has been supported by the National Aeronautics and Space Administration, most recently through NASA grants NAG8-1014, NCC2-883, and NCC5-228 (which funds TSU's Center for Automated Space Science), and the National Science Foundation, most recently through NSF grant HRD-9550561 and HRD-9706268 (which funds TSU's Center for Systems Science Research). The planetary search program is also partially supported by the Richard Lounsbery Foundation.

References

Baliunas, S. L., Donahue, R. A., Soon, W., & Henry, G. W. 1997a, in The Tenth Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. R. A. Donahue & J. A. Bookbinder (San Francisco: ASP), in press

Baliunas, S. L., Henry, G. W., Donahue, R. A., Fekel, F. C., & Soon, W. H. 1997b, ApJ, 474, L119

Baranne, A., et al. 1996, A&AS, 119, 373

Boss, A. P. 1997, Science, 276, 1836

Butler, R. P., & Marcy, G. W. 1996, ApJ, 464, L153

Butler, R. P., et al. 1996, PASP, 108, 500

Butler, R. P., et al. 1997, ApJ, 474, L115

Cochran, W. D., & Hatzes, A. P. 1994, Ap&SS, 212, 281

Cochran, W. D., et al. 1997, ApJ, 483, 457

Gatewood, G. 1996, BAAS, 28, 885

Gray, D. F. 1997, Nature, 385, 795

Gray, D. F., & Hatzes, A. P. 1997, ApJ, 490, 412

Guillot, T., et al. 1997, in Astronomical and Biochemical Origins and the Search for Life in the Universe, ed. C. B. Cosmovici, S. Bowyer, and D. Werthimer (Balogna: Editrice Compositor), 343

Hatzes, A. P. 1996, PASP, 108, 839

Henry, G. W. 1995a, in ASP Conf. Ser. 79, Robotic Telescopes: Current Capabilities, Present Developments, and Future Prospects for Automated Astronomy, ed. G. W. Henry & J. A. Eaton (San Francisco: ASP), 37

Henry, G. W. 1995b, in ASP Conf. Ser. 79, Robotic Telescopes: Current Capabilities, Present Developments, and Future Prospects for Automated Astronomy, ed. G. W. Henry & J. A. Eaton (San Francisco: ASP), 44

Henry, G. W., Baliunas, S. L., Donahue, R. A., Soon, W. H., & Saar, S. H. 1997, ApJ, 474, 503

Henry, G. W., Baliunas, S. L., & Donahue, R. A. 1998, ApJ, in preparation

Latham, D. W., et al. 1989, Nature, 339, 38

Maran, S. P. 1997, Smithsonian, Sept, 73

Marcy, G. W., & Butler, R. P. 1996, ApJ, 464, L147

Marcy, G. W., & Butler, R. P. 1997, in The Tenth Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. R. A. Donahue & J. A. Bookbinder (San Francisco: ASP), in press

Mayor, D., & Queloz, D. 1995, Nature, 378, 355

Naeye, R. 1997, Astronomy, April, 42

Noyes, R. W., et al. 1997, ApJ, 483, L111

Saar, S. H., & Donahue, R. A. 1997, ApJ, 485, 319

Saumon, D., Hubbard, W. B., Burrows, A., Guillot, T., Lunine, J. I., & Chabrier, G. 1996, ApJ, 460, 993

Ward, W. R. 1997, ApJ, 482, L211

Wolszczan, A. 1994, Science, 264, 538

INTRANIGHT VARIABILITY DURING THE 1997 OUTBURST OF BL LAC

Sandra D. Clements, Michael T. Carini, John C. Noble
Center for Automated Space Science
Western Kentucky University

Abstract

The blazar BL Lac underwent a major outburst this summer. To search for intranight variability during the outburst, observations of BL Lac were made with the 30 inch telescope at the University of Florida's Rosemary Hill Observatory. During ten nights in July, nearly 300 CCD observations were made. BL Lac was seen to vary by more than 1.5 magnitudes over those ten nights. Substantial intranight variability was also observed. For example, on July 29, BL Lac increased in brightness by more than 0.5 magnitude in less than 2.5 hours. The observations of BL Lac taken in July are reported here and discussed in relation to the models intended to explain intranight variability. The authors gratefully acknowledge the support of NASA grant NCC5-228.

Intranight Variability

Intranight variability, or microvariability, of active galactic nuclei (AGN) is defined by Miller, Carini, & Goodrich (1989) as ". . optical variations on timescales significantly shorter than a day." In that same paper, the first conclusive evidence of microvariability was presented. Since then, it has been found to be typical behavior for blazars (BL Lacertae objects and optically violent variable quasars). Blazars are radio loud objects, indicating that their jets lie nearly along the observer's line-of-sight. On the other hand, microvariability is seen less frequently in radio quiet quasars whose jets are not thought to. be beamed toward the observer. Various mechanisms have been proposed to explain microvariability (Gopal-Krishna, Sagar, & Wiita, 1993). Among them, hot spots on an accretion disk surrounding the nucleus, shocks propagating through turbulence in a jet, and shocks in a jet whose direction to the line-of-sight varies. The responsible mechanism is still unclear. That microvariability is typical of radio loud objects but less common in radio quiet objects favors jet models. The microvariability is detected at all in radio quiet objects favors accretion disk models for at least some objects (Sagar, Gopal-Krishna, & Wiita, 1996). The microvariability we report here is for the blazar BL Lacertae (BL Lac) which was undergoing a major optical flare during the summer of 1997.

BL Lac

BL Lac is the prototype of a class of AGN characterized by non-thermal spectral energy distributions, weak or undetectable emission lines, and strong and variable polarized emission (Angel & Stockman, 1980). These characteristics of BL Lac objects are generally attributed to emission from a relativistic jet oriented at small angles to the observer's line-of-sight. The radio jet of BL Lac itself has been directly observed using VLBI and is seen to exhibit superluminal motion (Mutel & Phillips, 1984). That is, emission features in the radio jet appear to move faster than light - an optical illusion caused by relativistic motion of these features almost directly toward the observer, Hughes, Aller, & Aller (1989) showed that events seen in the radio light curve of BL Lac were correlated with features in the radio jet. Using a shock-in-jet model, they attributed the brightening in the radio light curve to the emergence of emission features from the core. The fading in the radio light curve was attributed to the fading of these feature as they propagated down the jet.

BL Lac has an optical brightness range of over 4 magnitudes in its long term light curve. In May of 1980 it was as faint as V = 16.7 (Carini, Miller, Noble, & Goodrich, 1989). It brightened to

V = 12.5 in July of 1997. Microvariability has been observed in BL Lac since 1986 (Miller, Carini, & Goodrich, 1989). Are any of these optical variations of BL Lac associated with its jet? When long term optical light curves were compared to long term radio light curves, a "fair" correlation was seen between optical and radio events prior to 1978 and thus prior to VLBI mapping of radio jets (Clements, Smith, Aller, & Aller, 1995). During this time, the radio events were seen to lag optical events by approximately 1 month. However, there was no correlation between optical and radio events seen in the post-1978 light curves when Hughes, Aller, & Aller (1989) found events in the radio light curve to be correlated with features in the VLBI map of its jet. So the question of whether optical variations arise in the jet is still uncertain.

Observations and Data Reduction

In the summer of 1997, BL Lac underwent a major outburst. To investigate the microvariability behavior of BL Lac during outburst, observations were made on ten nights in July with the 30 inch telescope at the University of Florida's Rosemary Hill Observatory. Exposures were made with a Photometries Star-1 CCD camera placed at the f/4 Newtonian focus of the telescope. The camera has 384 x 576 pixels. At the Newtonian focus, the field of view of the camera is 10' x 15'. Nearly 300 exposures were made through V and R filters. Exposures through the R filter were typically 30 seconds but were sometimes as long as 60 seconds. With the V filter, exposures were typically 180 seconds but occasionally as long as 300 seconds. Multistar photometry was performed on BL Lac and four comparison stars in the same field as BL Lac using the *apphot* package of the IRAF* data reduction software.

Results and Discussion

Over the ten nights observed in July, BL Lac's brightness varied by at least 1.7 magnitudes from V = 14.97 and R = 14.27 at its faintest on July 4 to V = 13.29 and R = 12.73 at its brightest on July 29. (See Figure 1.) A single exposure on July 8 showed R = 12.48, but observing conditions were poor as high humidity had resulted in a fogged mirror. Variations were detected each night multiple exposures were made. Most nights, little structure was seen in the variations. That is, BL Lac either brightened throughout the observing session or faded throughout the observing session. Complete flares were not detected on most nights. More structure was seen on July 29 when at least four flares were observed. One reason more flares were detected this night was because it was the longest observing session lasting over 6.5 hours. The other observing sessions only lasted between 2 and 4 hours. The largest variation detected during the ten nights of observing occurred on July 29 when BL Lac brightened by 0.57 R magnitudes in 2.2 hours. The greatest rate of change also occurred on July 29 when BL Lac brightened by 0.46 magnitudes per hour during one flare. This is among the fastest rates ever detected.

Of the days multiple exposures were made, BL Lac was most active on the day it was brightest, undergoing -0.6 flares per hour on July 29 when V- 13.3. This activity is comparable to that found 10 years earlier (~0.5 flares/hr) when BL Lac was more than a magnitude fainter at V = '14.6 (Miller, Carini, & Goodrich ,1989). That BL Lac exhibited the same level of activity at these different brightness levels suggests that it maybe inappropriate to attribute the microvariability of this object to hot spots on an accretion disk since you might expect such hot spots to be more detectable when the object is fainter and therefore its jet contribution reduced (Noble & Miller, 1996).

A comparison of the V and R microvariability data indicates a strong correlation with little or no lag time between events in the two light curves. However, the two light curves are not merely

offset in magnitude one from the other. On July 4, a small event is seen in V but no corresponding event seen in R. In addition, the rate at which the brightness varies in V is not always the same as the rate it varies in R. On July 3, for example, BL Lac brightened faster in R than in V. Then on July 4, it faded faster in V than in R. Yet, over the ten days observed in July, BL Lac was seen to become bluer as it brightened. (See Figure 2.) Models for microvariability will need to address these observational constraints.

---

*IRAF is distributed by the National Optical Astronomy Observatories, which is operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

References

Angel, J. R. P. & Stockman, H. S. 1980, ARAA, 8, 321

Carini, M. T. 1990, PhD dissertation, Georgia State University

Carini, M. T., Miller, H. R., Noble, J. C., & Goodrich, B. D. 1992, AJ, 104, 15

Clements, S. D., Smith, A. G., Aller, H. D., Aller, M. F. 1995, AJ, 110,529

Gopal-Krishna, Sagar, R., & Wiita, P. J. 1993, MNRAS, 262,963

Hughes, P. A., Aller, H. D., & Aller, M. F. 1989, in *BL Lac Objects,* edited by L. Maraschi, T. Maccacaro, and M.-H. Ulrich (Springer-Verlag, Berlin), p.30

Jang, M. & Miller, H. R. 1995, ApJ, 452,582

Mutel, R. L. & Phillips, R. B. 1984, in *IAU Symposium 110, VLBI and Compact Radio Sources,* edited by R. Fanti, K. Kellerman, and G. Setti (Dordrecht, Reidel), p. 117

Noble, J. C. & Miller, H. R. 1996, in *Blazar Continuum Variability,* edited by H. R. Miller, J. R. Webb, and J. C. Noble (Astronomical Society of the Pacific, San Francisco), p.30

Sagar, R., Gopal-Krishna, & Witta, P. J. 1996, MNRAS, 281, 1267

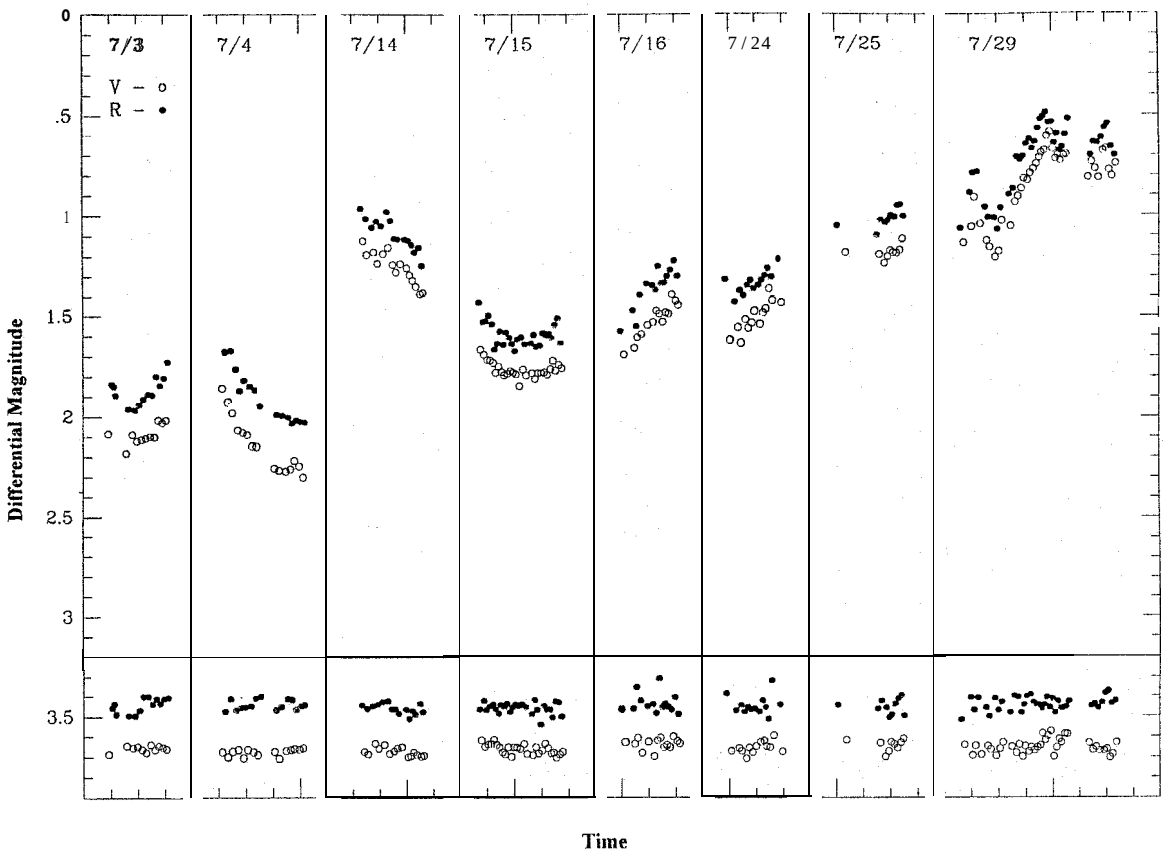


Figure 1- Differential Magnitude versus Time

Differential V and R magnitudes are shown for eight nights in July of 1997. The top panel shows differential magnitudes for BL Lac (magnitude of BL Lac - magnitude of comparison star 2). The bottom panel shows differential magnitudes for comparison star 3 (magnitude of comparison star 3- magnitude of comparison star 2). The scatter in the differential magnitudes in the lower panel indicates the overall observational uncertainty in the data (Carini, 1990).

The smallest increment on the x axis represents 0.05 days or 1.2 hours.
The smallest increment on they axis represents 0.1 magnitudes.

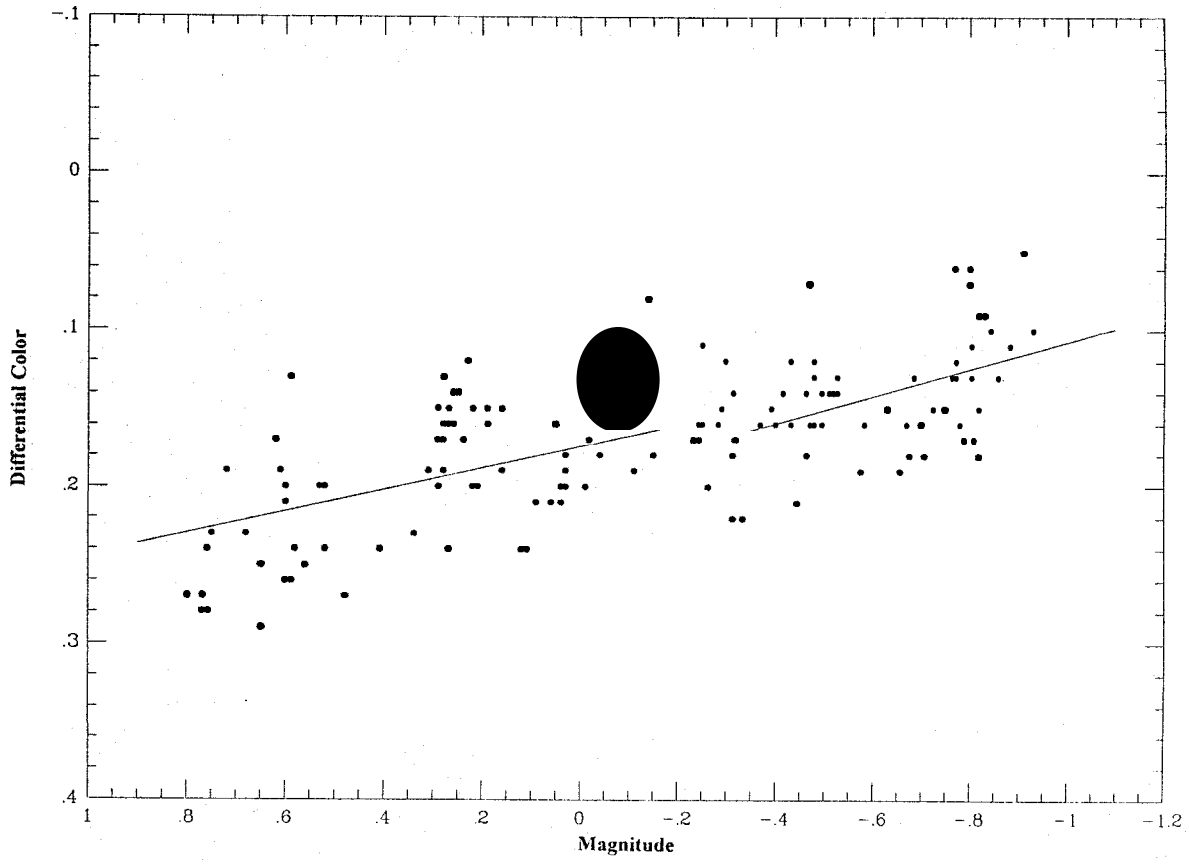


Figure 2- Differential Color versus Magnitude

The differential colors are obtained by subtracting differential R magnitudes from differential V magnitudes. This is equivalent to subtracting the color of comparison star 2 from the color of BL Lat. Differential colors are used to account for color variations of the atmosphere and interstellar medium. In this particular version of this plot, R magnitudes are interpolated to the times of the V exposures prior to computing differential colors. The differential colors were then plotted against actual (not interpolated) differential V magnitudes. A clear correlation is seen between color and magnitude with BL Lac becoming bluer as it brightens.

The smallest increment on the x axis represents 0.05 magnitudes.
The smallest increment on they axis represents 0.02 magnitudes.

RANDOM SPOT ON CHROMOSPHERICALLY ACTIVE STARS

Tamara S. Williams and Joel A. Eaton
Center for Automated Space Science (CASS)
Tennessee State University, Nashville, TN 37209

ABSTRACT

Observations of rapidly rotating chromospherically active stars provides the basis of the theory that these stars are highly spotted. Previously, we thought that the stars contained only two large spots at any one time. Eaton, Henry, and Fekel proposed that what appears to be two spots could actually be as many as 25 spots randomly distributed on a differentially rotting star. Initial tests of this random spot model produced all of the characteristic photometric variations of highly spotted stars. We have tested the random spot theory by measuring the differential rotation of the spots (k), the lifetimes (τ), and maximum and minimum periods of migration in approximately fifty data sets. These data sets consist of spots with an average lifetime of 2 years (cases 1- 21),5 years (cases 26- 38), and 10 years (cases 39- 49), Each data set spans 14 years, which is comparable to real stars that have been observed for about 20 years. The derived spot lifetimes in the two-spot model increased with assumed lifetimes of the random spots; specifically, median two-spot lifetimes were 0.7, 2.2, and 2.3 years, respectively for the actual random spot lifetimes of 2, 5, and 10 years. Eaton, et al. found a median two-spot lifetime of 1.5 years for some actual stars. Therefore, if the random spot model is correct, actual spot lifetimes are probably somewhat shorter than 5 years. Furthermore, this result shows that random spots with reasonable lifetimes can fully reproduce the persistence of spots in the actual stars' light curves.

I. introduction

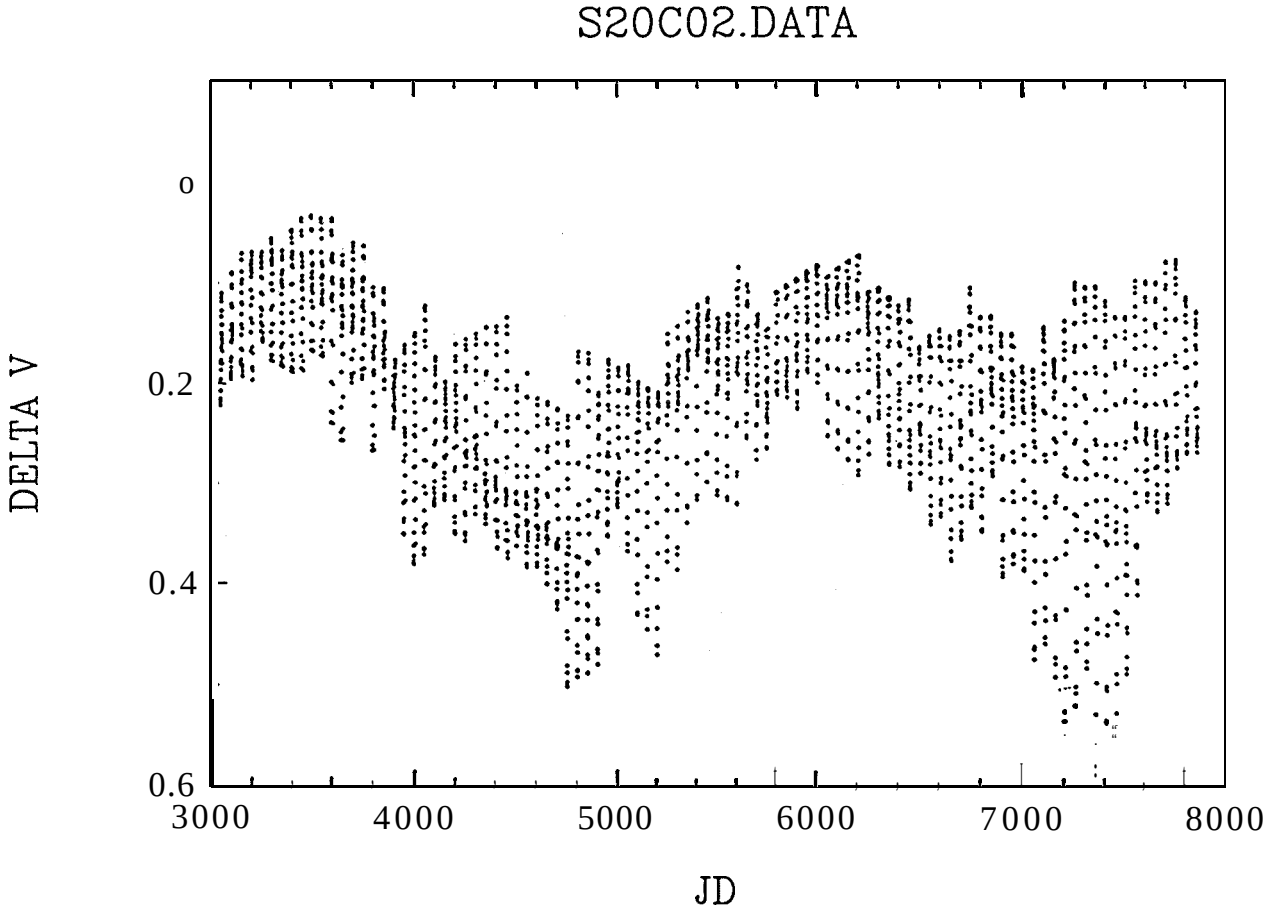
In recent years, Douglas Hall of Vanderbilt University, Joel Eaton, Gregory W. Henry, and Francis C. Fekel, all of Tennessee State University's Center of Automated Space Science, have collaborated their research efforts to explain the rotational light curves of magnetically active cool stars. The tool that allowed these researchers to fit the light curves of these stars was the two-spot model. This simple curve-fitting scheme devised by Hall approximates the effect of the spots as the lower lobe cosine function and fits a light curve with one or two such dips to find the spots" longitudes and strengths. The hypothesis gives a solid observational framework for interpreting the light variations of such highl y spotted stars.

Chromospherically active refers to the excess heating of the layer of a star's atmosphere located between the photosphere and the corona. The chromospherically active stars appear to be highly spotted. Eaton has calculated the light curves of stars with 20 to 25 spots on a rapidly rotating star as an alternative to the two-spot hypothesis. Eaton, Henry, and Fekel states, "We find this approach (1) gives rotational light curves that look like observations of chromospherically active stars, (2) provides two-spot solutions similar to those for actual stars,

and (3) gives appropriate changes in mean light level from season to season if spots have a limited lifetime, of the order of one year." The data has been calculated to test Hall's hypothesis illustrates that more than two spots can be present on the surface of a star.

II. Calculations

Eaton, et al. calculated light curves for circular spots placed randomly over the surface of a star of temperature 3500 K. The axis that these stars are rotating on is at an angle of 70° to the line of sight. The rotational period of the star was assumed to be 10 days at a latitude of 30° to divide the star into equatorial and polar zones of equal area. Eaton, et al. Calculated every fifth light curve while afterwards I analyzed every single one of those light curves. **Figure 1** illustrates a plot of a typical 14-year light curve where ΔV represents the brightness of the star and JD represents the years in Julian days.



When the theoretical light curves are calculated, we could analyze them with the two-spot method. Using a program previously written in FORTRAN, I was able to first, identify the maximum magnitude of the light curve by marking the highest point on the curve and setting its limits. The second step was to mark the two lobes of the cosine function to identify the two spots and set their limits. This prompted the computer to calculate a curve to fit the spots, thus the final step was to calculate a least squares fit of the plot. This procedure was performed -60 times for each star over the period measured in JD. If one of the spot's happened to have a negative amplitude the data had to be discarded and that particular panel's calculations were repeated. The plot of a light curve that has been fit by the least squares fit is shown in **Figure 2.** The abscissa is the rotational phase of the star which ranges from O to 1.0 year cycles of rotation. Note the list of calculations of the amplitudes and phases below the curve.

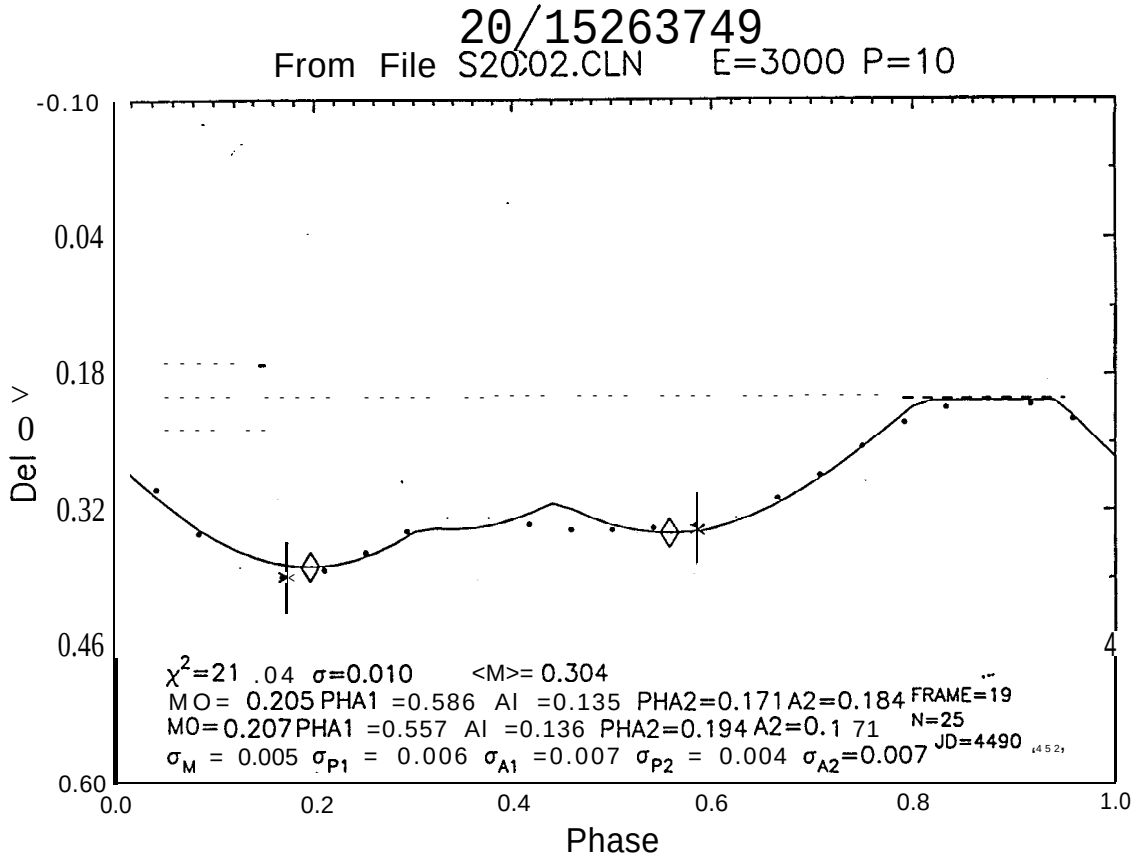


Figure 2

We observed that the amplitude of variation decayed which was a result of differential rotation (k). The differential rotation was 0.01 in the cases that I observed which were more similar to that of actual stars compared to k=0 .03 used by Eaton, et al. It was very important to notice that the differential rotation was proportional to the rate of migration for the two-spot model. The changing dips must migrate through the light curve at an appropriate rate if the model was to reproduce phenomena in actual stars. A spot rotating faster or slower than an assumed rotation period appears either later or sooner in successive light curves. As a result, the spot migrates through the light curve, and this migration is what a two-spot analysis measures most accurately.

Three distinct sets of data were tested in this experiment. For the first set, which consists of cases 1 through 21, the life span of the spots were 2 years; cases 26 through 38 had a life span of 5 years; and cases 39 through 49 had a life span of 10 years.

RESULTS FROM MIGRATION CURVES

TABLE 1.1

Case	#of spots	τ_{max}(yrs.)	τ_{min}(yrs.)	Median τ	k
1	16	3.56	0.40	0,76	0.018
2	19	2.19	0.28	0.55	0.018
3	20	2.47	0.28	0.55	0.015
4	15	3.55	0.40	0.96	0.024
6	21	2.48	0.28	0.55	0.021
7	22	3.01	0.27	0.55	0.039
8	20	2.48	0.27	0.41	0.021
9	17	3.43	0.28	1.10	0.013
10	17	3.41	0.28	0.55	0.019
11	15	3.41	0.27	1.57	0.023
12	19	2.46	0.27	0.96	0.017
13	21	2.19	0.26	0.54	0.025
14	17	2.47	0.27	0.55	0.027
15	16	3.01	0.28	0.55	0.011
16	17	2.47	0.40	0.55	0.018

17	23	1.65	0.28	0.55	0.024
18	19	2.47	0.27	0.56	0.021
19	19	2.47	0,27	0.55	0.016
20	19	2.47	0.27	0.42	0.040
21	16	2.46	0.27	1.31	0.024
26	7	6.16	0.97	4.52	0.006
27	8	5.34	0.83	2.68	0.008
28	9	6,57	0.41	2.19	0.012
29	9	6.17	0.40	2.46	0.008
30	16	2.47	0.28	1.30	0.015
31	8	7.39	0.40	2.46	0.022
32	11	3.28	0.55	2.05	0.016
33	10	4.52	0.42	2.46	0.012
34	12	4.52	0.40	1.44	0.015
35	11	6.30	0.27	1.50	0.012
36	11	4.51	0.42	2.19	0.011
37	15	5.47	0.27	0.55’	0.025
38	13	4.11	0.41	0.55	0.012
39	9	4.38	0.55	2.47	0.0071
40	11	3.57	0.55	1.37	0.0067
41	10	6.30	0.27	3.29	0.0072
42	14	3.41	0.27	1.50	0.015
43	11	4.52	0.42	1.10	0.004
44	8	9.59	0.55	2.0	0.011
45	7	6.98	0.55	3.3	0.0063
46	11	3.01	0.55	2.5	0.0058

47	10	4.52	0.28	3.0	0.012
48	8	8.35	1.23	3.1	0.003
49	10	4.52	1.10	2.0	0.003

AVERAGE VALUES FOR MEAN LIFETIMES AND DIFFERENTIAL ROTATION
τ =2.5, and 10 yrs.
TABLE 1.2

	τ	k
τ = 2 years	0.70 ± 0.31	0.017 ± 0.009
τ = 5 years	2.18 ± 1.23	0.013 ± 0.005
τ = 10 years	2.33 ± 0.793	0.0074 ± 0.0038

Eaton, et al. noted that the spot lifetimes for a star with k=0.0 1 were in the range of 0.27 to 2.46 years with a median value of 1.23 years. From the data that I have calculated, the range was from 0.27 to 3.56 years and a median of 0.646 years. Eaton, et al. assumed that the rotational period of a star, P_{rot}, was 10 days. The equation used to calculate he rotational period using the value of P and the migration period of the star is

$$\frac{1}{P_{mig}} = \frac{1}{10} - \frac{1}{P_{rot}} \qquad (1.1)$$

Example Calculation:

Case 6:
Measurement of Star's Rotational Period:

$-1/P_{mig} = 1/10 - -1/P_{rot}$

P_{mig} = -881 days

$1/881 = 1/10 - 1/P_{rot}$

P_{rot} = 10.11 days

Measurement of Spot's Rotational Period:

P_{mig} = 865 days

$-1/988 = 1/10 - -1/P_{rot}$

$P_{rot} = 9.90$ days

Next, we calculated and compared the experimental differential rotation to the actual differential rotation. The actual value for k in the cases that were analyzed was 0.01.

$P(\beta) = P_{eq} / (1 - k\sin^2(\beta))$

$\Delta P = P_{eq}\, k / (1-k)$
$k = \Delta P / \langle P \rangle$

$$k = \frac{2\,|10.11 - 9.90|}{|10.11 + 9.90|}$$

$k = 0.021$

The values of k ranged from 0.006 to 0.040. We found that out values were similar to the 0.2-6.5 year range with a median of 1.5 years, found by Henry, et al. for 53 actual spots. An example of a completely analyzed migration curve and its calculations are attached to this document.

III. Results

The first inquiry made was to determine how accurate the light curves were calculated for the random distribution of spots in comparison to actual stars. The criteria we applied in deciding are as follows: First, do the calculated light curves for different numbers of random spots actually have the right shapes? Second, do the calculated light curves have the correct amplitude within a given year and from year to year as actual stars? Third, can the two-spot model be used to fit a series of these light curves toidentify "spots" that migrate in rotational phase with the light rate and persist as long as those in actual stars?

We have previously shown that calculated light curves have the same shapes and amplitudes as actual observed light curves. The contribution of this study is to show that indeed the random spot model can give two-spot lifetimes similar to those derived for actual chromospherically active stars. According to Table 1.2, twenty random spots having assumed lifetimes of 5 or 10 years will give two-spot lifetimes of 2.18 or 2.33 years, respectively. Henry, et al. found a range of 0.2 -0.5 years with a median of 1.5 years for 53 actual "spots" on four stars. This implies that a random-spot lifetime of a little less than 5 years will satisfy observations of actual stars.

REFERENCES

Eaton, J. A., Henry, G. W., Fekel, F.C. 1995, ApJ Letters
Henry, G. W., Eaton, J. A., Harrier, J., Hall, D.S, 1995 ApJ

Geologic Mapping of the Cd. Chihuahua, Mexico, area from Satellite Imagery

Perez, A. E., Rodriguez, J. A.. Pingitore, N. E.. Keller. G R.. Penn. B. S.
Pan-American Center for Earth and the Environmental Sciences. Department of Geological Sciences
500 W. University El Paso, Texas 79968
University of Texas at El Paso
perez@geo.utep.edu
(915)747-5599

Introduction

The City of Chihuahua in Northern Mexico has been experiencing a water shortage in the last decade due to the great demand of its growing population. About a quarter of the municipal water is supplied by the Tabalaopa-Aldama aquifer located to the North-East of the city. The Tabalaopa Valley has been used for hundreds of years as agricultural land that is irrigated with intermittent irrigation from the Chuviscar River (Fig I). Because of the scarcity of water and the expense of chemical fertilizers some of the crops are irrigated with untreated sewage water from Chihuahua City. Although this poses no direct health threat given that the crops are for animal consumption only, the danger lies in the possible contamination of the underlying water. Percolation of the contaminated waters and the proximity of the water table present a conflict between the need of the farmers and the health of the community. The area is of great hydrologic interest because of its potable water potential as well as its agricultural resources. One of the first steps in addressing the problem is to better understand the geology of the area. Outlining the rock types present is the focus of this study. By integrating the preexisting geologic maps as well as TM satellite imagery we intend to redefine the geologic map in greater detail.

Geology

The Tabalaopa Valley is part of the Basin and Range Province. The province is characterized by extensional fault-block mountains and deep sediment filled basins. The valleys are formed by regional extension and grabens later filled by alluvial sediments. These sediments serve as reservoirs for water.

The base of the stratigraphic column begins with Precambrian metagranites, amphibolites, and gneisses. Overlying these are Cretaceus carbonates and evaporates which in turn are unconformably overlain by Tertiary volcanic rocks and volcaniclastic sediments. The basin is composed of Quaternary, Pleistocene, and Holocene alluvial unconsolidated deposits. The Tabalaopa-Aldama aquifer occurs in a sedimentary environment dominated by alluvial fans. These sediments are composed mainly of gravel, boulders, sand, and some lenses of clay that were observed during drilling (Rodriguez, 1997).

Classification analysis

The study area consisted of four main Sierras that were treated as separate entities for classification purposes. The first step involved comparing existing geologic maps to a 1986 TM image. Several band combinations were examined to find the best contrasts between the different lithologies. We found that for our purposes bands R:5, G:3, B: 1 showed the best contrasting colors for the lithologies

present. From the geologic map key areas were identified on the image. These areas contained known lithologies recognized by previous workers. These key areas were used to perform a supervised classification in which the known areas of the image are used to find the closest match in the rest of the unclassified image. Thus, by using an area's average digital number (DN) or reflectance, the program's maximum likelihood classification defines areas of similar lithologies for the entire image. Areas that did not match are unclassified and appear as black patches in the image. To define unclassified areas spectral slices were obtained through the unclassified patch and their reflectance was plotted against wavelength. The results were then used for comparison with the spectral library to identify the rock type (Fig.II). By taking the classified image, the existing geologic map, and the TM image we were able to redefine the lithologic boundaries of the study area.

Classification Results

Figure III shows the final geologic map produced by the techniques described above. Most of the lithologies within the Sierras were classified, although a few areas still remain unknown. The major faults in the valley were defied by following a geologic map. Although most of them had been previously defined, we did identify some smaller splays that do not appear in any of the published material. All of the faults appear consistent with gravity surveys conducted in the area. These results are currently being correlated with the geochemistry, hydrology and gravity studies.

Conclusions

A geologic map produced from this study provides the basis for continuing geochemical, geophysical, and hydrologic investigation of water quality and availability for the residents of Chihuahua City, Mexico. The data obtained will be fieldtested via a hand held spectral radiometer in order to examine our results and identify the unknowns present on the map. We will continue to use satellite imagery to monitor the environment of the aquifer in the Tabalaopa Valley.

References

Chavez A., 1993 Hidrogeologia fisica y quimica de la portion centro-occidental del vane de Tabalaopa-Aldama, Chihuahua. Master's thesis. Fat. de Ingenieria, Universidad Autonoma de Chihuahua.

Maldonado D., Megaw P. 1983 Geology of the Santa Eulalia Mining District, Chihuahua, Mexico. Geology of the mineral resources of north-central Chihuahua. Guidebook for the 1983 field conference, El Paso Geological Society, p. 367.

Mauger R.L., 1993 The geology and volcanic stratigraphy of the Sierra Sacramento block near Chihuahua City, Chihuahua, Mexico. Geology and mineral resources of north-central Chihuahua. Guidebook for the 1993 field conference, El Paso Geological Society, October 1993. p. 137.

Rodriguez J.A., 1997 Pingitore N. E., A hydrogeologic and remote sensing investigation of aquifer contamination by nitrates from *aguas negras in* Chihuahua, Mexico. 1997 URC-TC Conference. p. 639.

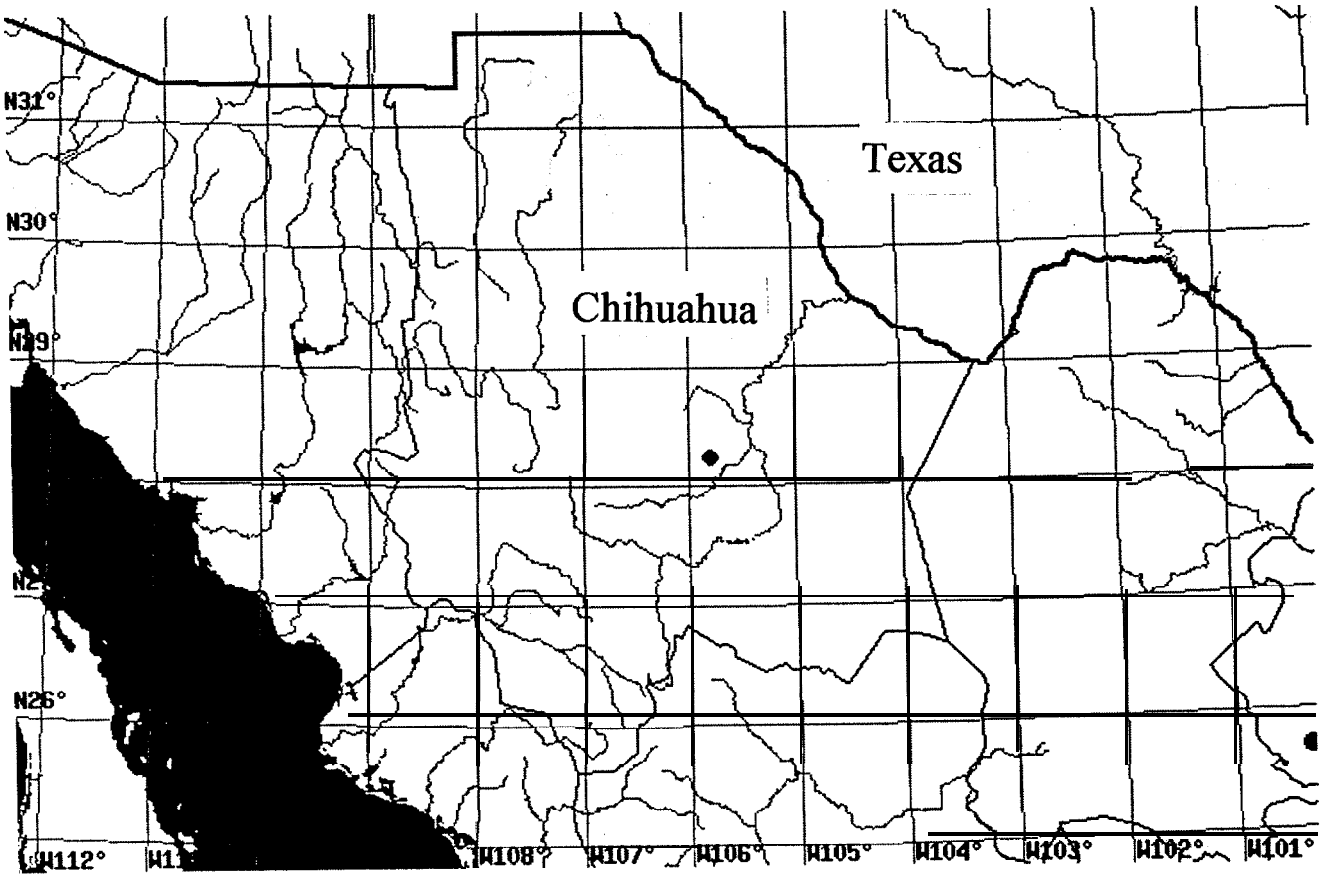
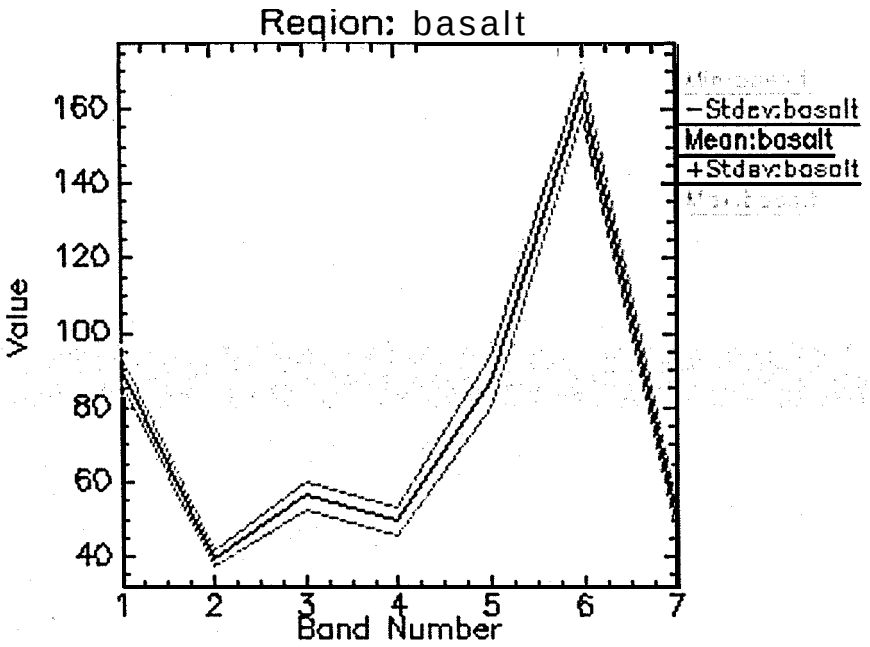
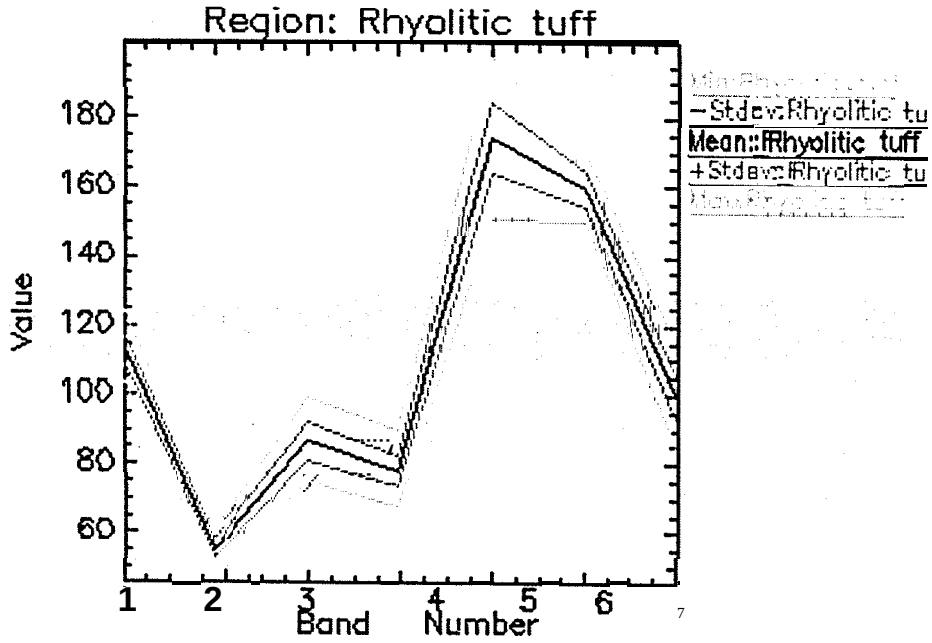
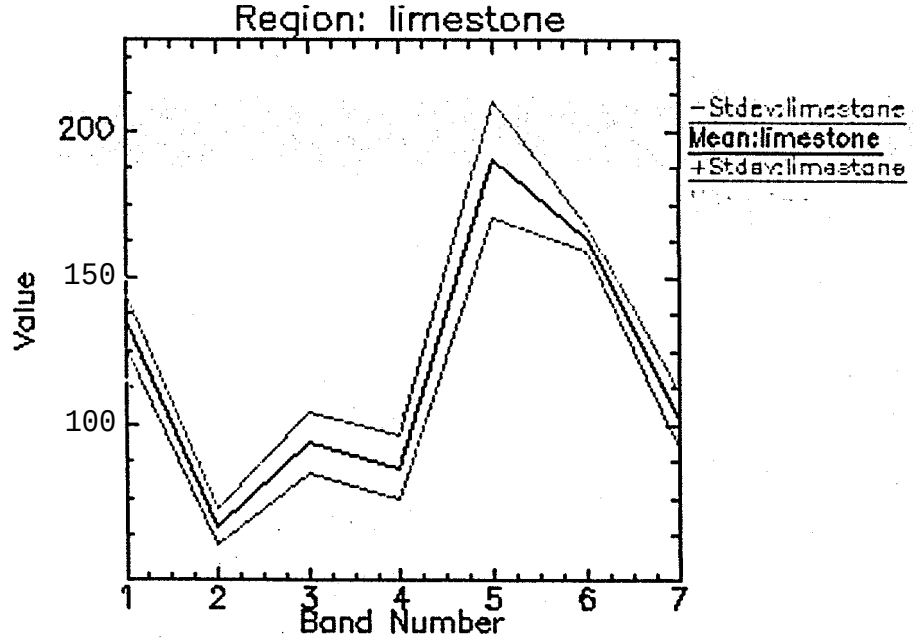


FIG 1 Area of study

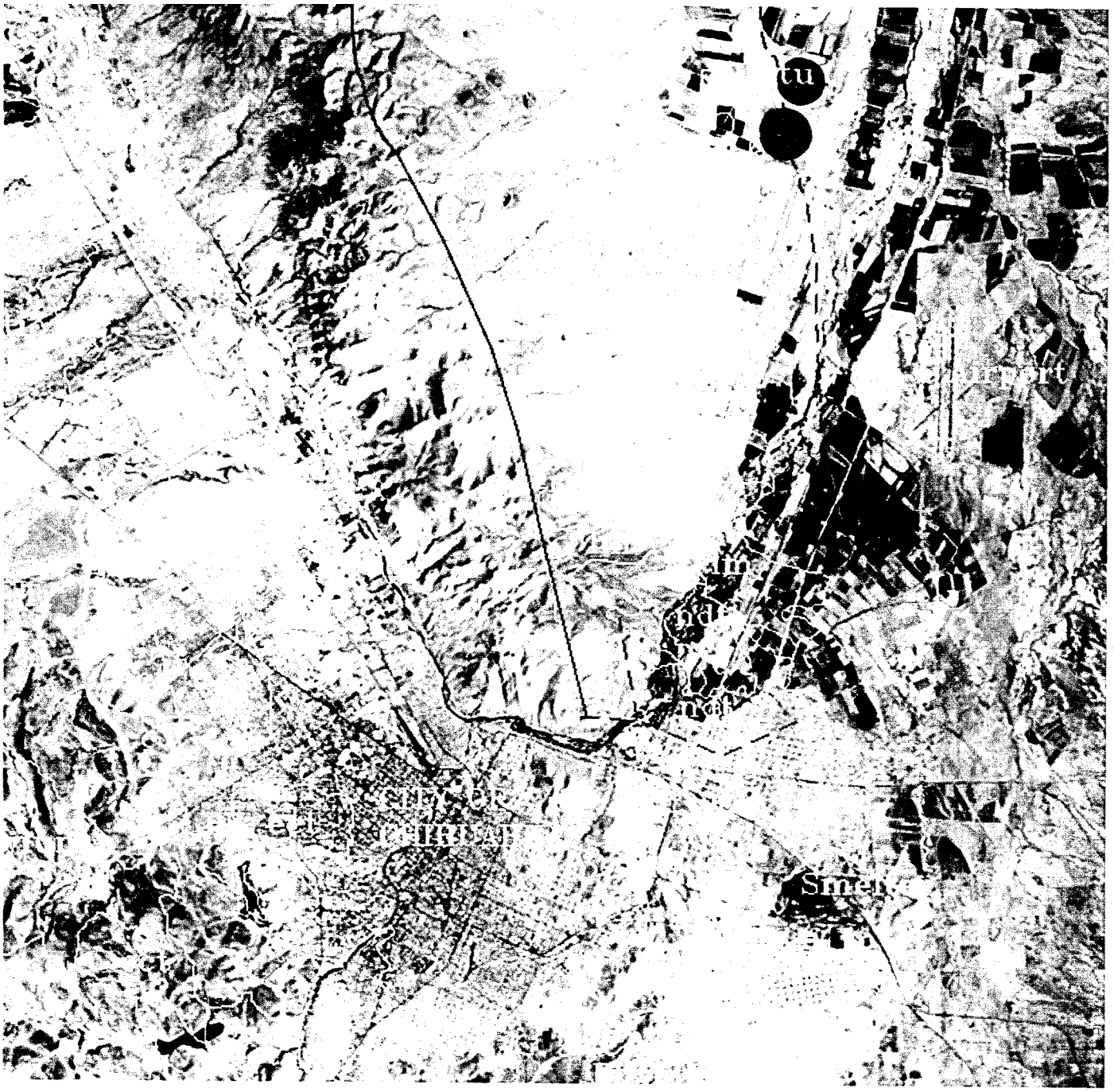






'FIG II

Plots showing the **minimum, maximum,** and mean values of reflectance for three of the major **lithologic** units in the area. The regions were first selected on the basis of identification on both the geologic map and the satellite image. The statistical mean values were then calculated for the entire region in order to obtain mean values. These data were acquired for bands 1-7.



FIGIII

TM image 04/02/86. Sacramento Range is depicted here N-W of the City of Chihuahua. The Geology of the major rock units is shown as well as a major fault that runs through the middle of the range. Note the Sacramento River that comes into the city from the N-W and flows to the N-E to the Tabalaopa Valley by the airport.

An Integrated Geophysical Analysis of the Albuquerque Basin

Cindy L. Gillespie, G. Randy Keller
Pan-American Center for Earth and Environmental Studies
PACES
The University of Texas at El Paso

Introduction

The Albuquerque Basin is one of the largest and deepest basins associated with the Rio Grande rift. Several deep wells and tens of kilometers of seismic reflection data, interpreted by water utilities and oil and gas companies, provide a good model of the structure and stratigraphy of the area. Seismic reflection data, for example, define large scale features such as the geometry of the basins and the presence of faults (Russell and Snelson, 1994). Deep well data allow stratigraphic correlations within the basin (Lozinsky, 1994). Integration of these results with gravity and remote sensing tools provides a "3D" picture of the basin.

Water usage from the Albuquerque Basin has increased due to industrial and residential use. It has been proposed that most of the shallow, highly permeable part of the aquifer has already been "tapped out (Kelly, 1994)." Locating alternative aquifers, to supply the growing demand for more water, are costly.

Regional Geology

Late in the Cenozoic Era, a series of Rio Grande rift related basins were formed. The Albuquerque basin is located south of the Española Basin and north of the Socorro Basin. It is bound to the west by the Colorado Plateau and to the east by a series of mountains and uplifts. It is thought to be the most complex of all the Rio Grande rift basins. It covers a 6,000 km^2 area and is subdivided by the Tijeras Accommodation Zone, a major strike/slip fault zone (Chapin and Cather, 1994) (Figure 1).

Oligocene to Middle Pleistocene sediments of the Santa Fe Group form the major components of the synrift basin fill. The basin fill ranges in thickness from 1-2,000 meters along the basin margin to 4,400 meters in the center. Compositionally, the basin till consists of intertongueing piedmont sediments (poor to well sorted), fine to coarse-grained sandstone, siltstone, conglomerate, and claystone. In addition, sediments derived from mafic bodies, alluvial fans, and fluvial (specifically axial river) deposits are found. The Santa Fe group overlies a "basement" of Mesozoic and Paleozoic sedimentary deposits overlying a Pre-Cambrian crystalline basement.

Groundwater Resources

Seismic reflection (covering an area of 200 km) and well (12 deep wells, ranging up to 4 km deep) data were used in this study to construct a preliminary model of the Albuquerque basin. Our model shows that the Albuquerque basin is a characteristic graben bounded by normal and listric faults. The primary aquifer unit belongs to the Santa Fe group and is segmented by these grabens (Figure 2). The most permeable zone is located in the upper part of the basin (Figure 2). Scientists estimate the aquifer to be as much as 6,000 feet thick. However, these authors are criticized for withholding the fact that the water is saline and non-potable. Scientists are beginning to recognize the productive zones are much thinner and less aerially extensive than originally estimated.

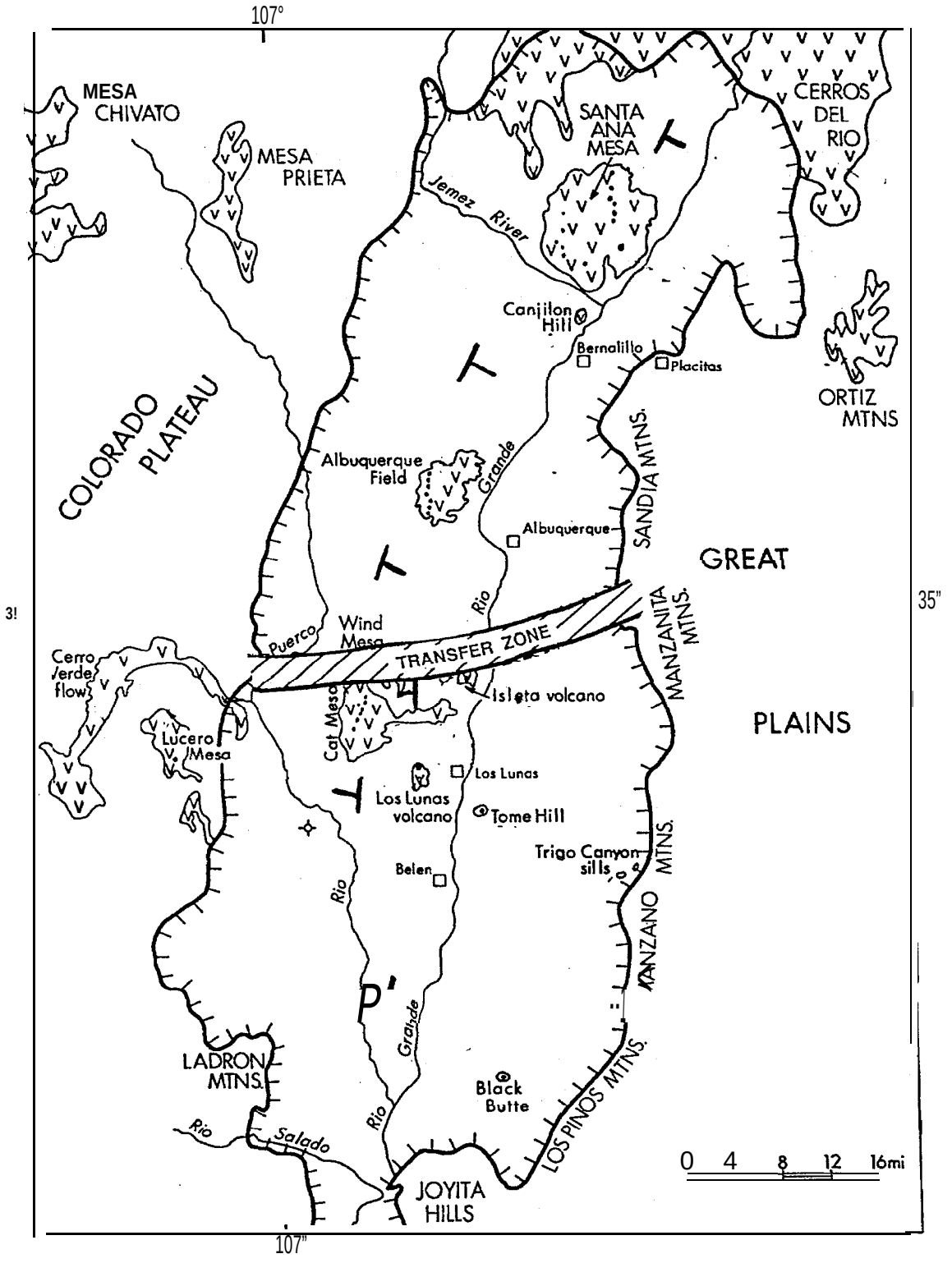


Figure 1. An index map of the Albuquerque Basin Area. The basin is subdivided into two sub-basins divided by the Tijeras Accommodation Zone. The north basin dips southeast and the southern basin dips to the southwest.

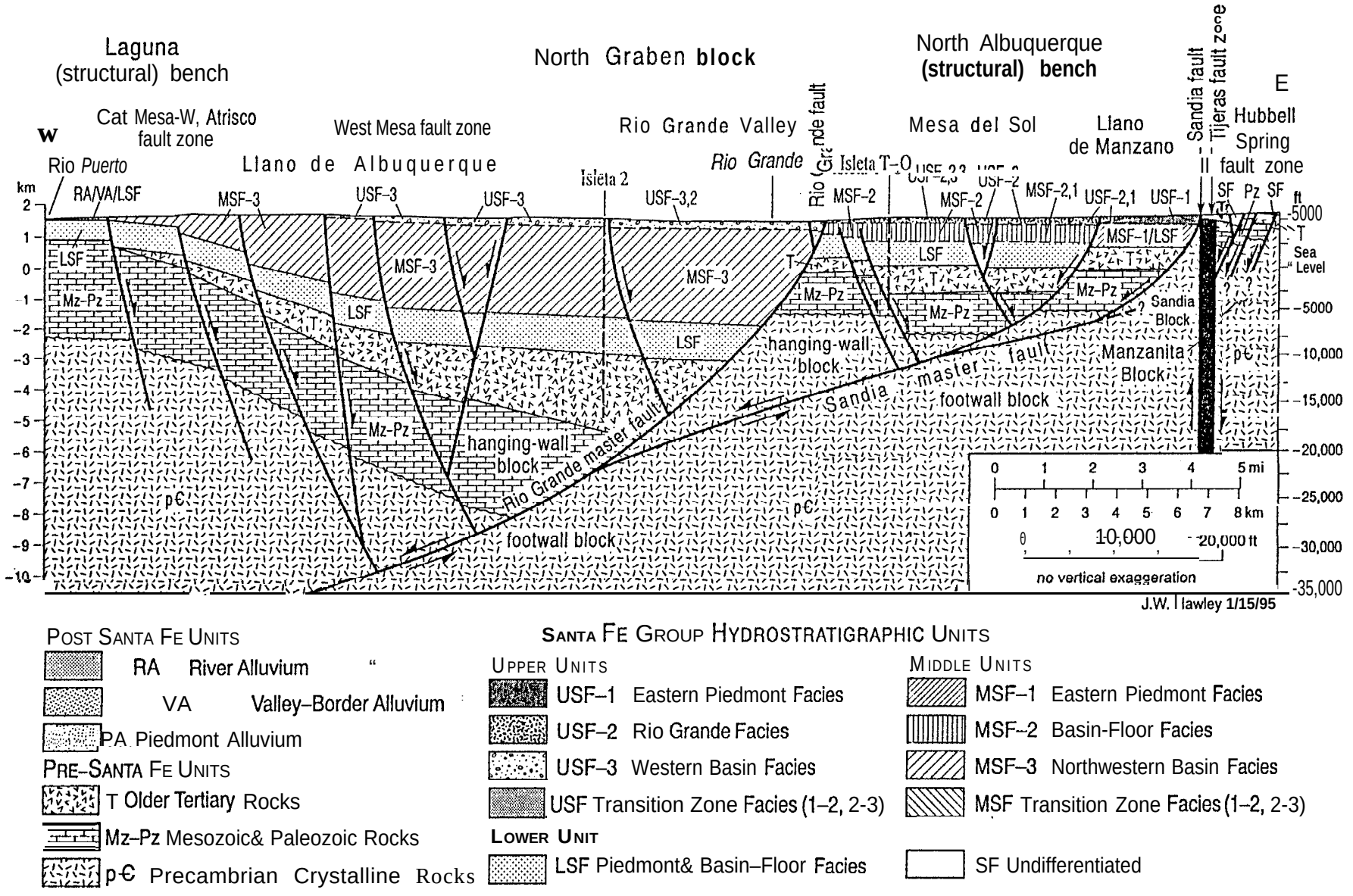


Figure 2. **Stratigraphic profile** across the North Albuquerque sub-basin. Basin fill is about 10,000 ft. thick in the central part of the **graben**. The **aquifer** is in the upper to middle Santa Fe Group reeks. Permeability decreases and mineral **content** increases with depth (from Hawley et al, 1994).

Kelly (1994) asserts that the most prolific *portion* of the aquifer is already depleted. Zones west of the Rio Grande are proposed for alternative aquifer locations; the area, however, is higher in mineral content including high arsenic levels in the ground water (Kelly, 1994). A second alternative water source involves deepening existing wells. However, the increase in mineral content with depth and decrease in permeability and water quality makes this an impractical alternative. A third alternative is to import water from outside the Albuquerque Basin to resolve water shortages in the area.

Data Analysis

Knowledge of the basin geometry is from interpretations of Shell Oil seismic reflection data. Basin fill information (stratigraphy correlation of deep wells and reports) comes from USGS concerns over the aquifer and ground water longevity of the Albuquerque Basin. Gravity data for this study is from the UTEP database and is supplemented with data collected during the summer of 1997 using standard gravity field collection methods. Standard corrections accounting for latitude, elevation, and terrain were applied to the data. A density of 2.67 gm/cc was used for the Bouger reduction. The resulting gravity map is shown in Figure 3. A gravity model constructed across the northern sub-basin is shown in Figure 4.

Summary and Conclusions

One result of this study is the interpreted outline of the deep portion of the basin shown in Figure 3. Another is a cross-sectional model of the basin structure confirming the presence of a deep basin filled with unconsolidated sediments (Figure 4).

The fear of the basin "drying up" may not be warranted in the case of the Albuquerque Basin. The water situation in the Albuquerque Basin may seem dim since freshwater is rapidly being depleted with very little recharge coupled with increasing TDS with depth. However, the future may not be so bleak. Water quality of the deeper aquifer units can be improved for usage. Also, importing water from other locations is an alternative to solve the water crisis.

References

Chapin, Charles E. and Cather, Steven M., 1994, Tectonic Setting of the Axial Basins of the Northern and Central Rio Grande Rift, In Keller, G. Randy and Cather, Steven M. (Eds.), Basins of the Rio Grande Rift: Structure, Stratigraphy, and Tectonic Setting, Geologic Society of America Special Papers, 234, p. 5-26.

Hawley, John W.; Haase, C. Stephen; and Lozinsky, Richard P., 1994, An Underground View of the Albuquerque Basin, In Proceedings of the 39* Annual New Mexico Water Conference Nov. 3-4, 1994, Albuquerque Convention Center, p. 37-56.

Kelly, T. E., 1994, Where Has All the Water Gone?, In Proceedings of the 39th Annual New Mexico Water Conference Nov. 3-4,1994. Albuquerque Convention Center, p. 169-178.

Lozinsky, Richard P., 1994, Cenozoic Stratigraphy, Sandstone Petrology and Depositional History of the Albuquerque Basin, Central New Mexico, in Keller, G. Randy and Cather, Steven M. (Eds.), Basins of the Rio Grande Rift: Structure, Stratigraphy, and Tectonic Setting, Geologic Society of America Special Papers, 234, p. 73-82.

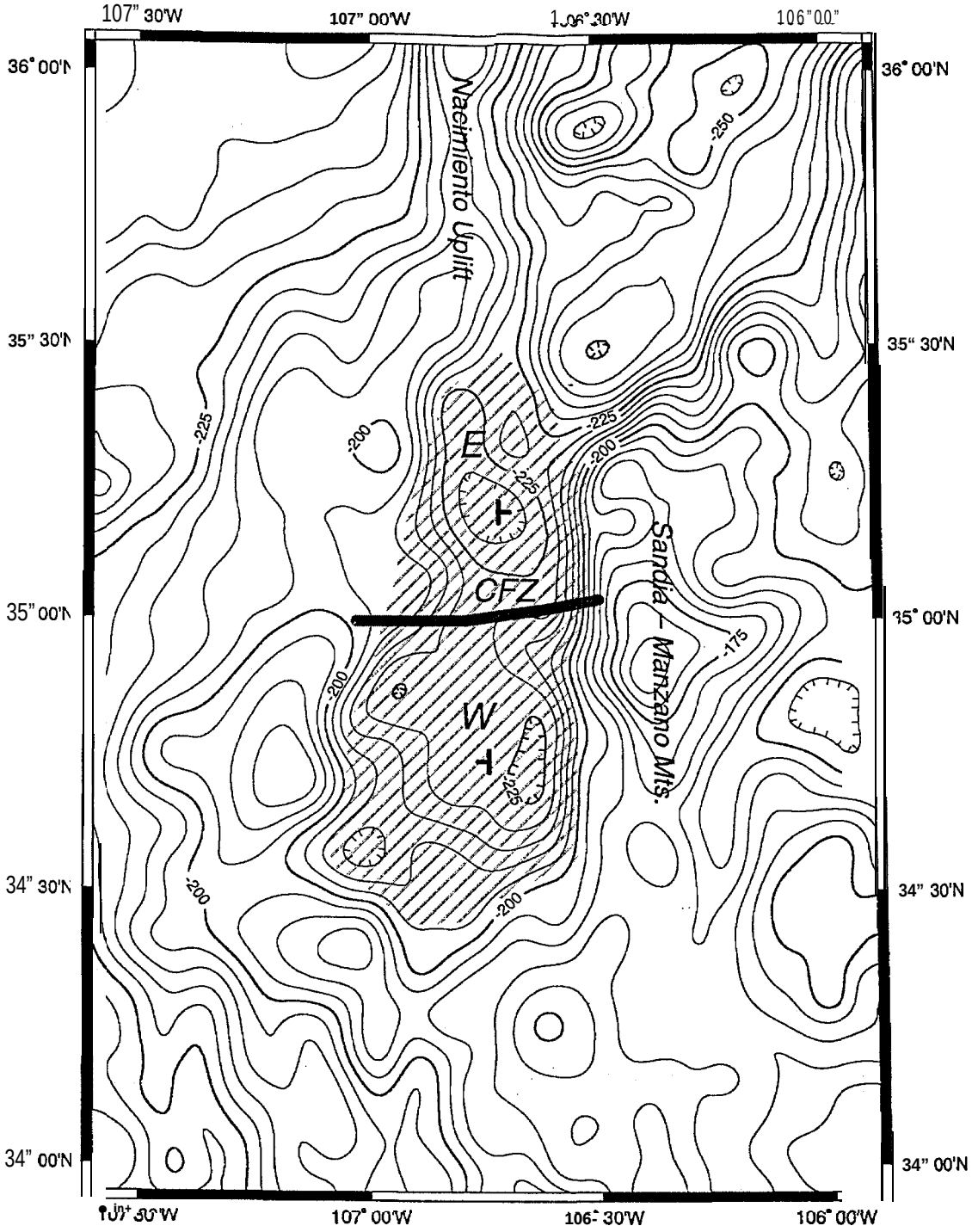


Figure 3. Gravity Map. E-denotes the East dipping northern sub-basin, W- denotes the West dipping southern sub-basin, CFZ-the Central Fault Zone.

Olsen, B. E., and J.E. Rakowski, 1979, Core Seismic Reflection Methods Verification Report, United States Dept. Of Energy San Francisco Operations Office.

Murphy, B. P., 1991, Determination of Shear Wave Velocity Structure in the Rio Grande Rift Through Receiver Function and Surface Wave Analysis, Dept. of Geological Sciences, unpublished Masters Thesis, The University of Texas at El Paso, El Paso, Texas.

Russell, L. R and Snelson, S., 1994, Structure and Tectonics of the Albuquerque Basin Segment of the Rio Grande Rift, In Keller, G. Randy and Cather, Steven M. (Eds.), Basins of the Rio Grande Rift: Structure. Stratigraphy, and Tectonic Setting, Geologic Society of America Special Papers, 234, p.83-112.

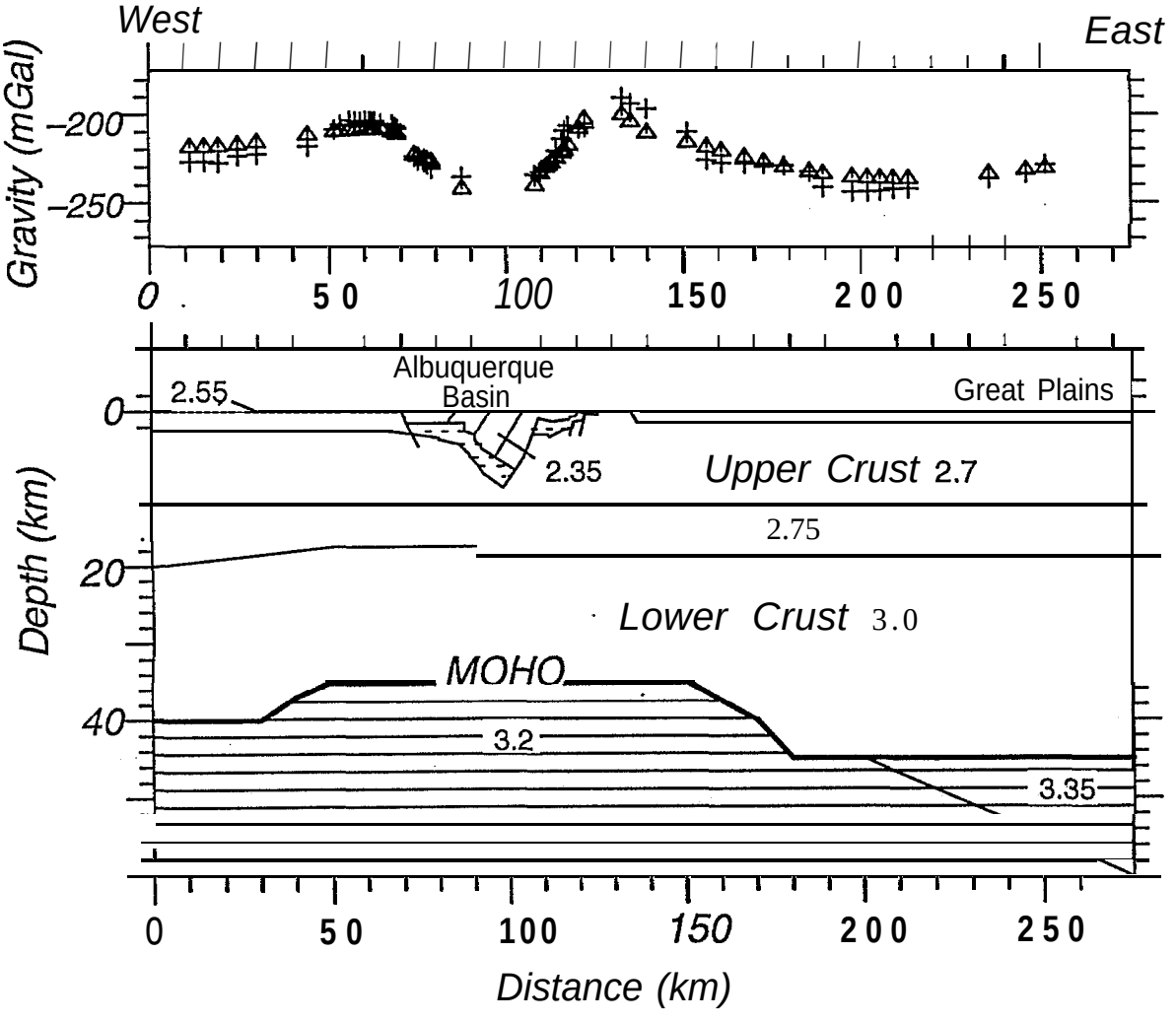


Figure 4. Computer model of a gravity profile across the Northern Albuquerque sub-basin. The gravity profile uses (Murphy, 1991) and (Olsen and Rakowski., 1979) as seismic constraints.

Spectral Characterization of Water Stress Impact on Some Agricultural Crops
I. Conceptual Model

Safwat H. Shakir and B. Girmay-Gwahid
Texas Gulf Coast Environmental Data Center
Prairie View A&M University
P.O. Box 235
Prairie View, TX 77446

Introduction

Potential crop growth and yield are usually not achieved because of constraints imposed by the plant environment. Water is an essential plant growth factor. Its reduced amount in the plant affects the plant photosynthesis which is in turn limits yield potential. Therefore adequate soil moisture has a large influence on plant growth. i.e., it promotes root growth and an abundant leaf area, increases the number of branches per plant. A number of plant traits are related with water stress particularly to drought resistance such as by stomatal resistance, osmotic adjustment, leaf orientation, size and rolling (Morgan, 1992).

In detecting stress on plants the response of the components of the plant canopy are important particularly for spectral identification. Studies have shown that the visible (red) and near infrared portion of' the spectrum give indications of the vegetation condition in terms of biomass, greeness and leaf area index. A combination of spectral bands has given a method of characterization. These include the greeness vegetation index (GVI)(Kuth and Thomas, 1976; Jackson, 1983), Normalized Difference Vegetation Index (NDVI: Rouse et al., 1974), Ratio index (NIR/R) (Tucker, 1979). Later a Transformed Vegetation Index (TVI) was used to avoid working with negative ratio values. Prepedicular Vegetation Index (PVI) (Richardson and Wiegand,1977)was also developed that account for soil background variations. Kauth and Thomas (1976) developed a technique for transforming Landsat MSS information in four-dimension data space using the four MSS bands. From this they calculated a Soil-Brightness Index (SBI). In addition, a Difference Vegetation Index (DVI) (MMR5-MMR6 (Shibayama and Akiyama, 1989) and a Product Ratio Vegetation Index (PRVI) [MMR3 x MMR4 x MMR6/MMR1 x MMR7] were also developed and tested. Satterwhite and Hanley (1987) have stated that rationing of spectral bands is effective for surfaces with high interband spectral contrast but is ineffective for surfaces with low contrast. They further indicated that the NIR/Red and NDVI ratios show an association with plant available water gradients and the drought tolerance or drought avoidance mechanisms. Traditional] y the above ratio indices were used with Thematic Mapper and Multispectral scanners. With the advent of hyperspectral spectrometers more detailed data could be obtained about stress in vegetation by using a combination of very narrow bands in the NIR and Middle infrared region.

The spectral vegetation indices, particularly the NDVI, arc most often **used to** observe vegetation pattern and dynamics from satellite sensors (Tucker *et al.,* 1985). At local scale, various plant parameters such as LAI, biomass production, and percent cover (Asrar *et al., 1984)* arc related to NDVI. Although NDVI has shown to minimize view angle effects relative to variations in the red and NIR bands, it remains sensitive to atmosphere, soil, and view sun angle conditions limiting its qualitative capabilities to characterize the vegetation surface (Huete *et al.,* 1992). Atmospheric influences lower the NDVI of vegetated areas whereas dark or wet soil backgrounds raise the NDVI for a given canopy. In partially vegetated surfaces illuminated from small solar zenith angles have lower NDVI values due to a higher proportion of sunlit soil (Huete, 1987). At larger solar zenith angles, the percent of direct flux reaching the soil biomass is small, and the scattering properties of the vegetation dominate, resulting in higher vegetation index values (Kimes *et al., 1985).* Huete *et al.*, (1985) have stated that observed spectral patterns involving constant vegetation amounts with different soil backgrounds could not be explained nor predicted by either the ratio or orthogonal greeness measures. All their greeness measures were found to strongly depend on soil brightness. Soil-induced greeness changes became greater with increasing amounts of vegetation up to 60% green cover. The results suggest that soil and plant spectra interactively mix in a non-additive, partially correlated manner to produce composite canopy spectra (Huete *et al.,*

1985), In order to minimize these problems, Huete (1988) hypothesized a soil adjusted vegetation index (SAVI) and later Huete *et* al., (1992) introduced SAVI -view angle response to be minimized with a cosine function.

In the broad Thematic Mapper (TM) band 5 region (1.55- 1.75μm) the information collected is useful in crop drought and plant vigor studies. Rees (1990) presented a figure on reflectance of upper surface of Sycamore leaf at different moisture contents in which these data showed that the middle infrared band (1,3-1 .8μm) had differentiated the impact of water reduction on the reflectance of the Sycamore leaves,

From the above discussion, we can see the validity of using the near infrared and middle infrared bands in detecting the crop stress. Then with the advent of the new Hyperspectral spectrometer, more detail data could be obtained about the stress in vegetation by using the combination of a very narrow bands in the NIR and MIR region. Therefore; in the present paper we will introduce a new conceptual model to evaluate the crop stress or vegetation based on combinations of very narrow bands in NIR and MIR region.

Theory

The radiance leaving from a plant canopy is a complicated mix of signals due to many components. These include the spectral absorbance and transmittance by plant components, spectral reflectance from plant parts, spectral reflectance from background (soil), scattering by plant and soil components. Therefore, the plant leaving radiance (L) is assumed to be composed of the following 3 components: plant canopy parameters, soil reflected and external parameters (*eg*. sensor and atmosphere).

Remotely sensed reflectance from a healthy or stressed crop canopy can be calculated or predicted from a combination of crop and vegetation indices. This will account the ratio indices, normalized indices and transformed indices. The most common reflectance parameters that most researchers applied are the near infrared and red spectra. As the reflectance spectra curves from a scene vary in shape and magnitude over a wide range of wavelengths the use of indices is selected primarily in the red and infrared regions and further extended to other wavelengths of importance with more data such as that obtained from hyperspectral imagery. This will account to the number of spectral components in the soil and vegetation population, with the response of each component at a particular wavelength and the percentage contribution of each component in a particular scene.

Hypothesis and Analysis

A narrow band resolution of TRWIS III and derived indices will detect vegetation condition (stress) better than data obtained from broad bands such as Multispectral Scanner, Thematic Mapper and AVIRIS.

Turgidity in leaves (unstressed) crops is assumed to have a maximum peak of extinction (*i.e.,* the absorption of light at a particular wavelength is greatest). Therefore, the maximum reflectance at the near infrared by a higher and unstressed vegetation cover could correspond with the lowest dip in the middle infrared and is assumed to be useful in vegetation stress index studies. It is anticipated that the use of vegetation stress index normalizes the differences accounted for by sensor variations. Therefore, the idea is to solve for a new index while combining all relevant parameters that could also be used to give an explanation for the nature of crops in the field,

For this the new index is named as a Hyperspectral Structure Component Index (HSCI). This model will combine a stress ratio(SR: NIR/MIR) of the bands at the NIR (0.76 -0.90μm) and MIR (1.3 - 2.5um) in addition to other established ratio indices.

In developing this index we used the simple canopy reflectance relationship that has been given by Baret *et al., (1989)* where an extinction coefficient **(k)** relates to leaf area index (**LAI**) as follows:

$$R_c = R_\infty + (R_\infty - R_s)e^{-k.LAI} \quad [1]$$

Where R_c = Canopy reflectance
R_s = Soil reflectance
R_∞ = Infinite canopy reflectance (when LAI is infinite)
k = Reflectance extinction coefficient

While incorporating all components of vegetation, background reflectance and sensor parameters in a simplified model, at the same time this approach will simplify and help us visualize based on a single value. This also can provide measurement to predict the status of a vegetation stress. For this, the stress ratio is combined with the R_∞ in synthesizing the single value. Table1 &2 show hypothetical and extracted data from other publications (Huete *et al,,* 1985, Major *et al.,* 1992, Rees, 1990) which incorporate one or more of the following parameters: LAI, k, NIR, MIR values in the range of extinction coefficients (0.3-1 .3), LAI (0-7), and NIR at 0.76μm and MIR at 1.45μm. The transformation of equation [1] into the index takes the form

$$HSCI = (RI-R_\infty) + (R_\infty - R_s)e^{-k.LAI} \quad [2]$$

where RI refers to Ratio Index (NIR/MIR).

Fig 1 & 2 show the visibility of the index for measuring stressed and unstressed crops. In this respect, the HSCI index has shown the sensitivity towards the moisture status in the. crops that will help detect the problem before the vegetation are wilted and stressed permanently. This new index will eliminate the disadvantage of IR/Red index which is less responsive to water stress in plants such as alfalfa (Moran *et al.,* 1989).

Table 1. Hyperspectral Structure Component Index of unstressed and stressed leaves at low coefficient of extinction.

LAI	K	Low R_S	R_C	HSCI unstressed	HSCI Stressed
0.00	**0.30**	**0.25**	**0.75**	1.13	0.26
0.10	**0.30**	**0.23**	**0.76**	1.12	0.25
0.50	**0.30**	**0.22**	**0.74**	1.24	0.37
1.00	0.30	0.21	0.71	1.17	0.30
1.50	0.30	0.20	0.69	1.17	0.30
2.00	0.30	0.19	0,67	1.18	0.31
2.50	0.30	0.17	0.65	1.19	0.32
3.00	0.30	0.15	0.64	1.20	0.33
3.50	0.30	0.11	0.63	1.22	0.35
4.00	0.30	0.09	0.62	1.23	0.36
7.00	0.30	0.08	0.55	1.28	0.41

Conclusions

The index proposed is promising in that the values of unstressed and stressed crops could be differentiated. The contrast is clear that the HSCI index values of stressed are above 1.0 while the stressed are below 1.0. This hypothesis will be tested and further validated from remotely sensed hyperspectral reflectance and images. Research work under this direction is being undertaken with hand-held radiometer measurements and in comparison with AVIRIS data. This will give the potential to explore the capabilities of the hyperspectral spectrometer in detecting stress in agricultural crops such as the TRWIS III and the forthcoming Lewis Satellite.

Fig. 1 Relationship between the Hyper-spectral Structure Component Index (HSCI) and Leaf Area Index (LAI) at high extinction coefficient

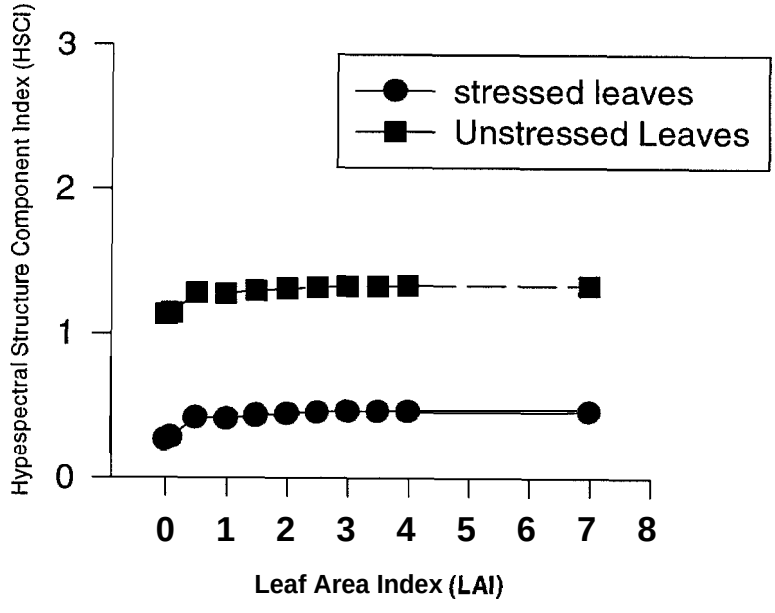


Fig. 2 Relationship Between Hyper-spectral Structure Component Index (HSCI) and Leaf Area Index (LAI) at low extinction coefficient

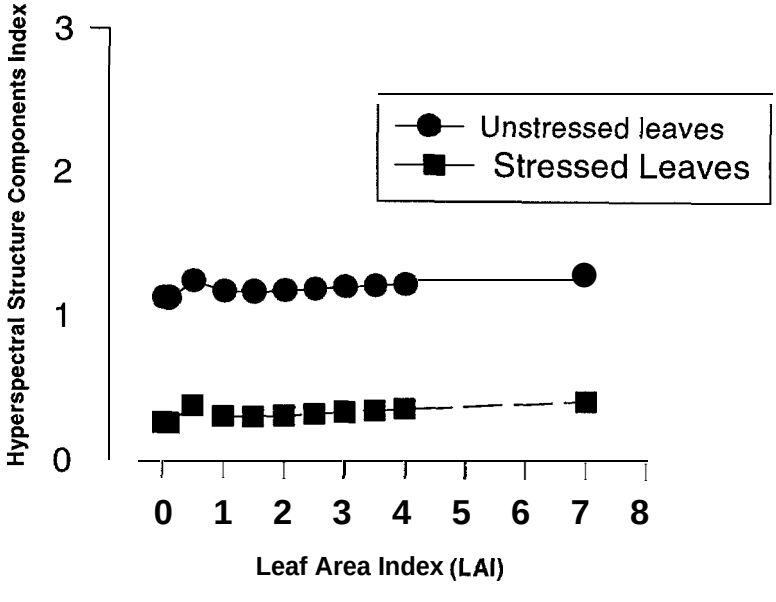


Table 2. Hyperspectral Structure Component Index of unstressed and stressed leaves at high coefficient of extinction.

LAI	K	High	R_S R_C	HSCI unstressed	HSCI Stressed
0.00	1.30	0.25	0.75	1.13	0.26
0.10	1.30	0.23	0.74	1.14	0.27
0.50	1.30	0.22	0.65	1.27	0.40
1.00	1.30	0.21	0.58	1.27	0.40
1,50	1.30	0.20	0.54	1.29	0.42
2.00	1.30	0.19	0.52	1.30	0.43
2.50	1.30	0.17	0.51	1.31	0.44
3.00	1.30	0.15	0.51	1.32	0.45
3.50	1.30	0.11	0.50	1.32	0.45
4.00	1.30	0.09	0.50	1.32	0.45
7.00	1.30	0.08	0.50	1.33	0.46

References

Asrar, G. Fuchs, M., Kanemasu, E. T. and Hatfield, J. L. 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. *Agron. J.* 76:300-306.

Baret, F., G. Guyot and D.J. Major (1989). Crop biomass evaluation using radiometric measurements. *Photogrammetria (PRS) 43:241-256.*

Huete, A. R. 1987. Soil and sunangle interactions on partial canopy spectra. Int. Jour. *Remote Sens.* 8:1307-1317.

Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25:209-309.

Huete, A. R. and Escadafal, R. 1991. Assessment of biophysical soil properties through spectral decomposition techniques. *Remote Sens. Environ.* 35: *149-159.*

Huete, A. R. And Warrick, A. W. 1990. Assessment of vegetation and soil water regimes in partial canopies with optical remotely sensed data. *Remote Sens. Environ. 32:155-167.*

Huete, A. R., G. Hua, . Qi, A. Chehbouni. 1992. Normalization of multidirectional red and NIR reflectances with the SAW. *Remote Sens. Environ. 41:143-154.*

Huete, A. R., R.D. Jackson and D.E. Post. 1985. Spectral response of a plant canopy with different soil backgrounds. *Remote Sens. Environ.* 17:37-53.

Jackson, R. D. 1983. Spectral indices in n-space. *Remote Sens. Environ. " 13:409-421.*

Kauth, R.J. and G.S. Thomas. 1976. The tasselled cap a graphic description of the spectral temporal development of agricultural crops seen by Landsat; Proceed. Symposium Machine Processing of Remote Sensing Data, LARS, Purdue University.

Kimes, D. S., Norman, J. M., and Walthall, C.L. 1985. Modeling the radiant transfer of sparse vegetation. IEEE Trans. Geosci. *Remote Sens. Environ.* 12:247-254.

Major, D. J., F.J. Larney, and C.W. Lindwall. 1992. Spectral reflectance characteristics of wheat residues. In Remore Sensing Science for the nieties (R. Mills, ed. Vol. 1), 10th Intern. Geoscience and Remote sensing Symposium. The University of Maryland, College Park, MD. Pp.603-607.

Moran, M. S., P.J. pinter, Jr., B.E. Clothier and S.G. Allen 1989. Effect of' water stress on the canopy architecture and spectral indices of irrigated alfalfa. *Remote Sens. Environ. 29:251-261,*

Morgan, J.A. 1992. Drought stress in winter wheat: mechanisms and strategies. In 1992 Wheat technology Conference. USDA-ARS. Fort Collin, Colorado, USA.

Rees, W.G. 1990. Physical Principles of Remote Sensing. Cambridge University Press. Cambridge, UK.

Richardson, A. J., and Wiegand, C.L. 1977. Distinguishing vegetation from soil background information. *Photogramm. Eng. Remote Sens. 43:1541-1552.*

Rouse, J. W., Jr., Haas, R. H., Schell, J.A. and Deering, D.W. 1974, Monitoring vegetation systems in the great plains with ERTS. In Third ERTS Sympos., NASA SP-351, U.S. Govt. Printing Office, Washington, DC, Vol. I, pp309-317.

Satterwhite and Hanley 1987. Spectral Characteristics of' selected soils and vegetation in Northern Nevada and their discrimination using band ratio techniques. *Remote Sens. Environ. 23:155-175.*

Shibayama, M. and Akiyama, T. 1989. Seasonal visible and near-infrared spectra of rice canopies in relation to LAI and above-ground dry mass. *Remote Sens. Environ. 27:119-127.*

Tucker, C. J. 1979, Red and photographic infrared linear combinations for monitoring vegetation. *Remote sens. Environ. 8:127-150.*

Tucker, C. J., Vanpraet, C.L., Sharman, M.J. and Van Ittersum, G. 1985. Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel. *Remote Sens. Environ. 17:233-249.*

USING VIDEOGRAPHIC REMOTE SENSING AND SPATIAL INFORMATION TECHNOLOGIES FOR DETECTING AND MAPPING PLANT SPECIES AND INSECT INFESTATIONS

J. H. Everitt, D. E. Escobar, R. S. Fletcher, and M. R. Davis
USDA, ARS
Remote Sensing Research Unit
2413 E. Highway 83
Weslaco, TX 78596
e-mail: j-everitt@tamu.edu
Fax: 956-969-4893

ABSTRACT

This paper presents an overview on the application of aerial videography with global positioning system (GPS) and geographic information system (GIS) technologies for natural resource management. Applications demonstrated include detecting and mapping: (1) saltcedar (Tamarix chinensis Lour.) infestations in a riparian area; (2) black mangrove [Avicennia germinans (L.) L.] populations in a coastal region; and (3) citrus blackfly (Aleurocanthus woglumi Asby) and harvester ant (Pogonomyrex barbatus F. Smith) infestations in agricultural areas and grazing lands, respectively. Video imagery (color-infrared and normal color) could be used to detect insect and plant infestations/populations. The integration of a GPS with the video imagery permitted latitude and longitude coordinates of insect and plant infestations/populations to be recorded on each image. The GPS coordinates were entered into a GIS to map insect and plant infestations/populations on a regional scale. The integration of videography, GPS, and GIS are valuable tools that can enable resource managers to develop maps showing the distribution of insect and plant infestations/populations over large areas. The digital video imagery can serve as a permanent geographically located image data base for monitoring future contraction or spread of insect and plant infestations/populations over time.

INTRODUCTION

Over the past several years videography has shown considerable potential as a remote sensing tool for natural resource assessment (Meisner and Lindstrom, 1985; Nixon et al., 1987; King and Vlcek, 1990; Everitt et al., 199 1). Video technology has many attributes that are attractive for remote sensing, but the most prominent are its real-time monitoring capability and immediate availability of the electronic signal for both visual interpretation and digital processing. In addition, video surveys can be conducted for approximately one-half the cost of photographic surveys (Everitt et al., 1994b).

Recently, aerial videography and global positioning system (GPS) technology have been integrated and shown to be useful tools for detecting and monitoring insect activity over forested areas (Bobbe and Ishikawa, 1992; Myhre, 1992). The latitude/longitude data provided by the GPS were entered into a geographic information system (GIS) to georeference forest pest problems (Myhre, 1992). Richardson et al. (1993) entered aerially obtained GPS coordinates into a GIS to map the distribution of undestroyed cotton (Gossypium hirsutum L.) fields in a regional management program for boll weevil (Anthonomus grandis Boheman) in south Texas (Summy and King, 1992). Everitt et al. (1993, 1994b) used airborne videography and GPS technology to detect brush and weed infestations on rangelands and entered the GPS data into a GIS to map noxious plant populations over an extensive area.

The objective of this paper is to present an overview on the integration of videography, GPS, and GIS technologies for entomological and botanical applications. Specific applications include detection and mapping: (1) saltcedar infestations in riparian areas; (2) black mangrove populations in a coastal region; (3) citrus blackfly infestations in an agricultural area and (4) harvester ant infestations in grazing lands.

GENERAL PROCEDURES

Data presented here has been previously published, Additional information on videographic, GPS, and GIS equipment, software, etc., can be obtained from the literature citations.

Imagery of black mangrove and citrus blackfly was acquired with a high resolution multispectral video system (HRVS) (Everitt et al., 1991). The system had three black-and-white (B&W) visible/near-infrared (NIR) (0.400 to 1.100 pm) light sensitive CCD video cameras and four super (S)-VHS recorders. Visible and NIR filters were placed over the camera lenses giving the system the capability to record selected wavelengths in the visible/NIR light region. Visible red (0.644 to 0.656 μm) and yellow-green (0.543 to 0.552 μm) filters were placed on two of the cameras, and a NIR (0.8 15 to 0.827 μm) filter was used on the third camera. The system generates two kinds of simultaneously synchronized video images: (1) real-time color-infrared (CIR) composite and (2) three band black-and-white imagery. Imagery of the harvester ant was obtained with a three-camera multispectral digital video imaging system (Everitt et. al., 1995). This digital system, an upgraded version of the original HRVS, had a computer with an image digitizing band (640 x 480 pixel resolution), and only one super (S)-VHS portable recorder. The NIR, red, and yellow-green B&W image signals from the cameras were subjected to the RGB inputs of the computer digitizing board, giving a CIR composite digital image. The hard disk can store 1000 CIR composite images. In addition, the cameras' signals were also subjected to a color encoder that provides an analog CIR composite and stored on the S-VHS recorder. The analog CIR imagery recording served as a back-up in the event the computer malfunctions. Imagery of saltcedar infestations was acquired with a conventional color camcorder and a three-camera color video system that provided wide angle coverage (Everitt et. al., 1996a). These video systems were mounted vertically in the floor of a fixed-wing aircraft. Altitudes from which imagery was acquired and the geographic locations are given with each application.

A Trimble* Transpack [1 GPS was integrated with the video systems. The GPS was equipped with a navigation system which constantly received data from GPS satellites and readily calculated and displayed continuously the flight direction (bearing), altitude above sea level, time, ground speed, date, and latitude/longitude coordinates of the aircraft location above ground. Two modes are used to obtain the GPS data, either single or simultaneously. In one mode a Compix interphase was used with the GPS which permitted the transfer and recording of this continuous information on the last two lines of the imagery. The latitude/longitude coordinates on the video correspond to approximately the center of each scene. The second mode was the "way point" in which "bull's eye" coordinates were manually entered and stored in the GPS memory (a storage capacity of 999 coordinate points). With this mode the coordinate data were entered directly into the computer. The accuracy of the GPS was approximately ±100 m from the center coordinates of each video scene, which was adequate for detecting most insect and plant infestations/populations. A differential GPS can provide submeter accuracy, but these systems are costly.

Personal computer ARC/INFO-GIS and ATLAS-GIS software were used to generate regional and detail maps based on the 1990 post census TIGER/Line files for the state of Texas. The U.S. Census Bureau developed and trademarked a machine readable referenced map base called TIGER (Topologically Integrated Geographic Encoding and Referencing) for the 1990 census. The TIGER map-based system was constructed using USGS 1:100,000 scale digital line graph maps. The ATLAS-GIS maps provided greater detail of the counties such as major highways, roads, hydrography and other landmarks. These maps were produced to geographically map locations of insect and plant infestations/populations in the counties using airborne survey GPS data.

Ground truth data were collected at the study sites at or near the time imagery was obtained. Data were recorded relative to insect and plant infestations, plant species, plant height, density, cover, soil type, and soil surface conditions, Ground photographs were obtained to help interpret the remote sensing data. Photographs of video images shown here were taken from a high resolution image display monitor.

RESULTS AND DISCUSSION

SALTCEDAR

Saltcedar, also known as Chinese tamarisk, is a shrub introduced to the United States from Asia for use as an ornamental and for erosion prevention of streambanks (Baum, 1967). Saltcedar is an invader of riparian sites in the southwestern United States and northern Mexico where it forms dense, low thickets that displace

* Trade names are included for the benefit of the reader and do not imply an endorsement of or a preference for the product listed by the U. S. Department of Agriculture,

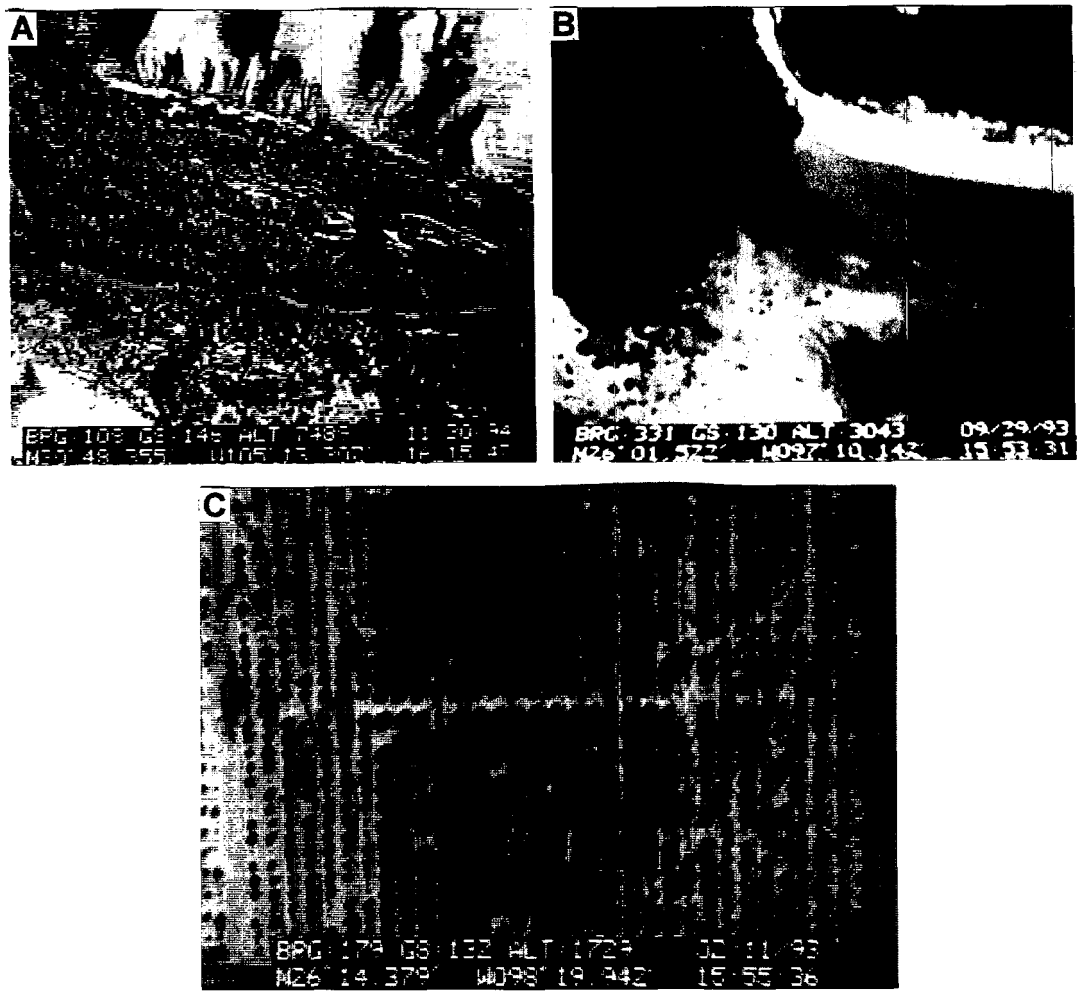


Figure 1. Conventional color (A) video image of a saltcedar infestation on the Rio Grande River near Candellaria, TX. The image was obtained at an altitude above ground level of 1500 m in November 1994. Color-infrared (B) video image of a black mangrove community near South Bay on the lower Texas gulf coast. The image was obtained at an altitude above ground level of approximately 915 m in September 1993. Color-infrared (C) video image of a citrus orchard near Mission, TX infested with citrus blackfly. The image was obtained at an altitude above ground level of 525 m in February 1993. The arrows on the prints (A, B, and C) point to the yellow-orange tone of saltcedar, the bright red image tone of black mangrove, and the dark tonal response of trees infested with citrus blackfly, respectively, The GPS data is shown at the bottom of each image.

native vegetation, impede water flow, increase sedimentation, use excessive water, and increase soil salinity (Horton and Campbell, 1974; Deloach, 1990). Saltcedar communities are also much less valuable for wildlife than are the native riparian communities they displace (Kerpez and Smith, 1989; Deloach, 1990). During late fall to early winter, the foliage of saltcedar turns a yellow-orange to orange-brown color prior to leaf drop. Previous research showed that saltcedar has higher visible light canopy reflectance than other associated plant species during this phonological stage, and subsequently, can be distinguished at this time on conventional color photography (Everitt and Deloach, 1990).

A project was recently completed using conventional color videography, GPS, and GIS technologies to distinguish and map saltcedar infestations in riparian regions of the southwestern United States (Everitt et al., 1996a). Study areas were along the Colorado River in southwestern Arizona and the Rio Grande and Pecos Rivers in west Texas. In this paper, only data obtained along the Rio Grande River will be presented.

Figure 1A shows a conventional color video image of a saltcedar infestation along the Rio Grande near Candellaria, Texas. Saltcedar has a conspicuous orange-brown image that is easily distinguished from other associated plant species, soil, and water. The distinct signature of saltcedar was due to the foliage turning a yellow-orange to orange-brown color. Saltcedar had an image tonal response similar to that shown in Figure 1A at all locations along the Rio Grande and could be easily distinguished at every site.

Figure 2 (upper right) shows a regional GIS TIGER map of extreme west Texas. The bold symbols along the left margin of the map represent GPS latitude-longitude coordinates of saltcedar populations along a portion of the Rio Grande. Areas with stars represent high populations of saltcedar, those with triangles have medium populations, and those represented by a plus sign have low populations. Population levels were assigned after a qualitative analysis of the video imagery of the area. Criteria for population levels were: >50% cover, high; 25-50% cover, medium; and <25% cover, low. Each symbol represents a composite of 2-3 video scenes because of the small map scale. A detailed GIS map of an area with several high saltcedar populations is shown in the center left portion of Figure 2. The map in the lower left portion of Figure 6 provides more detail of roads and hydrography associated with the saltcedar populations.

BLACK MANGROVE

Black mangrove, a widespread species of the American tropics and subtropics, is the only mangrove species that occurs along the Texas gulf coast (Correll and Johnston, 1970). Sherrod and McMillan (1981) reported on the distribution of black mangrove along the Texas coast and suggested that three major concentrations, i.e., Cavallo Pass in Calhoun County, Harbor Island in Nueces County, and Port Isabel-South Bay-Boca Chica of Cameron County, represented areas of continuous mangrove occurrence throughout the twentieth century. During periods of favorable weather, distribution expanded from these three centers and during unfavorable periods the populations contracted (Sherrod and McMillan, 1981).

Two severe freezes in December 1983 and 1989, respectively, did extensive damage to black mangrove communities along the Texas gulf coast (Lonard and Judd, 1985, 1991). Color-infrared aerial photography was used to determine the distribution of black mangrove following the 1983 freeze (Everitt and Judd, 1989). Everitt et al. (1996b) used CIR videography, integrated with GPS and GIS technologies, to determine the extent and distribution of black mangrove after the 1989 freeze. Figure 1B shows a CIR video image of a black mangrove community near South Bay, on the lower Texas gulf coast. The arrow on the print points to the distinct bright red image response of black mangrove compared with the light to dull shades of red and magenta of other associated vegetation. Dry bare soil has a whitish image tone, wet bare soil and litter have a gray tone, and water has various shades of blue. The GPS data is displayed at the bottom of the video scenes. Aerial videographic data showed that black mangrove occurred at 23 locations on the lower Texas coast (Port Isabel - South Bay - Boca Chica) and 15 locations on the lower-mid Texas Coast (Harbor Island). Ground truth reconnaissance confirmed the presence of black mangrove at all locations. The plants appeared to have fully recovered from the 1989 freeze, were actively growing, and had a good fruit crop.

Analysis of the video imagery obtained of the barrier islands in the Cavallo Pass area on the upper-mid Texas coast suggested that few black mangrove plants remained in this area. Ground visits to the area confirmed that none of the major mangrove communities remained there. However, several scattered plants were found around the perimeter of the islands with fewer plants found in the interior of the islands. Plants ranged in height

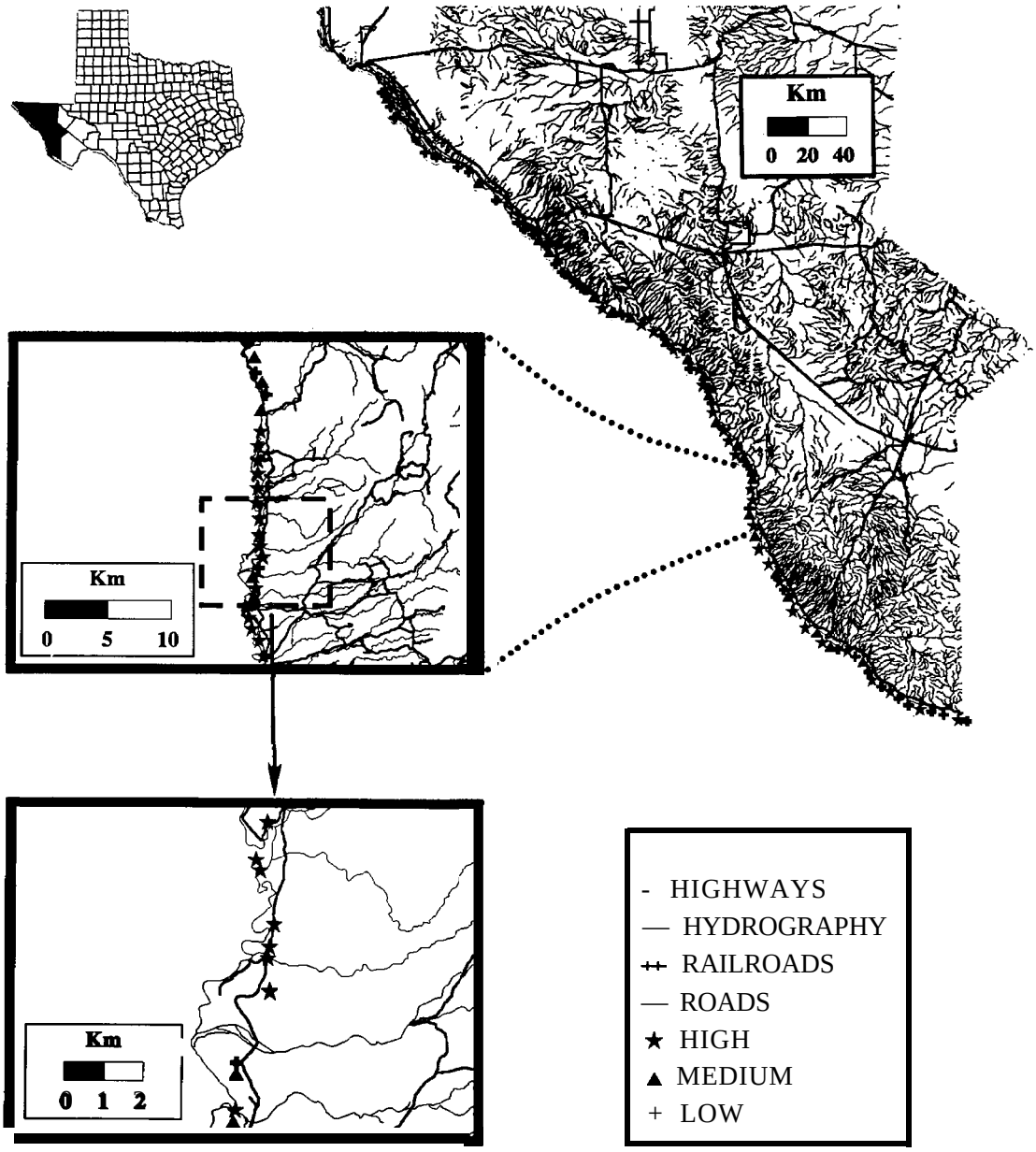


Figure 2. Regional GIS TIGER map (upper right) of extreme west Texas depicting GPS locations where saltcedar infestations occur along a portion of the Rio Grande River. The stars represent high populations of saltcedar, triangles are medium populations, and plus signs have low populations. Each symbol represents a composite of 2-3 video scenes because of the small detail of the map. A detailed map (left center) depicts an area with generally high populations of saltcedar. The lower left map shows even greater detail on the highly populated area.

from 54 to 140 cm, with a mean height of 87 cm. We observed no fruit on any of the plants. Based on a comparison of the current video imagery and ground observations to earlier maps, photography, and ground data of this area (Sherrod and McMillan, 1981; Everitt and Judd, 1989), the 1989 freeze reduced mangrove populations in the area to < 2% of what they had been prior to the 1983 freeze. The 1983 freeze reportedly reduced these populations to <5% of what they were before that freeze (Everitt and Judd, 1989).

No black mangrove plants could be detected in imagery acquired of Galveston Island on the upper Texas coast; however, a few isolated plants ≤ 75 cm tall were found in a ground visit. No fruit were observed on these plants. Black mangrove had been previously reported from this area (Sherrod and McMillan, 1981), but apparently the freezes of 1983 and 1989 have nearly eliminated this species from the upper Texas coast.

Figure 3 shows a regional GIS TIGER map of the entire Texas gulf coast. The approximate locations where black mangrove occurs are denoted: (1) Port Isabel-South Bay; (2) Harbor Island; (3) Cavallo Pass; and (4) Galveston Island. A detailed TIGER map of the Port Isabel-South Bay area is shown in the lower left part of the figure. The GPS latitude-longitude data provided on the video imagery of the area has been integrated with the GIS to georeference populations of black mangrove. The triangles depict the 23 locations where black mangrove occurs in this area. A detailed map (right center) is also shown for the Harbor Island area depicting the GPS georeference data for the 15 locations where black mangrove occurs in this area. A local site map presenting even greater detail of a highlighted area in the Harbor Island area is shown in the lower right portion of the figure. No detailed maps are provided for the Cavallo Pass and Galveston Island areas since few black mangrove plants occur there.

CITRUS BLACKFLY

The citrus blackfly, an exotic pest of citrus indigenous to the Asian tropics can cause extensive damage to citrus (Hart, 1978; Summy et al., 1983). Damage to citrus foliage caused by the citrus blackfly is of two types - actual feeding injury and the physiological damage caused by a sooty-mold fungus (Capnodium citri B. & D.) that develops on the honey dew excreted by the developing insect (Hart et al., 1976). When the sooty-mold deposits become heavy, they block sunlight and reduce the trees' ability to carry on photosynthesis, resulting in lower fruit production. The extent of damage by feeding injury can also be significant. Hart et al. (1973) used aerial color-infrared (CIR) photography to detect the presence of sooty-mold deposits caused by citrus blackfly activity in citrus foliage.

More recently, Everitt et al. (1994a) have used airborne CIR videography, GPS, and GIS technologies to detect and map citrus blackfly infestations in citrus orchards in the Lower Rio Grande Valley of extreme southern Texas. Figure 1-C shows a CIR video image of a citrus orchard near Mission, Texas. The arrow on the CIR print indicates the gray-black image response of the citrus blackfly-produced deposits of sooty-mold fungus on the tree foliage. The dark signature of the infested trees can be easily distinguished from the typical red-magenta tone of trees with no blackfly infestations. Trees with the heaviest sooty-mold deposits (upper center) have the darkest tonal response.

An analysis of the video imagery of the citrus growing areas of the Lower Rio Grande Valley showed that 27 orchards appeared to have citrus blackfly infestation (as indicated by leaf sooty-mold deposits). Ground truth surveys verified that 25 of these orchards were infested with brown soft scale (Coccus hesperidum L.), which also produces sooty-mold deposits on the foliage and yields the same video image tonal response as that of the citrus blackfly. Most of the blackfly infestations could be detected on the imagery acquired at an altitude of 1675 m, which had a horizontal ground pixel size of 1.8 m. However, infestations in newly planted orchards comprised of small trees could not be detected at this altitude because of their smaller canopies (<2m) and greater exposure of soil background. Likewise, infestations on replacement trees in mature orchards could not be detected at this altitude. The imagery illustrated in Figure 1-C was obtained at 525 m and provided a horizontal ground pixel size of 0.6 m. The lower altitude imagery was adequate for assessing both large and small trees. These results are in agreement with those of Hart et al. (1973), who used CIR aerial photography to detect citrus blackfly infestations.

The GPS latitude-longitude data was used with GIS technology to georeference citrus blackfly-infested orchards on a regional basis. The merger of the GPS data with a GIS is shown in Figure 4-A where a three-county (Cameron, Hidalgo, and Willacy counties) map based on ARC/INFO GIS TIGER/line file information

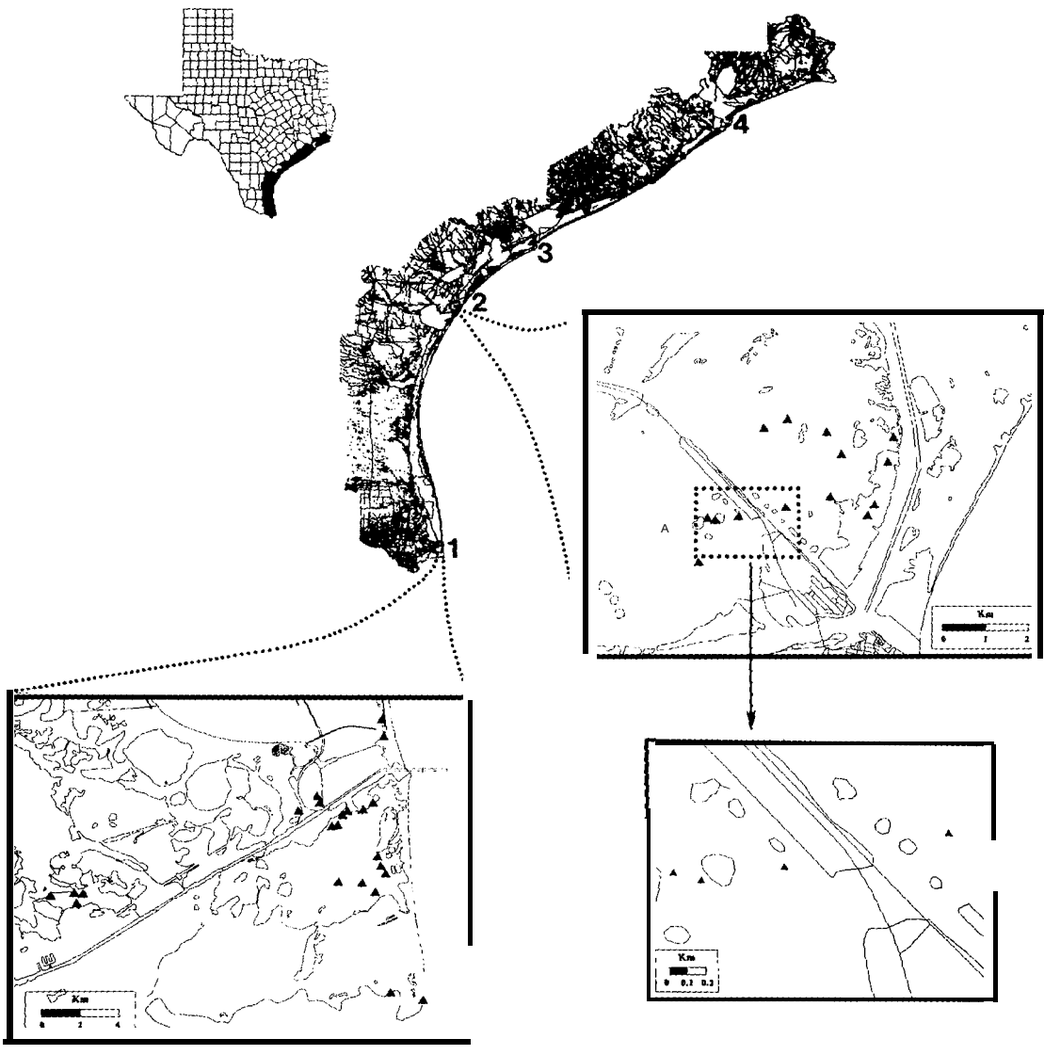


Figure 3. Regional GIS TIGER map (upper center) of the entire Texas gulf coast. The four locations where black mangrove occurs along the Texas gulf coast have been denoted: 1) Port Isabel - South Bay; 2) Harbor Island (near Port Aransas); 3) Cavallo Pass (near Port O'Connor); and 4) Galveston Island. Detail TIGER maps of the Port Isabel - South Bay area (lower left) and Harbor Island (right center) depicting GPS locations (triangles) where black mangrove populations occur in each area. A local site map (lower right) of the Harbor Island area showing greater detail of a portion of this area,

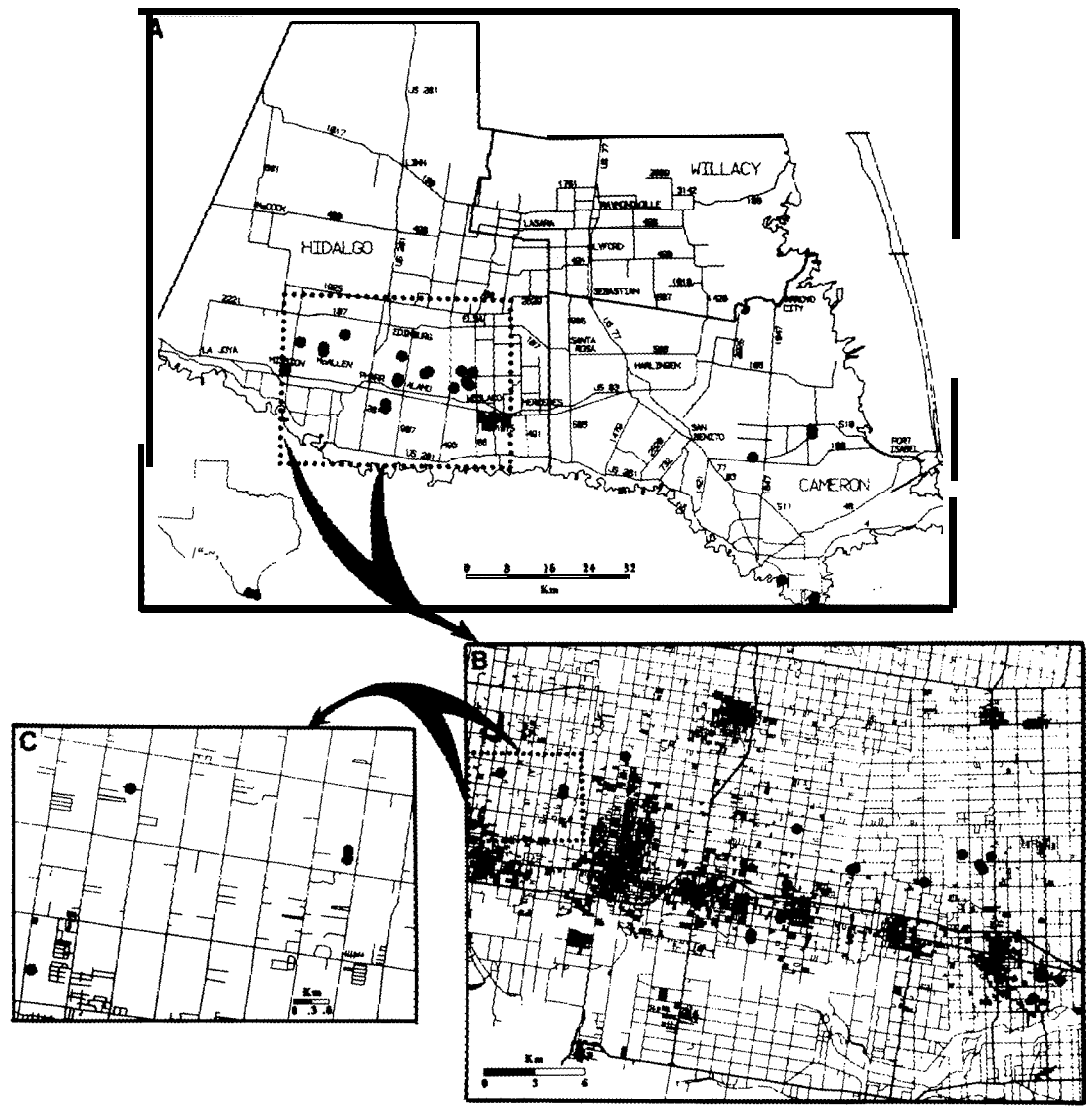


Figure 4. Regional GIS TIGER map (A) of the three-county citrus growing area of the Lower Rio Grande Valley of Texas depicting GPS locations (dots) of citrus orchards infested with citrus blackfly. The specific locations were obtained from GPS latitude-longitude data collected in aerial videographic surveys. Print B shows a more detailed TIGER map showing the locations of citrus blackfly-infested orchards in Hidalgo County. Print C shows a local site map showing even greater detail on the location of selected orchards in Hidalgo County.

was produced depicting locations of citrus orchards (dots) infested with citrus blackfly. The map also shows primary roads and the approximate locations of towns.

A more detailed ATLAS-GIS TIGER map showing the locations of blackfly-infested orchards in Hidalgo County is shown in Figure 4-B (area highlighted in Figure 4-A). With this map one can begin to associate general street/road addresses with the GPS marked location of each orchard where citrus blackfly infestations occur. A local site map showing even greater detail on the location of the selected orchards is illustrated in Figure 4-C.

HARVESTER ANT

The visible red narrowband black-and-white digital video image of a pasture near Elsa, Texas with a heavy infestation of harvester ants is shown in Fig. 5. The arrow on the print points to the conspicuous light gray to white tonal response of a harvester ant mound. The generally circular shape of the ant mounds delineates them from the sparsely vegetated areas which have a similar image response. Grassland vegetation has a gray to dark gray response. The image was obtained at an altitude of approximately 1380 m. The GPS data are shown on the bottom of the image. The harvester ant mounds could also be easily detected in the CIR composite image and could generally be delineated in the yellow-green image, but they could not be distinguished in the NIR image (other images not shown). Analysis of the video imagery of the 220 km^2 study area showed what appeared to be 44 ant infestations. Ground surveys confirmed the presence of harvester ants at every location.

The merger of the GPS latitude-longitude data obtained from the video survey of harvester ant infestations with a GIS is shown in Fig. 6. The lower portion of the figure shows GIS maps of Hidalgo and Willacy Counties with the study area delineated by a dotted boundary. The southern part of the experimental site was in Hidalgo County, whereas the northern portion was in Willacy County. The symbols (triangles) within the experimental area represented GPS latitude-longitude coordinates where harvester ant infestations occurred. Many of the locations are overlapped because of the small map scale. A more detailed GIS map of the study area is shown in the upper right portion of Fig. 6. A local site map showing even greater detail on the location of selected harvester ant infestations is shown in the upper left portion of Fig. 6. This capability of the GIS software is useful for obtaining more specific details on areas of interest.

CONCLUSIONS

The integration of videography, GPS, and GIS technologies are valuable tools for detecting and mapping insect and plant infestations/populations. These technologies can enable agricultural consultants and natural resource managers to develop maps showing the distribution of insect infestations and plant populations over large areas. The video imagery can serve as a permanent geographically located image data base to monitor future contraction or spread of insect infestations and plant populations. The GIS data base can be used to record attribute information for areas of interest. Thus, the consultant or resource manager could use the GIS to rapidly recall past video imagery for each georeferenced site and review recorded attribute information from the data base for that site. New information for each site can easily be entered into the data base as acquired. Information recorded could include climate, soil, weather, pest problems, or management actions taken by the resource manager.

ACKNOWLEDGEMENT

Thanks are extended to Angie Cardoza for typing the manuscript and Mario Alaniz for preparing the illustrations.

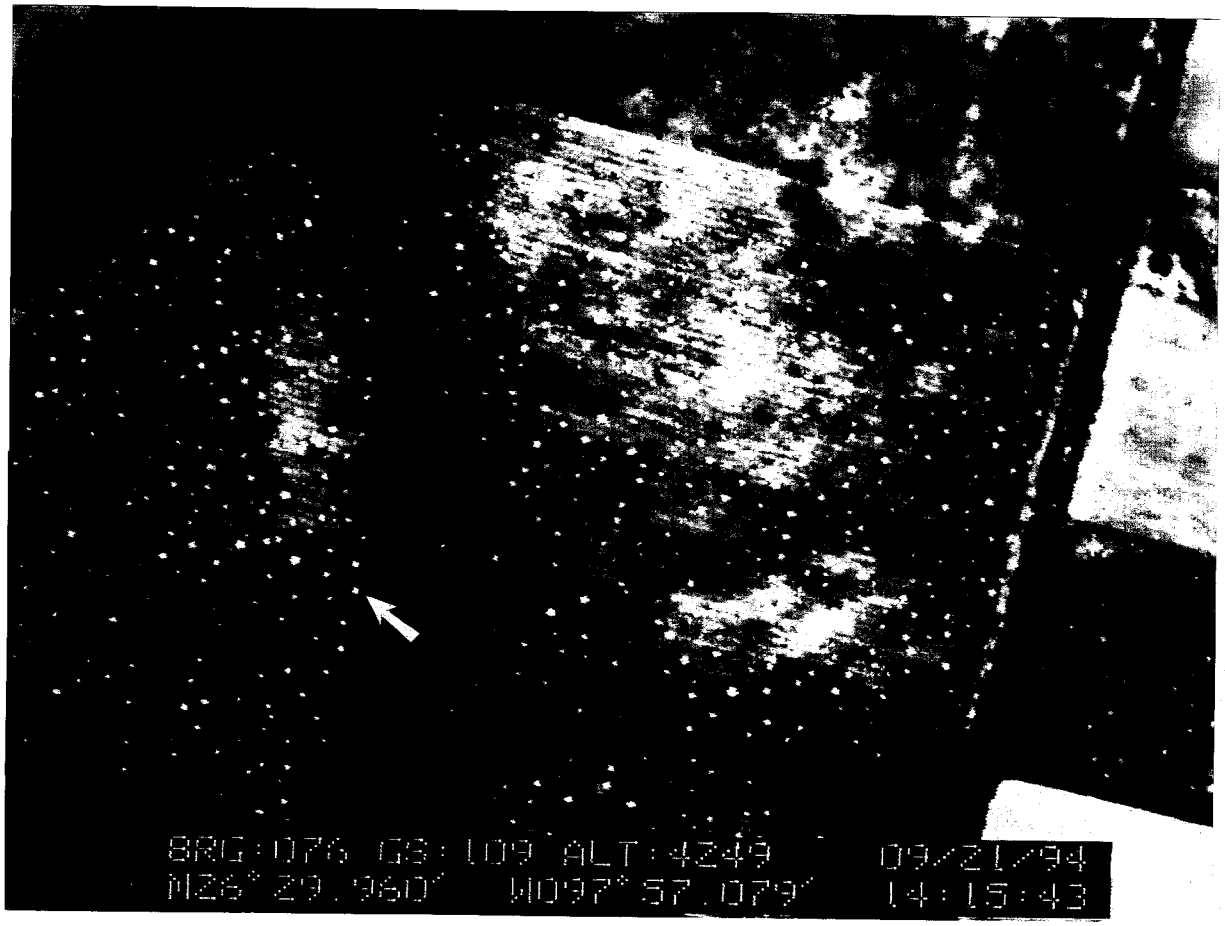


Figure 5. Black-and-white red narrowband video image of a pasture near Elsa, Texas infested with harvester ants. The arrow on the print points to the light gray-white response of a harvester ant mound. The global positioning system data shown at the bottom of the image includes the bearing (direction), ground speed, altitude, date, north latitude, west longitude, and time.

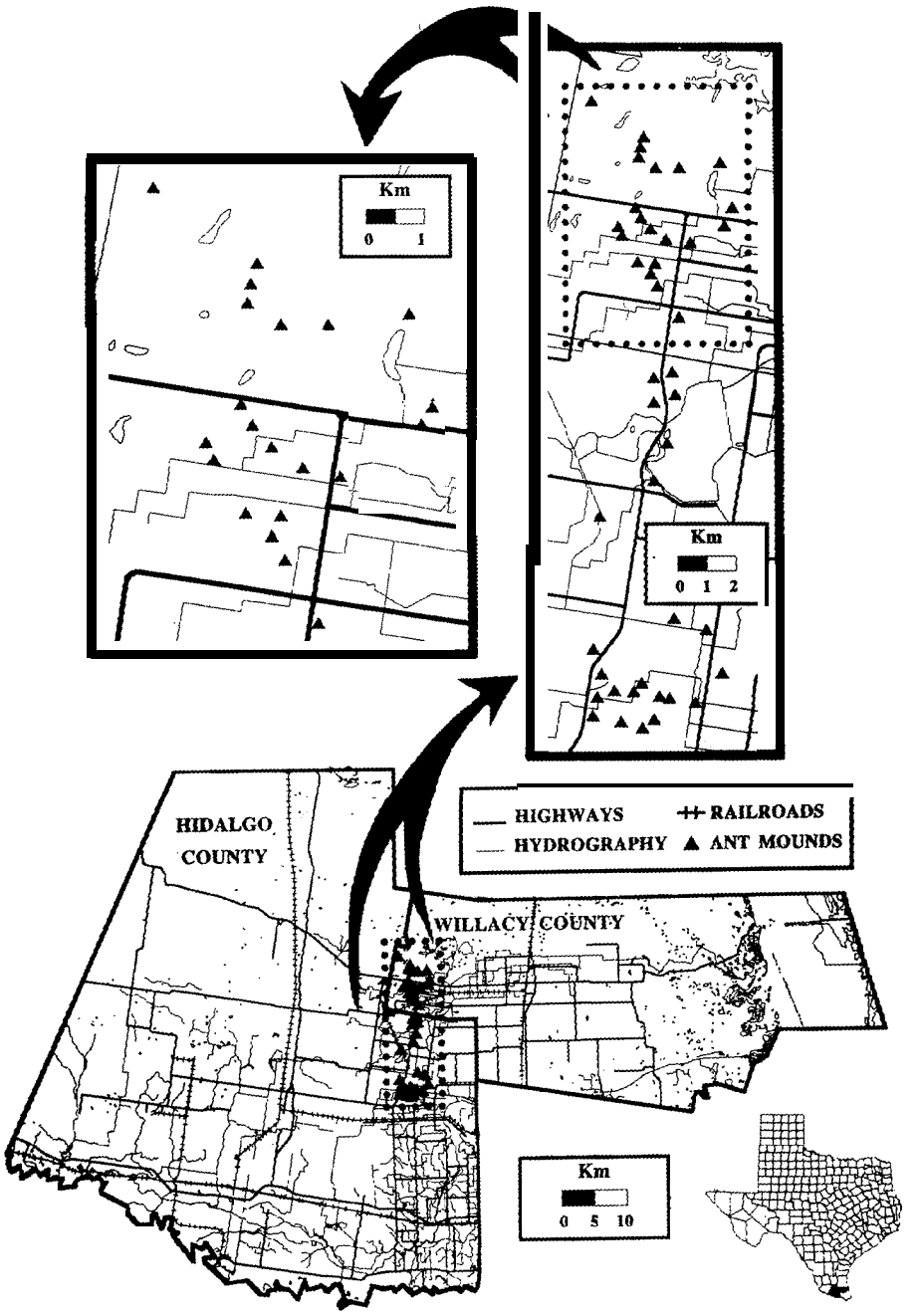


Figure 6. Geographic information system (GIS) TIGER map (lower) of Hidalgo and Willacy Counties with the harvester ant study area delineated with a dotted boundary. The symbols (triangles) within the experimental site represent global positioning system latitude-longitude coordinates where harvester ant infestations occurred. Each symbol represents a single pasture with a harvester ant infestation, A detailed GIS map (upper right) clearly depicts the locations of the harvester ant infestations within the study area. The upper left map shows even greater detail of a portion of the study area.

REFERENCES

B. R. Baum. Introduced and naturalized tamarisks in the United States and Canada. Baileya 15:19-25. 1967.

T. J. Bobbe and P. Ishikawa. Real-time differential GPS: An aerial survey-remote sensing application. Proc. Resource Technol. 92 Sympos. Amer. Sot. Photogramm. and Remote Sens. Bethesda, MD. p. 108-116. 1992.

D. S. Correll and M. C. Johnston. Manual of the vascular plants of Texas. Texas Research Foundation, Renner, Texas. 1881 p. 1970.

C. J. Deloach. Prospects for biological control of saltcedar (Tamarix spp.) in riparian habitats of the southwestern United States. Proc. VII. Int. Sympos. Biol. Contr. Weeds. March 6-11, 1988, Rome, Italy. pp. 307-314. 1990.

J. H. Everitt. Using aerial photography for detecting blackbrush (Acacia rigidula) on south Texas rangelands. J. Range Manage. 38:228-231. 1985.

J. H. Everitt and C. J. Deloach. Remote sensing of Chinese tamarisk (Tamarix chinensis) and associated vegetation. Weed Sci. 38:273-278. 1990.

J. H. Everitt, D. E. Escobar, and J. Noriega. A high resolution multispectral videosystem. Geocarto Int. 6:45-51. 1991.

J. H. Everitt, D. E. Escobar, K. R. Summy, and M. R. Davis. Using airborne video, global positioning system, and geographic information system technologies for detecting and mapping citrus blackfly infestations. Southwestern Entomology 19:129-138. 1994a.

J. H. Everitt, D. E. Escobar, R. Villarreal, M. A. Alaniz, and M. R. Davis. Integration of airborne video, global positioning system and geographic information system technologies for detecting and mapping two woody legumes on rangelands. Weed Tech. 7:981-987. 1993.

J. H. Everitt, D. E. Escobar, M. A. Alaniz, M. R. Davis, and J. V. Richerson. Using spatial information technologies to map chinese tamarisk (Tamarix chinensis) infestations. Weed Sci. 44:(In press). 1996a.

J. H. Everitt and F. W. Judd. Using remote sensing techniques to distinguish and monitor black mangrove (Avicennia germinans). Journal of Coastal Research 5:737-745. 1989.

J. H. Everitt, F. W. Judd, D. E. Escobar, and M. R. Davis. Integration of remote sensing and spatial information technologies for mapping black mangrove on the Texas gulf coast. J. Coastal Research 11 :(In press). 1996b.

J. H. Everitt, J. V. Richerson, M. A. Alaniz, D. E. Escobar, R. Villarreal, and M. R. Davis. Light reflectance and remote sensing of Big Bend loco (Astragalus mollissimus var. earlei) and wooton loco (Astragalus wootonii). Weed Sci. 42:115-122. 1994b.

W. G. Hart. Some biological control successes in the southern United States. Proc. Int. Citricult. 3:154-156. 1978.

W. G. Hart, H. W. Gausman, and R. R. Rodriguez. Citrus blackfly (Homoptera: Aleyrodidae) feeding injury and its influences on the spectral properties of citrus foliage. J. Rio Grande Valley Hort. Sot. 30:37-43. 1976.

W. G. Hart, S. J. Lingle, M. R. Davis, and C. Mangum. Aerial photography with infrared film as a method of surveying for citrus blackfly. J. Econ. Entomol. 66:190-194. 1973.

J. S. Horton and C. J. Campbell. Management of phreatophyte and riparian vegetation for maximum multiple use values. USDA For. Serv. Paper RM-1 17. 23 p. 1974.

T. A. Kerpez and N. S. Smith. Saltcedar control for wildlife habitat improvements in the southwestern United

States. USDI, U. S. Fish and Wildl. Serv. Resour. Publ. 169.16 p. 1989.

D. King and J. Vlcek. Development of a multispectral video system and its application to forestry. Canadian J. Remote Sens. 16:15-22. 1990.

R. I. Lonard and F. W. Judd. Effects of a severe freeze on native woody plants in the lower Rio Grande Valley, Texas. Southwestern Nat. 30:397-403. 1985.

R. I. Lonard and F. W. Judd. Comparison of the effects of the severe freezes of 1983 and 1989 on native woody plants in the lower Rio Grande Valley, Texas. Southwestern Nat. 36:213-217. 1991.

D. E. Meisner and O. M. Lindstrom. Design and operation of a color-infrared aerial video system. Photogramm. Eng. and Remote Sens. 51:555-560. 1985.

R. J. Myhre. Use of color airborne videography in the U. S. Forest Service. Proc. Resource Technol. 92 Symp., Am. Sot. Photogramm. and Remote Sens. Bethesda, MD. p. 145-152. 1992.

P. R. Nixon, D. E. Escobar, and R. L. Bowen. A multi-video false color imaging system for remote sensing applications. Proc. 11th Biennial Workshop on Color Aerial Photography and Videography in the Plant Sciences. Am. Sot. Photogramm. and Remote Sens. Bethesda, MD. pp. 295-305, 340. 1987.

A. J. Richardson, K. R. Summy, M. R. Davis, A. Gomez, and D. W. Williams. The use of 1990 Tiger/Line™ census files for monitoring the Rio Grande Valley cotton stalk destruction program. Proc. Application Advanced Information Technol. Sympos. pp. 231-239. 1993.

C. L. Sherrod and C. McMillan. Black mangrove, Avicennia germinans, in Texas: past and present distribution. Contributions in Marine Science 24:115-131. 1981.

K. R. Summy, F. E. Gilstrap, W. G. Hart, J. M. Caballero, and I. Saenz. Biological control of citrus blackfly (Homoptera: Aleyrodidae) in Texas. Environ. Entomol. 12:782-786. 1983.

K. R. Summy and E. G. King. Cultural control of cotton pests in the United States. Crop Protection 11:307-319. 1992.

A Novel Approach to Decision Making

98URC130

Jalal Baghdadchi
NASA Autonomous Control Engineering Center
North Carolina A&T State University
Email: bagh@ncat.edu

Abdollah Homaifar
NASA Autonomous Control Engineering Center
North Carolina A&T State University
Email: homaifar@ncat.edu

Abstract

Decision-making is the core of any control system. The driving metaphor for a traditional decision-maker is the existence of an explicit memory base and its ability to analyze and express any situation that the controlled plant experiences, by a set of mutually disjoint rules. The disjoint nature of the rules reflects the fact that in engineering world we tend to partition the problems to smaller tasks, deal with each individual task separately, and then reassemble the overall solution. Given that the majority of problems and situations in real-time do not lend themselves easily to this type of partitioning, we are proposing a decision making scheme which confronts the problems in their entirety, thus, functioning to a great extent like the human brain.

1. INTRODUCTION

The level of sophistication required from intelligent agents necessitate the existence of an on-board intelligence capability which emulates the information handling characteristics of the human brain. In the engineering world, when encountering a problem, we normally tend to try to partition the problem into smaller pieces, deal with individual pieces, and then put the whole thing back together. This strategy is referred to as the analytical method (Iran-Nejad et al., 1992). Since all real life issues do not lend themselves easily to this type of segmentation, this is not necessarily the best way of approaching a problem. Often, when dealing with individual pieces of a problem, we forget the "glue" which holds these pieces together.

Suppose we want to teach someone how to ride a bike. We can give the individual some insight about the pieces of a stationary bike such as the handle bar, pedals, braking mechanism, etc. Of course chances of the individual becoming able to ride a bike are not very good. Obviously, unless there is motion the knowledge about the pieces of the bike are not going to be, helpful in getting the individual to acquire the ability to ride a bike. On the other hand, if we give the potential bike rider a bike and have him/her just get on it and try to go forward by pedaling, chances of success will be considerably higher (Iran-Nejad et al., 1995). The first approach is the analytical method, and the second is a whole theme approach. In whole theme approach we try to approach the problem as a single item as opposed to analytical method, where we try to divide the problem to pieces.

The final objective is to construct a decision-making model capable of operating accurately in a partially known environment. Such a model will inevitably have to emulate the functioning of the human brain. We will proceed to show that the human brain tends to employ the whole theme approach rather than the analytical in performing instances of learning, remembering. This is particularly true when a decision is to be made in absence of full scope of supporting knowledge. The decision making model will employ a set of fuzzy rules which will govern its operation in a dynamic environment. We will first survey the existing cognitive theories of knowledge, then introduce our suggested decision making theory and model, followed by a brief comparison of main features.

The main activities of a real-time decision-maker are sensing, storing and retrieving of information. The equivalent terms for these activities in the field of psychology (cognitive theories of knowledge) are cognition, learning and remembering. Notice that the equivalency is approximate. We consider three major families of theories attempting to describe the functioning of the human brain: associationist theories, connectionist theories and the biofunctional model. The first two rely on explicit memory as the knowledge base, where as the third rests upon the whole theme approach. Information processing theory (IPT)(Atkinson, Shiffrin, 1968), an example of associationist theories, considers the human brain to be a predominantly passive storage/retrieval structure. The explicit memory structure consists of sensory registers (SR) and short and long term memories (STM, LTM). IPT accounts for intentional learning only (Iran-Nejad, Homaifar, 1992.) Adaptive resonance theory (ART)(Grossberg, 1995), an example of connectionist theories, functions based on perception matching. It executes learning both serially and in parallel. It offers an improvement over the former by not limiting learning to only intentional occasions. Inhibition, one of the refining mechanisms of ART, however, limits its effectiveness.

3. BIOFUNCTIONAL MODEL

Before we explore the biofunctional model, let's consider an example. Assume you wish to pick up an object using your hands. If you are right handed, like I am, your right arm will move towards the object in an attempt to pick it up. The neural network model suggests that there is a nerve connection between the brain and your right hand (a dedicated command line.) The command to pick up the object travels through this nerve from brain to the right hand.

Biological experiments however, reveal that the response of a certain organ to a command does not depend on the existence of a physical connection between the organ and the command-issuing center (brain or spine). The experiment with the trained salamander clarifies the point. In this experiment, a salamander was trained to push a little door to let loose his food. The salamander normally used his right front leg (hand) to push the door open. When the trainer tied salamander's right front leg to his body, he used the left hand to push the door open (Iran-Nejad et al., 1992). When both front legs were tied, the salamander resorted to using his head in an attempt to push the door open. Had the animal's reaction to command been dependent on the existence of a physical connection from brain to his right hand, he would not have attempted to execute the routine using alternative organs.

The example above suggests that the commands, news and information travel throughout the system and that biological activities (such as a certain organ performing a task) do not need a physical connection to/from the brain in order to receive the command. In other words, the awareness is universal. Once an event takes place, the news is available throughout the system. A learning model accommodating this important feature of biological entities is closer to real life and suits our purpose better.

Biofunctional Model: The core of the biofunctional model is the existence of two types of brain activities. Learning is achieved as a result of the interaction between these two activities. The two activities are on-going-brain-activity (OBA), and momentary constellation firing (MCF). OBA is an unbounded, ever present, ever evolving and ever rearranging contiguous body of knowledge. MCFs are short lasting appearance (detection of or reaction to) events.

A learning system includes sources of control, fictionally autonomous subsystems (organs), on-going-brain-activity (OBA), and the capability of generating MCFs. The subsystems can be thought of as bodily organs. OBA is the main (and only) body of knowledge. Note that OBA is not static. MCFs on the other hand, are momentary phenomena. At any given instant of time the system may or may not experience the appearance of any MCFs. We will further elaborate on OBA and MCFs through the following analogy: Consider a highway traffic arrow (of the kind, which are usually used while

construction goes on). The arrow is composed of many light bulbs. If all the light bulbs are on or off, we will see the shape of an arrow. If light bulbs go on and off with a certain pattern, we will experience the illusion that the arrow is pointing to a certain direction. The original arrangements of the light bulbs (independent of their on / off status) is analogous to OBA. The turning on and off of each individual light bulb can be thought of as MCFs. The combination of the background provided by light bulbs, and their going on and off result in the learning experience (the arrow pointing to a direction).

OBA and MCF in action; Two examples:

Consider yourself as the learner and someone who you know and regularly interact with as the subject to learn about. Let's define the summary of all your information about this individual, that is your opinion, as the OBA. Two months from now you may have a different opinion about the individual. The new experiences and interactions that you have had with the individual will then cause the change in the previous opinion. These experiences are the MCFs. The relevant experiences (MCFs) will change the composure of the previous OBA (rearrange it).

Again consider yourself as the learner. Suppose you are watching a mystery movie on TV. Your OBA at any given time, is what you have gathered from the story up to that point in time. As you are watching, many scenes are presented to you; these are the MCFs. These scenes have different levels of importance and apparent relevance to the main theme of story. Some of these scenes are judged important and leave a sharply outstanding trace on the OBA. The trace left by others may not be as vivid. As you continue to watch, your understanding of the story gets updated. The updated OBA is the result of interaction between OBA and MCFs.

Notice that OBA remains contiguous at all times. The relevant MCFs get integrated into the OBA. Integration of MCFs into OBA account for the "ever changing, ever evolving, ever rearranging,.." characteristic of the OBA. Also, the model has the capability to go back in time (dig into past experiences). Consider the case of watching mystery on TV. Often, as the story unfolds and the mystery is about to be solved, suddenly one or some of the past scenes (MCFs) which were not treated as important, get revived and play a significant role in the making of the conclusion and formation of the final experience. Thus, the OBA is seen to have the capability to mold old and new knowledge and come up with a more accurate account of the events.

Learning: Learning is defined as any change to OBA. Types of activities that the learning system (human brain) engages in from the time of first attention until the time of final conclusion in a learning episode are attention, inquiry, closure, combination, and knowledge creation. The learner pays attention to a certain process when, due to either internal activation or external excitation the system desires to learn about the process. Attention leads to inquiry; the system either mindfully or subconsciously becomes inquisitive about the process in question. Inquisitiveness leads to new data becoming available to the system (closure). This can be the result of external ly triggered MCFs as well as OBA trying to revive the past knowledge. Once all types of data (OBA digging into the past, externally triggered MCFs, etc.) from different sources (different subsystems) about a certain event are available to the learner, the learner tries to make sense of all data (combination), If the process leads to or includes an experience worthy of remembering, it gets integrated into the OBA (knowledge creation).

4. COMPARISON

The comparison of some of the features of Biofunctional model and the associationist and connect ionist theories is shown in the table 1. The comparison reveals that the characteristics and the functioning of the Biofunctional model closely resemble those of the human brain.

	Biofunctional	Associative and Connectionist
Paradigm	whole theme	piecemeal
Types of Learning	allowing both intentional and Incidental learning	intentional learning only (IPT)
Learning Mechanism	double activity hypothesis (OBA and MCFs)	assembly line approach (three sequential phases) for IPT, single pattern processing for neural networks
Processing Capacity	the capacity of human brain	subject to limitations of explicit memory
Information Routing	information is available throughout the system	dependent on the existence of physical connection between brain and the organ

5. BIOFUNCTIONALITY IN DETAIL

5.1. Biofunctionality and Remembering:

The following analogy and examples demonstrate that the biofunctional model describes the human experience of remembering very closely. Recall OBA; "... an unbounded, ever present, **contiguous body of knowledge".** This can be thought of as a rough surface with bumps, peaks and valleys just like one of those computer-generated terrain maps. Every feature (bump, peak or valley) of this surface corresponds to a certain experience (knowledge), of a network of interrelated (not entirely separable) experiences. Suppose this surface is made of rubber sheet. Now, consider this rough surfaced rubber sheet in a three dimensional Cartesian space. The vertical axis signifies the vividness of the memory. At the bottom of valleys, the awareness about a certain experience is almost non-existent (oblivion region). On top of the high peaks, the experience is vividly present in the memory (full-awareness region), Thus, remembering a certain event will be like pulling a valley or a low peak from oblivion or low awareness up to the full-awareness region. The higher the point comes up, more vividly it is remembered.

Remembering is not exact: Often, when trying to remember an old tune (a piece of music that you used to listen to some years ago), in addition to the main subject of recall (the piece of music), other memories of those days will come to life. You will probably remember the set up of your living quarters, taste of foods that you used to eat and other details. The surrounding items and details may not be recalled, as vividly, nevertheless, most likely more memories will be revived than you originally intended to recall. The rubber sheet remembering analogy can explain this phenomenon. When you target a point on the sheet and pull it up, a cone like segment of the sheet will come up with it. The apex (main focus of attention) attains the highest degree of vividness. The surrounding material will become livelier than before but not as vivid as the main item to be remembered.

Repeated Remembering: When you try to remember an event which took place a couple of years ago that you have not thought of ever since, you might have to make some effort to remember it. However, if you make a second attempt a few days after the first, the second time around the recalling process will require much less effort. The rubber sheet remembering analogy can explain this, Forgetting an experience is equivalent to sinking down of its representative feature (peak, bump,...) to the oblivion region. This is like resetting of a mechanical system, which is usually a function of time. Rubber sheet too, can be considered to be slow in returning to its position before being disturbed. At the time of the

second attempt, the newly risen peak has not totally retreated to its pre-remembering position. The longer the time lapse between the two remembering instances, the greater the effort for second remembering. This is of course in contrast with the explicit memory models. If you pick a book which has been sitting in a I ibrary shelf for the past few years, put it back and pick it up again the next day, the effort needed to withdraw the book, will be same in both cases.

5.2. **Sources of Regulation:**

Learning and remembering in the biofunctional model are regulated by internal and external sources of control. The biofunctional model accommodates both intentional and incidental learning. To account for these two types of learning we define two internal sources of regulation: active and dynamic. The active (executive) type of control regulates the intentional learning, while the dynamic (non-executive) control regulates the unintentional activities. An example of active control is the case when we deliberately initiate an action. A wide range of biological functions, such as heart beat, on the other hand are dynamically regulated.

Let us review the example of section 3. Consider yourself as the learner and your impressions of someone who you regularly interact with as the OBA. Any time that you notice the individual in question, or come in contact with him, your attention gets activated. If the individual utters some words towards you, naturally you will notice and make an attempt to hear him out. This accounts for the inquiry sub-function. The receiving of the message is the closure. Through the message the individual may alter your impression of him, remind you of a promise (cause you to remember and compare), or give you some information. These activities will all take place at the combination stage. If a conclusion is drawn or a piece of information worthy of future remembering is received, it will be integrated into knowledge base during the knowledge creation.

6. **CONCLUSIONS AND FUTURE WORK**

The comparison between the reviewed theories and biofunctionality reveals that the biofunctional model accommodates the characteristics and requirements of our originally intended learning model. These requirements were contiguous body of knowledge, universal information exchange, connection-independent execution environment, and the ability to perform in a dynamic environment with unknown parameters. The model is particularly suitable to dealing with complex situations with inter-dependent players, where the traditional single thread analysis may not yield the best result. We believe that the biofunctional model can successfully and accurately model the decision-making aspects of a multi-variable operational environment. We intend to employ the refined biofunctional learning and remembering model to a multi-robot task scheduling and task-performing scheme.

7. **ACKNOWLEDGMENT**

This work is partially funded by a grant from NASA Autonomous Control Engineering Center ACE, North Carolina A&T State University branch under grant number 48146. The authors wish to thank them for their financial support

8. REFERENCES:

1. Atkinson, R. C., Shiffrin, R. M., “Human memory: A proposed system and its control processes”, The Psychology of learning and motivation, 1968, Vol. 2, pp. 90-97.

2. Baghdadchi, Jalal, Homaifar, Abdollah, “A Biofunctional Approach to Decision Making”, NASA URC Technical Conference, Feb. 1997, pp. 87-92.

3. Campbell, J., “The Improbable machine”, New York, Simon& Schuster, 1989.

4. Grossberg, Stephen, ''The Attentive Brain", American Scientist, 1995, Volume 83, pp. 438-449.

5. Grossberg, Stephen, Carpenter, A. Gail, "Integrating and Neural Processing in a Self-Organizing Architecture", Artificial [intelligence and Neural Networks, 1994, pp. 387-421,

6. Iran-Nejad, Ali, Marsh, E. George, "The Figure and the Ground of Constructive Brain Functioning: Beyond Explicit Memory Processes", Educational Psychologist, 1992, 7(4), pp. 473-492.

7. Iran-Nejad, Ali, Homaifar, Abdollah, "Associative and Non-associative Theories of Distributed Learning and Remembering", S. J. Schmidt (Ed.), Frankfurt, Germany, 1992.

8. Iran-Nejad, Ali, "Associative and Nonassociative Schema Theories of Learning", Bulletin of the Psychonomic Society, 1989, 27(1), pp. 1-4.

9. Konig, Harold, Litz, Lothar, "Inconsistency Detection-a Powerful Means for the Design of MIMO Fuzzy Controllers", IEEE 1996,0-7803-3645, pp. 1191-1197.

Multiobjective Evolutionary Path Planning via Sugeno-Based Tournament Selection

Gerry Dozier*† Shaun McCullough Abdollah Homaifar†
Albert Esterline†

Abstract

This paper introduces a new tournament selection algorithm that can be used for evolutionary path planning systems. The fuzzy (Sugeno) tournament selection algorithm (STSA) described in this paper selects candidate pat hs(CPs) to bc parents and undergo reproduction based on: (1) path feasibility, (2) the euclidean distance of a path from the origin to its destination, and (3) the average change in the slopc of a path.

In this paper, wc provide a detailed description of the fuzzy inference system used in the STSA as well as some examples of its usefulness. We then use 12 instances of our STSA to rank a population of' CPs based on the above criteria. We also show how the STSA can obviate the need for the development of an explicit (lexicographic multiobjective) evaluation function and use it to develop multiobjective motion paths.

KEYWORDS: Sugeno Tournament Selection, Evolutionary Path Planning, Fuzzy Inference

1 Introduction

Recently, there has been a growing number of successful applications of Evolutionary Path Planners [1, 2,3,.5,7',8, 9]. However, many of these systems are concerned with finding the shortest path between a starting point and a destination for a robot to traverse[1]. In path planning, the shortest path may not always be the most efficient means of getting from start to destination. There are many other attributes of a path that may be desirable other than distance. one example would be the smoothness of a path. The development a multiobjective closed-form fitness equation for evolutionary path planning may be extremely difficult. Therefore, our concentration has been placed on the development of a multiobjective selection method based on Fuzzy Inference.

In this paper, we introduce a new tournament selection algorithm, called fuzzy (Sugeno) tournament selection (STS), that can be used for evolutionary path planning systems. The STS algorithm (STSA) described in this paper selects candidate paths (CPs) to be parents and undergo reproduction based on: (1) the euclidean distance of a path from the origin to its destination, (2) the sum of the changes in the slope of a path, and (3) the average change in the slope of a path. The remainder of this paper is as follows. In Section 2, we provide a brief description of tournament selection, and in Section 3, we discuss our new STSA. In Section 4, we present our experiment, and in Section 5, we present our results and conclusions.

2 Tournament Selection

Tournament selection [6] is typically the selection method of choice. In tournament selection, N individuals are randomly chosen from the current population and placed in a tournament (usually $N = 2$). The individual with the best 1 fitness in the IV-tournament is selected to be a parent.

Even though tournament selection is the method of choice it still cannot be adequately applied to optimization problems with multiple objectives. To this end, we have combined the concept of tournament selection with fuzzy logic to form a STSA for multiobjective evolutionary path planning.

*Department of Computer Science & Engineering, Auburn University, Dunstan Hall, Auburn University, Auburn, AL 36849, gvdozier@eng.auburn.edu

†The NASA Center for Autonomous Control Engineering, North Carolina A&T State University, Greensboro, NC 27411, {shaunm,homaifar,esterlin}@ncat.edu

[1] In [8], the authors use a fuzzy evaluation function for assigning fitnesses to paths based on time, distance, and work load

3 Fuzzy **Tournament Selection**

The overall objective of our STSA is to allow evolutionary path planners to evolve CPs that have: (1) relatively small distances from start to destination, (2) small sums of the changes in slope (SCS), and (3) small average changes in slope (ACS). Our STSA is actually a zero-order Sugeno Fuzzy Inference System [4]. It takes six inputs: the distances, the SCS , and the ACS of two CPs (CP_1 and CP_2) that are randomly chosen from the current population and returns one output which ranges between [-1.. 1]. This output corresponds to which CP should be selected to be a parent. Thus, any output that is less than zero means that CP_1 is selected while any output greater than zero means that CP_1 will be selected.

In what follows, we describe four attributes of our STSA: (1) the fuzzification method used to convert the six inputs into 3 fuzzy singletons, (2) the set of membership functions for each of the 3 fuzzy singletons, (3) the fuzzy rulebase, and (4) defuzzification.

3.1 Fuzzification

Each CP chosen randomly from the current population has three attributes: (1) the euclidean distance of the CP from start to destination, (2) the total sum of the changes of slope along the CP (SCS), and (3) the average change of slope along the CP (ACS). For example, consider two CPs. Let d_1 = 24.80, s_1 = 17.96, al = 1.80, d_2 = 25.37, s_2 = 11.46, and a_2 = 1.91, represent the distance, SCS, and ACS of CP_1 and CP_2. Three fuzzy singletons, *d, s,* and a, whose values are in [-1,1] can be created as follows:

$$d = \frac{24.80-25.37}{24.80+25.37} = -0.01,$$

$$s = \frac{s_1-s_2}{s_1+s_2} = \frac{17.96-11.46}{17.96+11.46} = 0.227 \text{ and}$$

$$a = \frac{a_1-a_2}{a_1+a_2} = \frac{1.80-1.91}{1.80+1.91} = -0.03.$$

Notice that for *d, s,* and *a,* if their values are less that zero, then CP_1 has the smaller attribute and vice-versa for CP_2. Therefore by closer inspection of *d, s,* and *a* one can see that CP_1 **is** slightly shorter distance and ACS than CP_2 while CP_2 has a smaller SCS than CP_1.

3.2 **Membership Functions**

Each of the fuzzy inputs *d, s, a,* has a domain of three membership functions. Figure 1 shows the three membership functions for *d*. They are as follows. If the values of *d* is non-positive then it is a member of set DP 1 with degree $DP1(d)$. Members of DP 1 represent the set of all tuples (CP_x, CP_y) such that $d_x < d_y$. That is CP_x is shorter than CP_y. Similarly, values of *d* that are non-negative are members of set DP2 with degree $DP2(d)$ and represent the set of tuples (CP_x, CP_y) such that $d_x > d_y$.

All values of $|d| < D$ are members of set DSM with degree $SM(D, d) = 1 - \frac{d}{D}$ if $D > d \geq 0.0$ and $SM(D, d) = 1 + \frac{d}{D}$ if $-D < d \leq 0.0$. Members of SM represent the set of all tuples (CP_x, CP_y) such that $d_x \approx d_y$. That is based on *D* CP_x is approximately the same length as CP_y. By changing *D* the STS has the ability to adapt its focus optimizing distance.

The membership function of s and a, shown in Figures 1 are similar to the membership functions of *d* with the exception that they correspond to the comparative SCS and ACS attributes of two CPs.

For our example, let D = S = A = 0.15. Therefore,

d = –0.01 has membership degree 0.01 in DP1, degree 0.93 in SM, and degree 0.0 in DP2.

s = 0.22 has membership degree 0.00 in SP1, degree 0.00 in SM, and degree 0.22 in SP2.

a = –0.03 has membership degree 0.03 in AP1, degree 0.80 in SM, and degree 0.0 in AP2.

3.3 The Fuzzy Rulebase and Inference Used

The fuzzy rulebase is as follows:

```
Rule 1:   If d is DP1                                  Then P = -1
Rule 2:   If d is SM(D) and s is SP1                   Then P = -i
Rule 3:   If d is SM(D) and s is SM(S) and a is API    Then P = -1
Rule 4:   If d is SM(D) and s is SM(S) and a is SM(A)  Then P = O
Rule 5:   If d is SM(D) and s is SM(S) and a is AP2    Then P = 1
Rule 6:   If d is SM(D) and s is SP2                   Then P = 1
Rule 7:   If d is DP2                                  Then P = 1
```

The Max-Product operator [4] was used to calculate the fire strength of the antecedents. Actually our rulebase consists of 3 rules. Rule 1, Rule 2, and Rule 3 can be view as one rule with three 'ORed' antecedents. Rule 4 could be viewed as a second rule and Rules 5-7 can be viewed as the third and final rule. For each of the seven rules, *P* represents the consequence of a rule. If the consequence of a rule is $P = -1$, then the rule has specified that CP_1 should be selected to be a parent. Similarly if a rules consequence is $P = 1$ then it has specified that CP_2 should be selected.

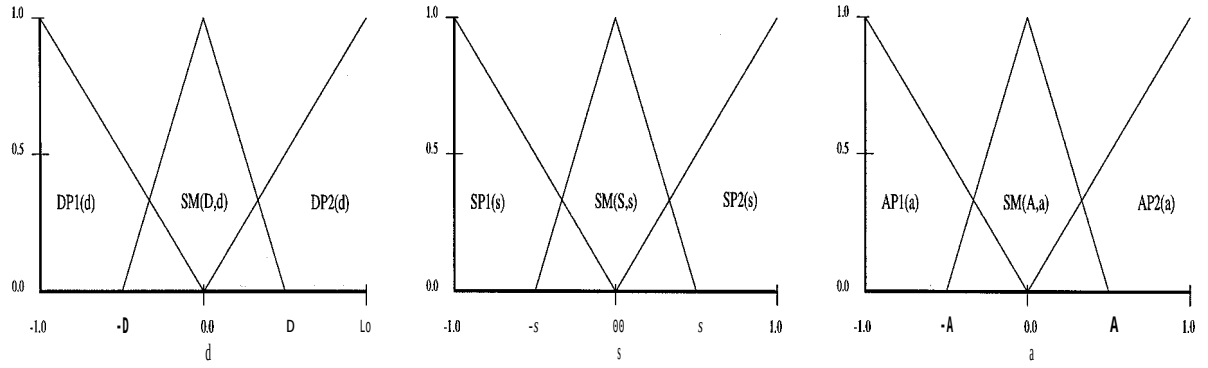


Figure 1: The Membership Functions of d, s, and a

In continuation our example, we can compute the fire strength of each of the antecedents. Let f_x denote the fire strength of Rule x. The fire strengths are as follows (where d = –0.01, s = 0.22, a = –().()3, and $D = S$ = A = 0.15):

$f_1 = DP1(-0.01) = 0.01,$

$f_2 = SM(D, -0.01) * SP1(0.22) = 0.93 * 0.00,$

$f_3 = SM(D, -0.01) * SM(S, 0.22) * AP1(-0.03) = 0.93* 0.00* 0.00,$

$f_4 = SM(D, -0.01) * SM(S, 0.22) * SM(A, -0.03) = 0.93* 0.00* 0.80,$

$f_5 = SM(D, -0.01) * SM(S, 0.22) * AP2(-0.03) = 0.93 * 0.00 * 0.00,$

$f_6 = SM(D, -0.01) * SP2(S, 0.22) = 0.93* 0.22,$ and

$f_7 = DP2(-0.01) = 0.00$

At this point, let $F_1 = max(f_1, f_2, f_3) = 0.01$, $F_2 = f_4 = 0.00,$ and $F_3 = max(f_5, f_6, f_7) = 0.20.$

3.4 Defuzzification

The defuzzification technique used is the mean of the maximums. Let P_x represent the consequence for F_x. Therefore PI = – 1, $P_2 = 0.0$ and P_3 = 1.0. Our defuzzification function is $P = \frac{\Sigma_i F_i * P_i}{\Sigma_i F_i} =$ **0.90,** where P represents the output of the our zero-order Sugeno fuzzy inference system.

4 The Experiment

A selection algorithm should be able to used as a mechanism for ranking members of a population. To test our STSA we have chosen seven paths shown in Figure 2. The paths all have the same start and destination point. These paths could have been created by any of the aforementioned evolutionary planning systems. The attributes for each of the seven paths are shown in Figure 3. Notice that the paths are sorted based on their distances with path1 being the shortest and path7 being the longest. Having the paths arranged in this manner will help see how the STSA goes about selection based on distance, SCS, and ACS.

In our experiment we allow each of the seven paths to compete in a tournament (tournament size = 2) with every other path. The STSA was used select the better of the two paths. When a path wins a tournament, a counter corresponding to the winning path is increment. When all of the tournaments are finished (a total of 21 tournaments in all), one can see how good a path is relative to the others.

We performed the above experiment twelve times; one for each of the values of D taken from the set {0.0, 0.025,0.05,0.1,0.15, 0.2,0.25,0.3,0.35, 0.4,0.45,0.5}. We kept S and A constant at 0.15.

5 Results

The results of our experiment are shown in Figure 4. In Figure 4, *D as* stated in Section 3 can be viewed a similarity threshold of the SM membership function for an attribute. Notice that when $D = 0.0$ the ranking of the paths is based purely on distance. This correspond to how a conventional tournament selection algorithm would assign wins. As we increase the range of similarity, the longer but smoother paths are increasingly preferred over the shorter paths. This is indicated by the fact that the longer path receive an more wins as D is increase to 0.35.

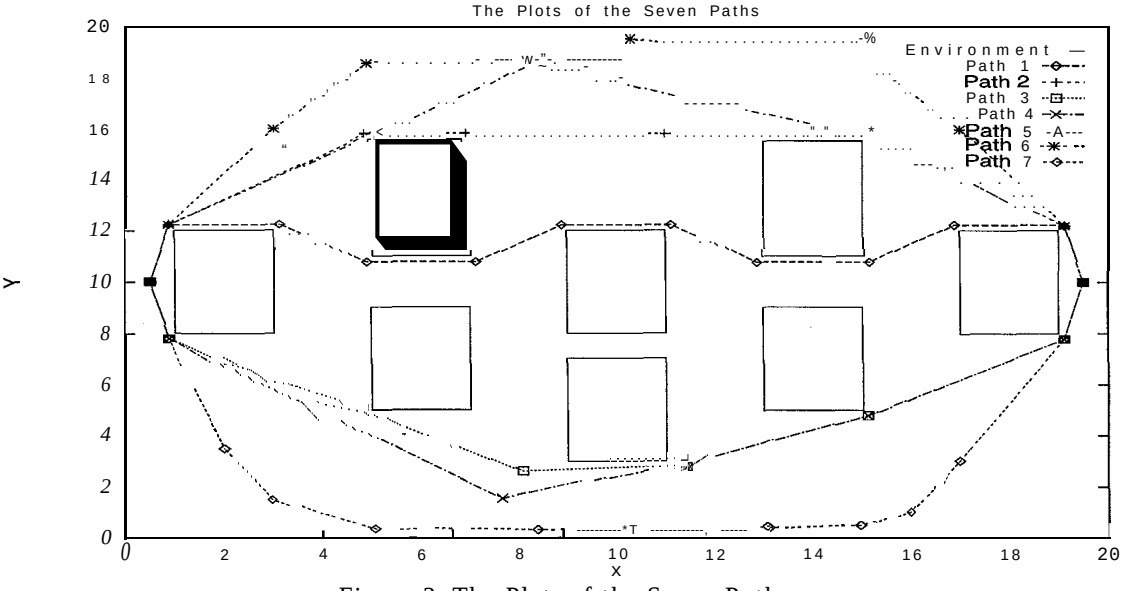


Figure 2: The Plots of the Seven Paths

Path	Distance	*SCS*	*ACS*
Path 1	24.79	17.96	1.79
Path 2	25.36	11.45	1.90
Path 3	25.90	11.46	2.29
Path 4	26.84	11.46	2.29
Path 5	26.92	11.45	2.86
Path 6	30.03	11.45	1.43
Path 7	32.00	11.46	1.04

Figure 3: The Attributes of the Seven Paths

A surprising and unexpected result can be seen upon closer observation of Figure 4! Notice that number of wins assigned to Path 1 and Path 7 decreases/increases monotonically as *D* is increased. However, this is not the case for Paths 2-6. This result is both shocking and encouraging! It seems that even though our STSA is simple and is composed of three rules it still exhibits some complex behavior. This behavior we see as result of changing the values of only one of three parameters.

Another important result is that our STSA (because it can reliably rank paths based on multiple objectives) obviates the need to develop a multiobjective evaluation function. In Section 3, one may have thought that Rule 4 was extraneous. Actually this is not the case at all! Rather than having Rule 4's consequence tied to 0.0 one could instead use the consequence that would signal for the values of *D*, *S*, and *A* to have their values modified. In other words, Rule 4 can be used to tell the convergence of the population. If Rule 4 fires regularly then this is a sign the the population is converging. By modifying the values of *D*, *S*, and *A*, the STSA causes the evolutionary path planner to place greater emphasis on other attributes. How *D*, *S*, and *A* are effective modified is a topic for future research.

Path	***D* = 0.00**	**0.025**	*0.05*	**0.1**	**0.15**	*0.2*	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**
Path 1	**6**	**4**	*2*	**2**	**2**	**1**	**1**	**0**	**0**	**0**	**0**	**0**
Path 2	**5**	**6**	*6*	**6**	**6**	*6*	**5**	**5**	**4**	**4**	**4**	**4**
Path 3	**4**	**5**	*5*	**5**	**4**	*3*	**3**	**3**	**3**	**3**	**3**	**3**
Path 4	**3**	**3**	*4*	**3**	**2**	*2*	**2**	**2**	**2**	**2**	**2**	**2**
Path 5	**2**	**2**	*3*	**2**	**1**	*1*	**1**	**1**	**1**	**1**	**1**	**1**
Path 6	**1**	**1**	*0*	**2**	**3**	*4*	**4**	**4**	**5**	**5**	**5**	**5**
Path 7	**0**	**0**	*1*	**1**	**3**	*4*	**5**	**6**	**6**	**6**	**6**	**6**

Figure 4: Path Rankings vs. *D*, with S = 0.15 and *A* = 0.15

Acknowledgements

The authors wish to thank the NASA ACE Center (under grant ACE-48146) and NASA Dryden (under grant NAG4-131) for their financial support.

References

[1] Ashiru, I. and Czarnecki, C. (1995). Optimal Motion Planning for Mobile Robots Using Genetic Algorithms, *Proceedings of the 1995 IEEE/IAS International* Conference on Industrial *Automation* and Control.

[2] Chang, T.-Y., Kuo, S-W., and Hsu, J. Y.-J. (1994). A Two-Phase Navigation System for Mobile Robots in Dynamic Environments, Proceedings *of the 1994 IEEE/RSJ/GI International* Conference on *Intelligent Robots* and Systems, pp. 297-300.

[3] Hocaoglu, C. and Sanderson, A. (1996). Planning Multi-Paths Using Speciation in Genetic Algorithms, Proceedings *of the 1996 IEEE International* Conference on Evolutionary *Computation,* pp. 378-383.

[4] Jang, J.-S. R., Sun, C.-T., and Mizutani, E.. *Neuro-Fuzzy* and *Soft* Computing, Prentice-Hall, Inc., 1997.

[5] Lin H-S, Xiao, J., and Michalewicz, Z. (1994). Evolutionary Algorithm for Path Planning in Mobile Robot Environment, Proceedings *of the 1st IEEE* Conference on Evolutionary Computation. *IEEE World* Congress on Computational *Intelligence,*

[6] Michalewicz, Zbigniew. (1994). *Genetic Algorithms + Data Structures = Evolution* Programs, 2nd Edition, Springer-Verlag Artificial Intelligence Series, 1994.

[7] Michalewicz, Zbigniew. (1994). Path Evaluation in the Evolutionary Planner/Navigator, Proceedings *of the 1995 International Workshop* on *Biologically* Inspired Evolutionary Systems, Tokyo, Japan, May 30–31, 1995, pp.45–52.

[8] Shibata, T., and Fukuda, T. (1993). Intelligent Motion Planning by Genetic Algorithm with Fuzzy Critic, Proceedings *of the 1993 International* Symposium on *Intelligent* Control.

[9] Xiao, J., Michalewicz, Z., and Zhang, L. (1996). Evolutionary Planner/Navigator: Operator Performance and Self-Tuning, Proceedings *of the 1996 IEEE International* Conference on *Evolutionary* Computation, pp. 366-371.

Hybrid Fuzzy PID Controller Design for PEBB Controlled DC Motor with Optimized Coefficients

V. Kunchithapadam, M.Bikdash, A.Homaifar

vlakshmi@ncat.edu, bikdash@ncat.edu, homaifar@ncat.edu

NASA-Autonomous Control Engineering Center

Department of Electrical Engineering

North Carolina A&T State University

Greensboro, NC 27411

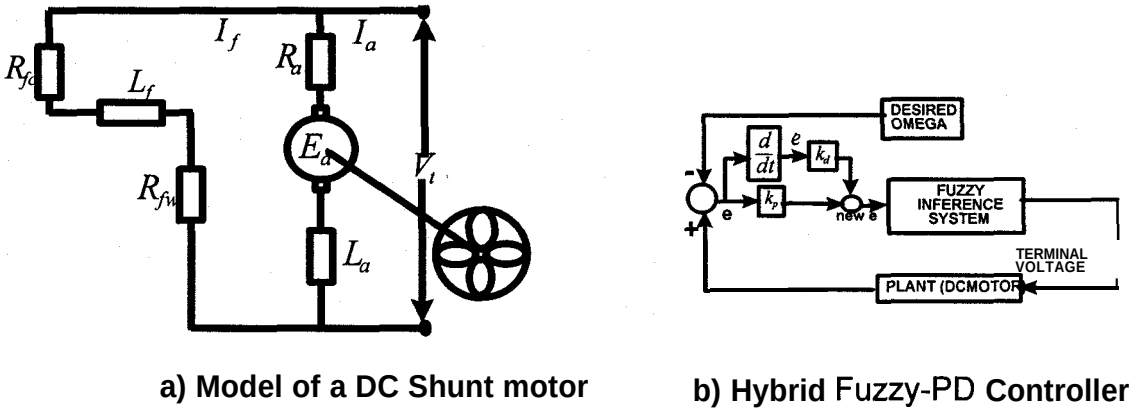
ABSTRACT: We propose a Hybrid Fuzzy Linear controller for a PEBB controlled DC motor. We have developed a nonlinear model of a DC Shunt Motor that includes the effect of magnetization saturation and field excitation. The three states considered here are the armature current, the field current and the angular velocity. The control input is the terminal voltage. We characterized the equilibrium state for any demanded power level. Fuzzy logic provides an accurate and efficient approach for systems which are difficult to model, have a strong nonlinearity, or are embedded in a changing environment with uncertainty. The input to the Fuzzy Logic Controller is the aggregate error between the required and the actual rpm. The output of the Fuzzy Engine is the control (terminal voltage). The design of the rules and membership functions are conducted using heuristic arguments and extensive simulation. The Fuzzy Logic Controller is designed to approximate a Bang-Bang Controller which is the usual solution for minimum-time control problem. We have also done a parametric study of the effect of overlap in the membership functions on the response in order to find the optimum tradeoff between short rise time, small steady state errors and control chattering. The performances of the openloop and closedloop systems are simulated in the presence of sinusoidal disturbances of different amplitudes. The Fuzzy controller is modified into a hybrid Fuzzy PID Controller to control the motor even in the presence of the disturbances. The best coefficients of the PID controller are obtained by an optimization process in MATLAB.

1. Introduction

DC motors are easily controllable and have dominated the adjustable speed drive field. AC motors are more expensive to control and are used in drive systems where special features of AC motors, such as the absence of commutators and brushes are desirable (Fitzgerald et al ., 1971). Speed control of DC motor above and below the rated speed can be easily achieved without power circuit switching, it is simple and less expensive. DC motors respond quickly to changes in control signals due to their high ratio of torque to inertia. Shunt motors are suitable where constant speed is needed at any preselected value or where appreciable speed range is required. Most shunt motors operate from adjustable voltage supplies and do not need auxiliary starting equipments. A stabilizing winding helps prevent the speed from increasing as the load increases at weak field settings (Gyorki et al., 1991). There are many methods of controlling the speed of a DC motor; one of them is using a thyristor. The armature voltage is adjusted by controlling the electrical angle within the ac wave at which the gate signal is applied to each thyristor. A thyristor is made to switch on and off at a fast rate so that the applied voltage divides between the armature and the switch, resulting in a controllable average armature voltage.

In our research, we adopted a simple nonlinear model of a DC shunt motor and from the state equations, we characterized the equilibrium state for any power level. The Fuzzy Logic Controller (FLC) is designed to approximate a bang-bang controller using Sugeno inference. A rough knowledge of the system behavior was needed. The input to the FLC is an aggregate error and the output of the FLC is the terminal voltage.

2. Model of DC Motor



a) Model of a DC Shunt motor **b) Hybrid Fuzzy-PD Controller**

Figure 1. Used Models.

In a shunt motor, the armature circuit and the field circuit are connected across a dc source of voltage V_t. The equations governing the operation of the DC motor are

$$u = V_t = I_a R_a + E_a + \frac{d}{dt}(I_a L_a) \quad (1)$$

$$It = I_a + I_f \quad (2)$$

$$u = V_t = I_f(R_{fc} + R_{fw}) + \frac{d}{dt}(I_f L_f) \quad (3)$$

where V_t is the terminal voltage and constitutes the control signal, I_a is the armature current, R_a is the armature resistance, E_a is the armature emf, Φ is the flux, w is the angular velocity, R_{fc} is the resistance of the control rheostat in the field, and R_{fw} is the resistance of the field winding. The flux Φ is assumed to be nonlinear in relation to I_f and that it saturates. We assume that a quadratic fan load is connected to the shaft, so that the Torque T is proportional to ω^2 and its power is proportional to w^3. The electrical torque of the system which equals the mechanical torque is

$$T_e = K_a \Phi I_a = \alpha_f \omega^2 + B\omega + J\dot{\omega} \quad (4)$$

where α_f is the fan coefficient, B represents a damping factor and J the rotor and fan inertia. The three states are the armature current, the field current and the angular velocity; the terminal voltage is assumed to be the control input. Hence we obtain the state-space model.

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} u - x_1 R_a - K_a \Phi(x_2) x_3 \\ u - x_2(R_{fc} + R_{fw}) \\ K_a \Phi(x_2) x_1 - \alpha_f x_3^2 - B x_3 \end{bmatrix} \quad (5)$$

The equilibrium condition of the motor is obtained by equating $\dot{x}$ to zero. Fans, pumps etc, are required to deliver a given power or torque at a certain time. Hence it is required to determine the corresponding states and the terminal voltage at any desired power or torque level. The angular velocity required for a particular power level can be determined from $P_f = \alpha_f \omega^3$. With x_3 = w given, the corresponding states and control can be found from Eq (6).

3. Fuzzy Logic Control of DC Motor

In recent years FLCs for electrical machines have attracted considerable attention. Fuzzy logic is a flexible method for implementing non-linear functions and it calls for expert knowledge, expressed linguistically or numerically (Reay et al ,1995). The operation of an FLC may be split into three stages : (a) Fuzzification of the crisp inputs (b) Fuzzy inference of the system using a knowledge base (c) Defuzzification of the inference to give crisp output.

Table 1: Parametric Study of effect of overlap in membership function

Expt no.	Amount of overlap(rad/sec)	Duration for maxV(sec)	Settling time (see)		Chattering
			Openloop	Closedloop	
A2	*30*	.21	1	.54	None
A3	*20*	*.25*	1	*.47*	None
A4	16	.27	1	.43	None
A5	8	.31	1	.33	None
A6	2	.36	1	.33	Slight
A7	*0.2*	*.38*	1	1.98	Medium
A8	*0.04*	*.38*	1	*2.00*	High
A9	*0,02*	*.38*	1	1.98	High
A10	0.01	*.38*	1	1.98	Very High
A11	*0.005*	*.38*	1	1.99	Very High

In this research, we have used a Sugeno-type controller which approximates any control law by mapping the states into control actions within an arbitrary accuracy (Clifton, 1996). The input to the Fuzzy Inference System is the error which is the difference between the actual speed and the desired speed. The membership functions of the error are trapezoidal as shown in Figure 2. The type of defuzzification used here is weighted average.

The closed-loop system consists of a Fuzzy Inference System (FIS) as defined by the MATLAB Fuzzy logic toolbox. The rules for our controller are very simple:

```
IF error is Positive THEN Voltage is Vmin
IF error is Zero THEN Voltage is Vdesire
IF error is Negative THEN Volatage is Vmax
```

These rules are based on the concept of quasi bang-bang controller. Bang-Bang control usually results when attempting to force a continuous-time system, x = $Ax + Bu$, from an initial state to the origin in *minimal time* under the constraint that the control signal magnitude is bounded. In most situations, the control signal *'bangs'* between its lower and upper limits (Malmgren, 1995). In our controllers however, the overlap of the membership functions, will result in interpolating the control action in the overlap region and thus smoothing it. Hence the *"quasi"* bang- bang form.

The model of the closed-loop system is shown in Figure 1(b). When K_p equals 1 and K_d equals O, it behaves like a simple FLC. The velocity of the motor is compared with that of the desired velocity (1200 rpm) to get the error. This error is input into the Fuzzy inference System. The FIS evaluates this error and decides the necessary control action to be taken. It is found that the closedloop controlled system has a shorter rise time and smaller steady-state errors when compared to the open-loop system.

SIMULATIONS: First we examine the open-loop response of Eq (6) under a large input step. This is simulated using the Runge-Kutta algorithms as implemented in ODE45 in MATLAB. Here we are trying to move the speed of the motor from 0.5 to 1.2 times the rated speed, without feedback. The initial speed is set for 500 rpm, and the initial state is found as the corresponding equilibrium. The openloop behavior shall be compared with the closedloop behavior in our subsequent plots. We conduct a parametric study of the effect of the overlap of the membership functions of the input (error), in the absence of disturbances. The study is summarized in Table 1.

The amount of overlap ω_{ov} was decreased from 30 rad/sec to 2 rad/see, and it is found that the settling time for the closed-loop decreases from 0.54 sees to 0.33 sees, but as the overlap is further decreased, the settling time increases to 1.98 sees (Figure 3). The duration for which the control voltage remains at the maximum value increased from 0.21 sees to 0.38 sees, but the voltage started to chatter V_t at an overlap of about 2 rad/sec and this chatter increases as the overlap is further

Table 2: Performance of a Hybrid Fuzzy PD Controller Here ω_{ov} represents the overlap in rad/see, and the control chatter is measured as std(diff(omega)))).

Disturbance	Controller	Exp	ω_{ov}	$K_d\Delta t$	K_p	mean(w) rad/sec	std(ω)	w Chatter
T_d =30 percent of Rated Torque $\omega_d = 2\pi/0.5$	Openloop	B1	8	0	0	124.9276	8.7408	1.0984
	FIS	B2	8	0	1	124.6245	2.5211	0.3530
	FIS/PD	B3	8	1	3	124.8452	1.5621	1.7125
	FIS/PD	B4	8	1	2	124.7519	1.6500	1.5992
	FIS/PD	B5	8	1	5	124.8318	1.6556	1.9940
	FIS/PD	B6	11	1	1	124.5800	2.8326	0.4024
	FIS/PID K_i=.2985	B7	8	-0.30	2.06	124.9924	1.5996	1.6262
T_d =10 percent of Rated Torque $\omega_d = 2\pi/0.5$	Openloop	B8	8	0	0	124.7992	2.8704	0.3625
	FIS	B9	8	0	1	125.3382	0.7846	0.1054
	FIS/PD	B10	8	0.01	1	125.3383	0.7844	0.1054
	FIS/PD	B11	8	0.75	1	125.3360	0.7658	0.1454
	FIS/PD	B12	8	1	1	125.2982	0.7336	0.2985
	FIS/PD	B13	10	1	1	125.2995	0.8917	0.1249
	FIS/PD	B14	11	1	1	125.2713	0.9518	0.1244
	FIS/PD	B15	8	6	1	116.5863	2.3769	2.2794
	FIS/PID K_i=2.1	B16	8	1.05	2.7	125.4942	0.3981	0.06067

reduced (Figure 4). The tradeoff between fast response and chattering is at 8 rad/sec.

We consider a sinusoidal torque disturbance $d(t) = T_d \cos \omega_d t$, where T_d refers to the amplitude of the disturbance added, and ω_d refers to the frequency of the disturbance. Different experiments are done with T_d = O. $1T_{rated}$ and $0.3T_{rated}$ (Figures 5 and 6).

HYBRID FUZZY PD CONTROLLER: The input to the FIS is now $K_p e + K_d \dot{e}\Delta t$. Table 2 summarizes our study with a disturbance of amplitude T_d. For a disturbance of O. $1T_{rated}$, different values of K_d *are* used, with K_p at a constant value of 1. The best performance with regards to less standard deviation was noticed at the expense of chattering (B12). To reduce the amount of chattering, the overlap was increased to 11 rad/see, and the chattering and the rpm oscillation were suppressed very well. The same combination of K_p and K_d worked reasonably well on a disturbance of $0.3T_{rated}$. However in this case, the best result was achieved in experiment B3 with K_p equal to 3 and K_d equal to 1..

OPTIMIZATION OF CONTROLLER COEFFICIENTS: The coefficients used in the above study was chosen in an adhoc fashion. In order to get the best value of the coefficients which reflect short rise time, less overshoot, chattering, and error in angular velocity, an optimization was done with an initial guess for the controller coefficients. Instead of a PD controller, a PID controller was considered here. The values of the coefficients for disturbance which is 10°/0 of the rated torque are:K_p= 2.0624, Kd= -.3018, K_i=.298. The values of the coefficients for a disturbance of 30% of the rated torque are: K_p= 2.7, Kd = 1.05, K_i =2.1. The performance results are given in B7 and B16. The results for B 16 are much better than the earlier experiments and the results for B7 can be considered as optimum. The performance of the best coefficients for a disturbance of 10°/0 of the rated torque is given in Figure(7).

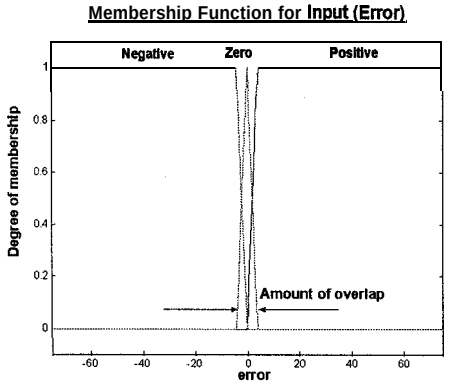


Figure 2. Membership function for input to FIS (Error in omega)

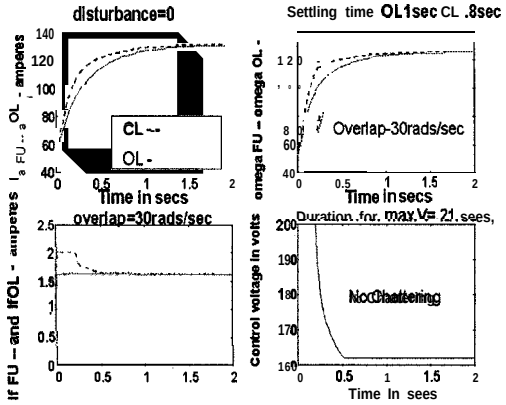


Figure 3. Comparison of openloop and closedloop performance for an overlap of 30 rad/sec

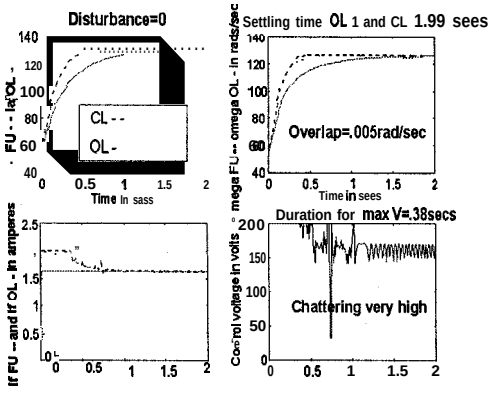


Figure 4. Reducing the overlap introduces significant chatter in the control

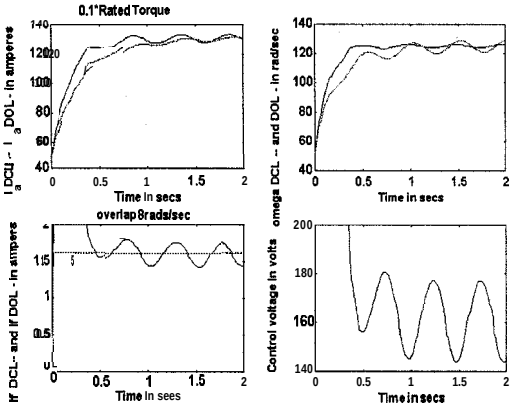


Figure 5. Effect of torque disturbance $(\mathrm{T}_d = 0.1\mathrm{T}_{rated})$

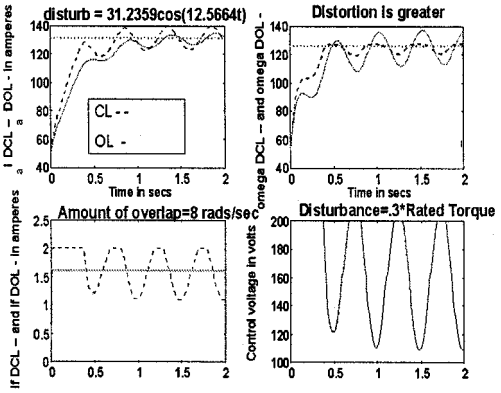


Figure 6. Effect of large torque disturbance $(\mathrm{T}_d = 0.3\mathrm{T}_{rated})$

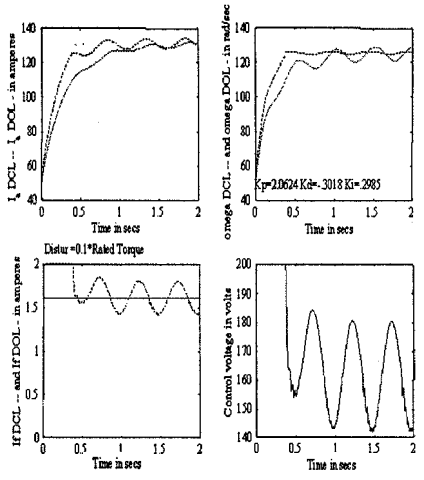


Figure 7. Performance of a PID controller with optimized coefficients $(\mathrm{T}_d = \mathrm{O}.1T_{rated})$

4. Conclusions

We used a nonlinear model of a DC motor with saturated magnetization and quadratic mechanical load. We designed a simple Sugeno quasi bang-bang controller to control the speed of the motor. It is found that the designed controller achieves fast response and there is no steady-state error. In the parametric study of overlap, we found that there is a trade-off in the amount of overlap, in terms of speed of response and chattering. We found that a disturbance of 0.1 T_{rated} can be controlled well at the expense of some chattering, but the same combination of the PD controller coefficients work reasonably well with a disturbance of higher magnitude, though better performance maybe achieved with other combinations.

We optimized the performance of the controller with the best value of the coefficients.

ACKNOWLEDGEMENT

This work is supported by the US office of Naval Research through contract No N 00014-96-1-1123. The authors would like to thank them.

REFERENCES

1. C. Clifton, "Design and application of generalized sugeno controllers," *MS thesis, NC* A&T State University, (1996).
2. A.E. Fitzgerald, C. Kingsley, Jr., and A. Kusko, *Electric Machinery,* McGraw-Hill Kogakusha, Ltd., (1971).
3. D.S. Reay, M.W. Dunnigan, M. Mirkazemi-Moud, and B.W. Williams, "A comparison between fuzzy and linear controllers applied to a switched reluctance motor based position servo," *IEE Colloquium on Advances in Control Systems for Electric Drives,* (May 1995) pp. 49-52.
4. W. Brogan, *Modern Control Theory,* Prentice Hall, Englewood Cliffs, N.J. 07632,(1991).
5. A. Malmgren, "Contraction based state feedback control," *Ph.D Thesis,* University of Uppsala (1995).
6. A.E. Bryson and Y. C. Ho, *Applied Optimal Control,* Hemisphere Publishing Corporation, NY, (1975).
7. C.C. Lee, "Fuzzy logic in control systems: fuzzy logic controller: Part I & Part II," *IEEE Trans Syst Man and Cybernetics,* vol.20, (March /April 1990) pp. 404-435,.
8. J.R. Gyorki and L.A. Berardinis, "Electrical and Electronic systems", *Machine Design,* Penton publication, (June 1991) pp. 324-326.

Production of High Intracavity UV Power From a CW Laser Source

R. T, David, T. H. Chyba, C. E. Keppel
Hampton University

D. Gaskell
Oregon State University

R. Ent
Thomas Jefferson National Accelerator Facility

Corresponding Author: Romuald David, Research Center for Optical Physics, Department of Physics, Hampton University Hampton, VA 23668
romuald@gprc.hamptonu.edu

The goal of this research project is to create a prototype high power CW source of ultraviolet (UV) photons for photon-electron scattering at the Thomas Jefferson National Accelerator Facility (TJNAF), Hall B. The facility will use optical resonant cavities to produce a high photon flux. The technical approach will be to frequency-double the 514.5 nm light from an Argon-Ion Laser to create 0.1 to 1.0 watt in the UV. The produced UV power will be stored in a resonant cavity to generate an high intracavity UV power of 102 to 103 watts. The specific aim of this project is to first design and construct the low-Q doubling cavity and lock it to the Argon-Ion wavelength. Secondly, the existing 514.5 nm high-Q build-up cavity and its locking electronics will be modified to create high intracavity UV power. The entire system will then be characterized and evaluated for possible beam line use.

1. Introduction

Using a frequency doubling method followed by a cavity enhancement technique, high energy photons will be produced for photon-electron scattering experiments at Thomas Jefferson National Accelerator Facility (TJNAF). Efficient CW frequency doubling in optical resonators has been demonstrated by several groups. In 1992, M. Oka and S. Kubota generated continuous-wave fourth-harmonic ultraviolet light to read out high density optical disks and to reduce the noise and diffraction in the beam [1]. They used a BBO (Beta Barium Borate) crystal in a linear external cavity to convert green power into ultraviolet power with 10°/0 conversion efficiency. Using intracavity frequency doubling, M. Oka *et al.* converted an IR beam from an Nd:YAG laser into green light [2]. This visible light was then sent into an external four mirror resonant cavity in which a BBO crystal generated C W ultraviolet radiation . The team used the Pound-Drever-Hall (PDH) technique [3] to lock the frequency of the green laser to the external resonator. A 54% conversion efficiency from incident green power to UV output light was obtained.

S. Owa *et al.* used an external ring resonator [4], rather than an intracavity technique to generate a deep ***W source with*** 53% conversion efficiency. They too have used the PDH technique to match the resonator frequency to the input light frequency. In their method, an EOM (Electro-Optical Modulator) modulates the frequency of the input light. A PZT (piezoelectric transducer) controls the cavity length. Using CW frequency doubling methods, excellent conversion efficiencies have been obtained with nonlinear crystals in general and more than 50% efficiency of ***W*** light generation from a BBO crystal in particular has been reported.

2. Objectives and Technical Approach

The goal of this research is to develop and test a prototype high power source of ***W*** photons. This system is now under construction in an optics laboratory at TJNAF. If suitable, this system will be modified for use in the electron beam. We will resonantly double the 514.5

nm line from an Argon-Ion laser to create 257 nm light. This low power source of UV photons is then resonantly enhanced by storing the photons in a high-Q ring cavity. In its nuclear physics application, the electron beam intersects the UV photons at the crossing point in the center of the high-Q (UV enhancement) cavity.

Figure 1 illustrates the optical layout. The Argon-Ion laser resonator is formed by a high reflector and a curved output coupler (OC). The gain medium is the electrically pumped plasma tube. The intracavity prism selects the 514.5 nm transition for the laser. The intracavity etalon maintains single longitudinal mode operation. The adjustable intracavity aperture is set to produce TEM_{00} operation with maximum output power cavity. The EOM-2 in the cavity provides fast feedback to the laser frequency to lock it to the external high-Q ring cavity. Once out of the Argon-Ion resonator, the light passes through an optical isolator to protect the laser from optical feedback. The 1/2 wave retarder rotates the output linearly polarized light to any linear orientation. The two pairs of mode matching lenses in the system couple the TEM_{00} spatial modes of the external cavities with those of the incoming beams.

The low-Q cavity will frequency double the 514.5 nm light to 257 nm. DET 1 is used for the low-Q cavity feedback loop. DET 2 monitors the total power out of the cavity, EOM 1 modulates the UV beam for the Drever-Pound loop which locks the high-Q cavity with the laser. The UV light is next sent through the Acousto-Optical Modulator (AOM) that serves as a fast on/off switch to measure the high-Q cavity decay time. An electro-optic polarization rotator can rapidly switch the polarization of the UV photons. The 257 nm light is then sent to the high gain cavity. The PZT's in the system are used for microscopic motion of the mirrors and to provide slow feedback to the cavities to follow the thermal drifts of the resonators. DET 3 measures the frequency modulated resonance signal from the high-Q cavity. DET 4 measures the power lost through one of the mirrors of the cavity, the cavity decay time and the finesse. The high-Q cavity is designed to create a beam waist at the crossing-point. At this location, the intracavity irradiance is a maximum.

2.1 Design and Construction of the Low-Q Cavity

For our prototype, an external design has been chosen as shown in figure 2. We will take advantage of the high hi-directional conversion efficiency (53%) obtained by Owa *et al.* [4] by coupling the green beam into the cavity through a mirror which is highly reflecting in the UV and the visible. Spherical mirrors will be used for the cavity, but elliptical mirrors will be tested if time allows. The advantage of elliptical mirrors [5] is that the irradiance can be high but the beam divergence can be kept low in the sensitive plane of the crystal.

All the mirrors in the low-Q cavity will be highly reflective at 514.5 run. Ideally, the input coupler will also have 100% reflectivity at 257 nm, to prevent the *W* light from exiting the cavity back towards the Argon-Ion laser. However, the second mirror will have 0% reflectivity at 257 nm in order to allow high transmission of the produced *W* out of the cavity. This low reflectivity also prevents the *W* from reflecting back through the doubling crystal a second time which will cause additional losses.

A Brewster-cut BBO crystal will be placed in the low-Q cavity to frequency double the 514.5nm light. The crystal is chemically stable. Its high transmission and large angular acceptance in the visible makes the BBO crystal an excellent choice for doubling the frequency of the Argon Ion laser. Table 1 summarizes its key properties. An AR coated BBO crystal has 85% transmission at a wavelength of 257 nm.

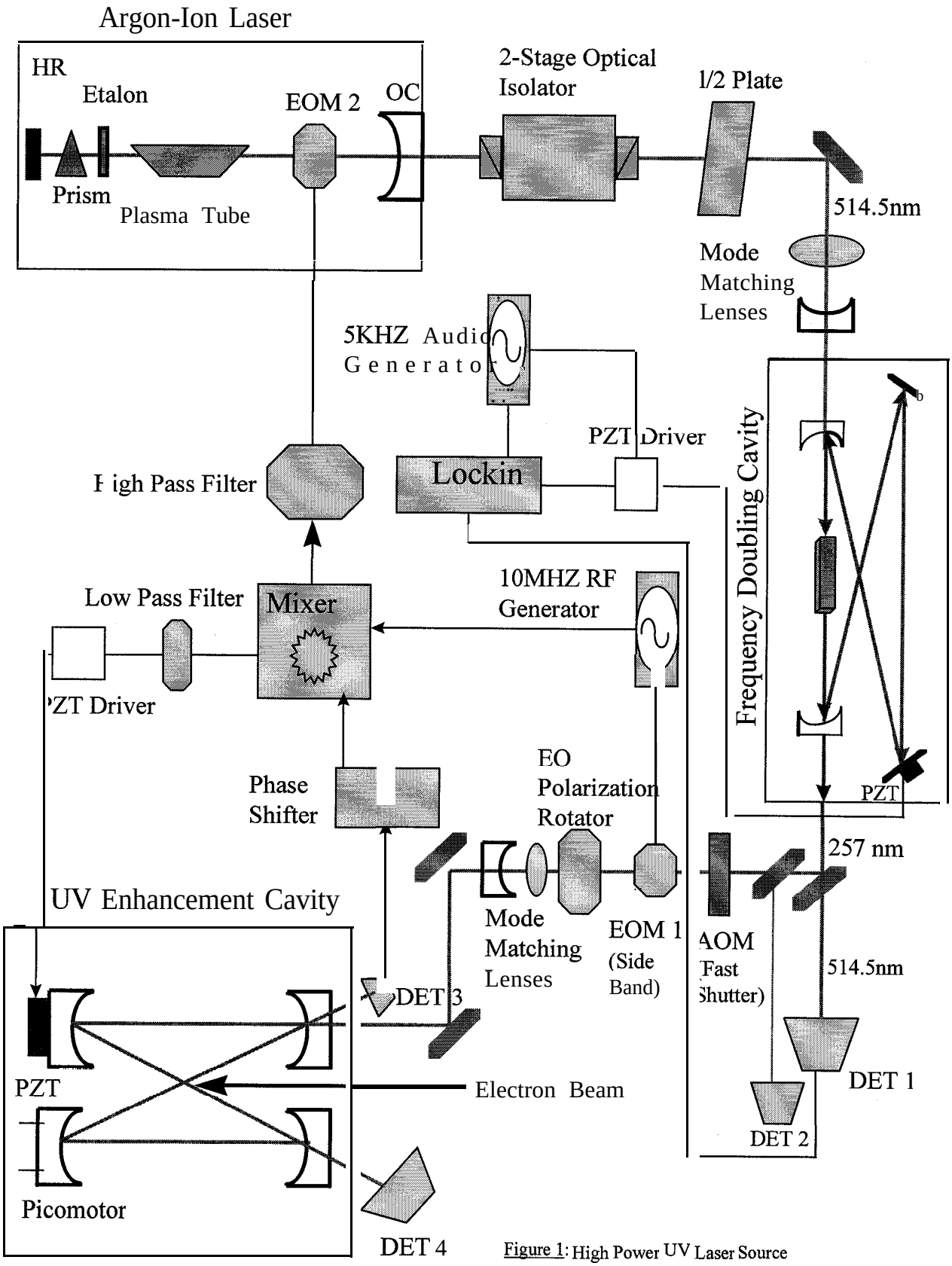
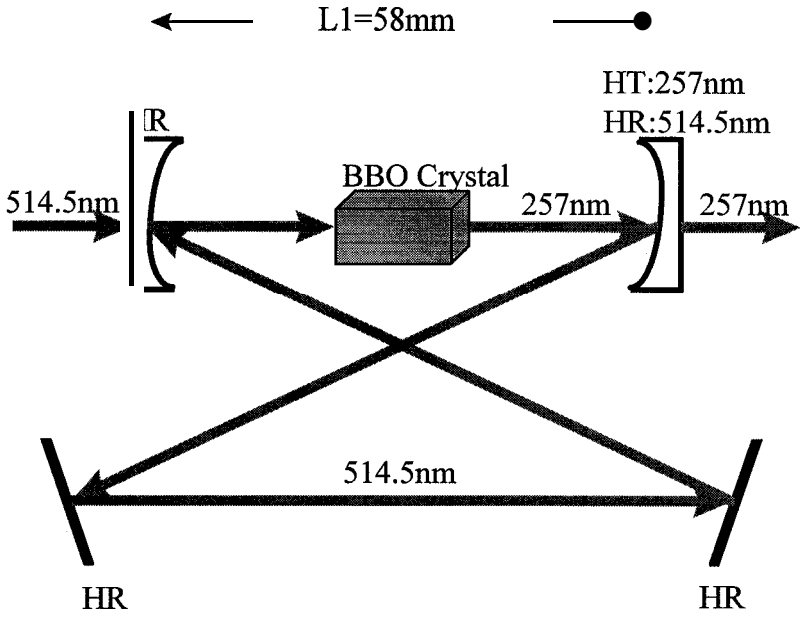


Figure 1: High Power UV Laser Source



Beam Waist Radius= 12 μm
Input Power, 1_0=1 W
Cavity is stable for L_{Tot}= 408 mm and R_c=50mm

Figure 2: Frequency Doubling Cavity

Physical Properties	Optical Properties	Nonlinear Properties
Melt. Pt. T_m=1095±5°C	Refract. Indices @266nm	Angular Accept. @532 nm
Transit. Temp.T_T=925±5°C	n_o=1.7571, n_e=1.6139	$\Delta\Theta_m$=0.5 mrad x cm

Table 1: Characteristics of the BBO crystal [6].

2.2 Preliminary Results

The Argon-Ion Laser has a single frequency output power "of about 1 W. The beam radius w, is about 0.6 mm at the output coupler and the far field divergence half angle θ_{FF}, is 0.68 mrad. Some preliminary results have been obtained for the high-Q resonator for green light. The value obtained for the finesse, F, is about 200 using mirrors with reflectivities R close to 99.9°/0. The high-Q cavity will be tested in the visible with mirrors whose reflectivities are greater than 99.99%. The design of the frequency doubling cavity is completed, and its construction has begun. The necessary electronics for locking the cavities are still under construction.

The stable region, the beam radii, and the Rayleigh range for the low-Q cavity without the BBO crystal have been calculated. These are plotted in Figures 3, 4, and 5 below. The total cavity length is $L_{Tot} = L_1 + L_2$, where L_1 is the separation between the curved mirrors. Four different values of L_2 (250 mm, 280 mm, 350 mm, 400 mm) have been used in the figures. In Figures 3 and 5, for a given curve, the cavity is stable for values of L_1 less than the value of L_1 at which that curve has a minimum. The cavity is stable for a total length of 408 mm using spherical mirrors of radius of curvature 50 mm. For L_1= 58 mm, the beam waist in the BBO crystal is 12 μm while the corresponding Rayleigh range, Z_R= 0.82 mm. The beam radius on each mirror has also been calculated and is illustrated in Figure 5. The corresponding beam radius on each mirror is 18 pm.

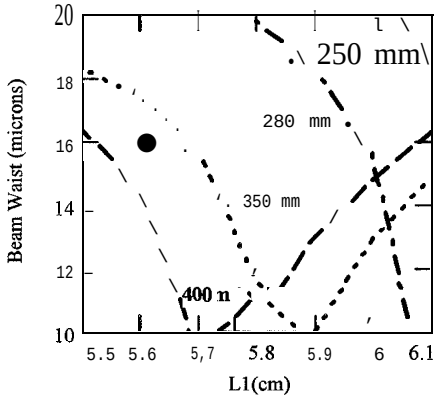


Figure 3: Beam Waist Radius Vs L_1 for 4 different L_2 values

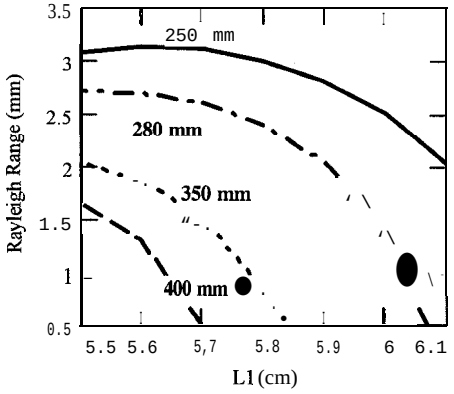


Figure 4: Rayleigh Range

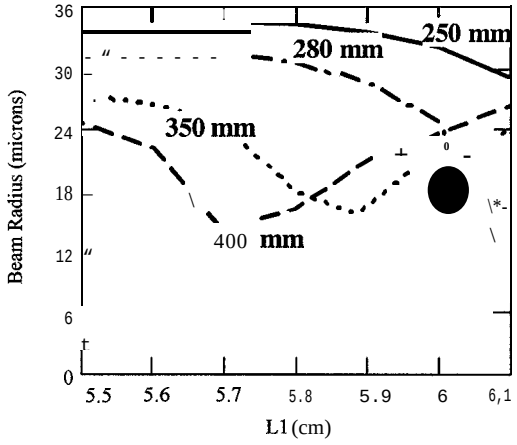


Figure 5: Beam Radius on each mirror

3. Future Work

After assembly of the low-Q cavity, different methods will be tested to lock its resonant frequency to the Argon-Ion frequency. Efficient frequency doubling requires high power which is only possible if both the spatial distribution and frequency of the Argon-Ion laser are matched to the low-Q cavity. The PDH technique could be used for this. After locking the cavity, we will measure its output power in the UV.

The existing ring resonator will be modified for UV use. The present mirrors will be replaced by new mirrors with high reflectivity (>99.9%) for *W* enhancement. Optical filters will be replaced by new ones suitable for *W* light. The present mode matching lenses and the acousto-optical modulator (AOM) will be replaced as well. Both the low-Q and high-Q resonators have to be simultaneously locked to the Argon-Ion laser. The same Drever-Pound technique used to lock the high-Q cavity in the visible will be used in the *W*. We will measure the intracavity *W* power in the *W* enhancement resonator.

As shown in figure 1, the high-Q cavity configuration will be similar to that currently in use at TJNAF for enhancing the 514.5 nm output from an Argon-Ion laser. After complete construction the prototype will then be tested.

Acknowledgments:

This work is sponsored by the DOE, the NSF, and by NASA grant NAGW2929.

References

1. Michio Oka and Shigeo Kubota, "Second-Harmonic Generation Green Laser for High-Density Optical Disks", Jpn. J. Appl. Phys., 31,513-518, (1992).
2. M. Oka, L. Y. Liu, W. Wiechmann, N. Eguchi, S. Kubota, "All Solid-State Continuous-Wave Frequency-Quadrupled Nd:YAG Laser" IEEE Journal of Selected Topics in Quantum Electronics, 1,859-866 *(1995).*
3. R. W. P. Drever, "Laser Phase and Frequency Stabilization Using an Optical Resonator", Applied Physics Journal B 31,97-105, (1983).
4. S. Owa, Y. Taira, S. S. Kane, "Efficient Second-harmonic generation using an external resonant cavity for C W ultraviolet radiation", Conference on Lasers and Electro-Optics, 11, 636-637, (Technical Digest Series, OSA, Washington, D. C., 1993).
5. Y. Taira, "Continuous 6-watts coherent deep *W* generation using a resonant frequency doubling technique", Conference on Lasers and Electro-Optics, 11,634, (Technical Digest Series, OSA, Washington, D. C., 1993).
6. R. Aimis, "Lasers, Optics, and Electronics catalogue", EKSMA Co., Vilnius, Lithuania.

Plasma Processing of Metallic and Semiconductor Thin Films in the Fisk Plasma Source

Gregory Lampkin, Edward Thomas, Jr.,[a] Michael Watson, Kent Wallace, Henry Chen, and Arnold Burger

NASA/Fisk University Center-for Photonic Materials and Devices
Department of Physics, Fisk University
Nashville, TN 37208

[a] Contact: etjr@dubois.fisk.edu

Abstract

The use of plasmas to process materials has become widespread throughout the semiconductor industry. Plasmas are used to modify the morphology and chemistry of surfaces. We report on initial plasma processing experiments using the Fisk Plasma Source. Metallic and semiconductor thin films deposited on a silicon substrate have been exposed to argon plasmas. Results of microscopy and chemical analyses of processed materials are presented.

1. Introduction

Plasma technologies have been applied widely throughout the semiconductor industry. Plasma processing techniques offer several advantages over the traditional chemical processing. First, since a plasma is composed of ionized particles, the transport of those particles can be modified by the application of external electric and magnetic fields. This facilitates a greater degree of control over the production process. Second, plasmas can achieve high particle kinetic temperatures (> 5000 K) for power inputs as low as 50 to 100 W. Third, plasmas are typically produced under vacuum conditions, plasma processing reduces the possibility of contamination of sensitive electronic components.

Because of these properties, plasmas are used in the preparation and cleaning of surfaces,[1] etching of electronic components,[2] low energy ion implantation,[3] and in the production of diamond-like thin films. The rapid growth of this field of plasma processing is driven by the fact that plasmas offer an economically and environmental y attractive method of production.[4] However, in spite of the widespread application of plasma processing technologies, fundamental plasma-material interactions are not fully understood.

In this paper, we present preliminary results of plasma processing experiments performed using the Fisk Plasma Source. In these studies, materials were exposed to argon plasmas. The surface morphology is characterized before and after plasma exposure using atomic force microscopy. Changes in the surfaces are presented for a thick film and a thin film deposited onto a silicon substrate. A model that describes the heating of the sample surface by the plasma is also presented.

2. Facilities

Plasma processing experiments are performed using the Fisk Plasma Source (FPS). The FPS device is a highly flexible, multiple configuration plasma device that produces moderate density, low temperature plasmas using DC glow discharge, thermionic filament, and, in the near future, RF/microwave techniques. The plasma source has been operational at Fisk University since Summer, 1997.5

The major hardware component of the FPS is a stainless steel 6-way cross vacuum vessel. The vacuum vessel has a large volume; it has a 10" inner diameter with 13.5" - ASA 8 flanges and a length of 20" to 24" along each axis. A schematic of the FPS device, in a configuration to produce plasmas using the DC glow discharge technique, is shown in Figure 1. Plasmas are generated by applying a potential difference between the anode and cathode. The plasma forms initially between the anode and cathode, then expands to fill the chamber.

The experimental setup of the FPS device for the studies described in this paper are listed in Table 1. Measurements of plasma parameters on the FPS have been made using Langmuir probes and by optical emission spectroscopy. A summary of those plasma parameter measurements are given in the paper by Thomas, *et. al.*, in these proceedings.

Table 1: Experimental configuration for plasma processing experiments

Gas Pressure (p_0)	100 mTorr
Anode Voltage (V_A)	168 V
Cathode Voltage (V_C)	-283 V
Plasma density (n)	1 - 3 x 10^{14} m^{-3}
Electron temperature (T_e)	6 - 10 eV

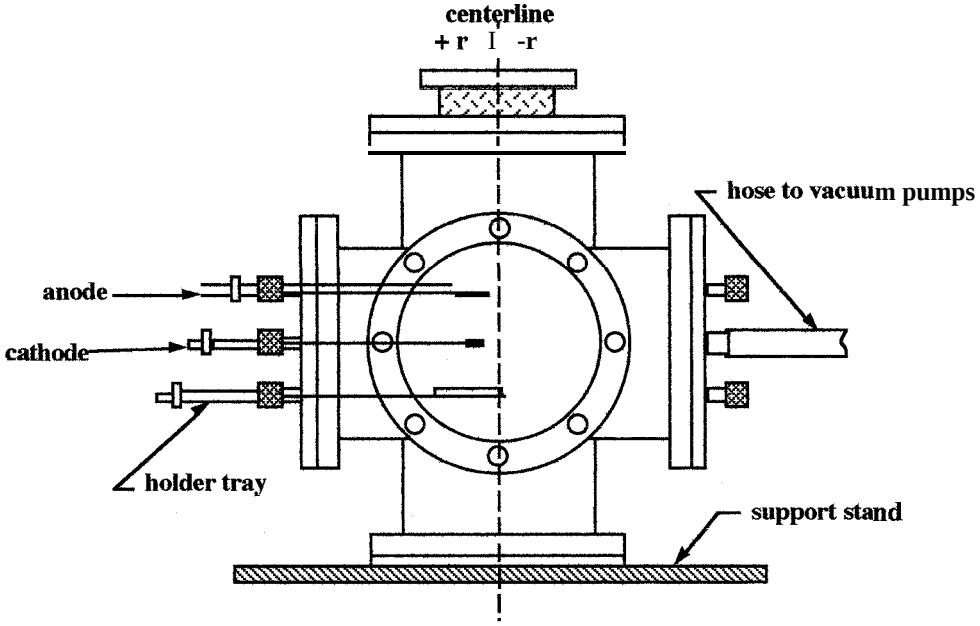


Figure 1: Schematic of the Fisk Plasma Source

3. Plasma exposure experiments

The objective of these first plasma exposure experiments was to test the feasibility of using the FPS device to perform controlled processing of materials. The specific task was to use argon plasmas to modify the surface morphology of films deposited on silicon substrates. The information gathered during these experiments will be used to develop new techniques for the fabrication of semiconductor-based devices.

A material sample that is to be exposed to the plasma is placed on a 5 cm x 5 cm brass tray which is inserted into the plasma. The tray can be moved radially in the FPS chamber. For the experiments described in this paper, the tray (and the sample) were centered in the plasma at a radius of r = 2 cm, at a height 4.5 cm below the cathode. The plasma density at the center of the tray was n = 2.5 x 10^{14} m^{-3} and the electron temperature was T_e = 10 eV.

Two samples were exposed to the plasma. The first was a sample of gold on a silicon substrate (Au/Si). The gold layer was sputtered onto the silicon by a magnetron sputtering technique in a 50 mTorr argon environment. The sputtering power was 50 W for a duration of 5 minutes. The second sample was erbium doped zinc selenide on a silicon substrate (ZnSe:Er/Si). This sample was also prepared by magnetron sputtering. Here, the sputtering conditions were a power of 90 W at a pressure of 170 mTorr for a duration of 2 hours. The sample was also heated to a temperature of 350 °C while being sputtered.

The surface of each sample was imaged using atomic force microscopy before and after plasma exposure. The mean surface roughness is used as a measure of the effect of the plasma exposure. A 500 nm x 500 nm AFM image of the surface of the Au/Si sample prior to plasma exposure is shown in Figure 2(a). A 1000 nm x 1000 nm AFM image of the surface of the ZnSe:Er/Si surface after plasma exposure is shown in Figure 3(a). The Au/Si

surface is a thick film of gold on the silicon substrate characterized by a mean surface roughness of 0.65 nm. The ZnSe:Er/Si surface is a thin film of ZnSe:Er on the silicon substrate characterized by a mean surface roughness of 4.5 nm.

Each sample was placed into the plasma as described previously. The Au/Si was exposed to an the argon plasma for 28 minutes. The ZnSe:Er/Si sample was exposed to the plasma for 10 minutes. The difference in the exposure times was to ensure that, the ZnSe:Er/Si sample, was not completely removed by the plasma. An AFM image of the surface of the Au/Si sample after plasma exposure is shown in Figure 2(b). The mean surface roughness after the plasma exposure is 0.23 nm. AFM image of the surface of the ZnSe:Er/Si sample after plasma exposure is shown in Figure 3(b). The mean surface roughness after the plasma exposure is 0.17 nm.

Exposure of both samples to an argon plasma causes a substantial reduction in the surface roughness. A comparison of the AFM images before and after plasma exposure indicates that the surface of each sample has been visually smoothed as a result of the plasma exposure. These results suggest that it may be possible to use argon plasmas to improve the smoothness of surfaces, thereby improving the adhesion of layers that may be deposited on the film. However, much work remains to be done to characterize changes in the electrical and optical properties of the plasma exposed samples.

4. Model for plasma-material interactions

In order to take advantage of the potential applications of plasma processing, it is necessary to develop a better understand the processes that occur at the boundary between the plasma and the materials. Plasma-surface interactions have been explored extensively in the realm of fusion-based plasma physics[6] in the context of energetic (i.e., $T_i, T_e > 100$ eV) particle impacts on plasma-facing walls. In recent years, these results have been applied to interpreting the complex physical and chemical processes that occur at the plasma-material boundary.

In these studies, we wish to understand the physical mechanisms that lead to the smoothing of the sample surfaces as a result of plasma exposure. On the macroscopic scale, it is possible to estimate the temperature rise of the plasma exposed surface due to the impacts from energetic particles from the plasma. However, the temperature rise is found to be - 1 to 10 C. This suggests that surface melting is not responsible for smoothing the surfaces during plasma exposure.

On the microscopic scale, particles from the plasma will impact atoms on the surface. As a result of the impacts, atoms on the surface will gain energy and will be free to move across the surface. Therefore, we wish to determine if there are sufficient interactions between the plasma particles and atoms on the material surface to account for the reduction in surface roughness.

Particles in a plasma with a density n and an electron temperature T_e will travel towards an electrically floating, plasma exposed surface at the Bohm speed, $v_B = (T_e/m_i)^{1/2}$. In a time, t_0, the particle flux (in units of particles/cm^2) towards the surface will be given by

$$\Gamma = n v_B t_0 \text{ (cm-z)} \qquad (1)$$

If the surface has a density of atoms given by U, then in a time t_0 the number of interactions between plasma particles and surface atoms will be given by:

$$N = \alpha\Gamma/\sigma \qquad (2)$$

where a is the probability y of the interaction. For simplicity, the assumption is made that a plasma particle will transfer all of its energy to an atom on the surface. It is further assumed that the total number of atoms in the film remains constant and interactions between atoms in the film are not yet included.

In the case of the ZnSe:Er/Si sample, the surface roughness was reduced from 4.5 nm to 0.17 nm. The typical lattice parameter for ZnSe is 0.6 nm corresponding to a value of $\sigma = 2.78 \times 10^{14}$ cm-z. For the argon plasmas used in these experimental studies, the flux (as given in Equation 1) during a 10 minute plasma exposure is: $\Gamma = 7 \times 10^{14}$ cm^{-2}. For $a = 1$, this gives N = 2.7 impacts/surface atom. Each impact would give the atom approximately E -4.8 eV of energy.

This result suggests that there may be sufficient energy carried by the plasma to allow substantial movement of surface atoms, thereby giving a preliminary explanation of the changes in the surface roughness. However, to fully understand the effect of the plasma on the film surface, many additional factors must be taken into account such as: possible surface damage caused by the impacts, the interaction between surface atoms, and the transport of plasma particles after colliding with surface atoms. Therefore, additional study is needed to understand the interactions between the plasma and the material surface.

5. Conclusions

The surface morphology of metallic and semiconductor films, deposited on a silicon substrate, and modified by argon plasmas has been studied. In both cases, the films were deposited onto the substrate using a magnetron sputtering technique. The samples were placed in DC glow discharge argon plasmas produced in the Fisk Plasma Source. The surface morphology of each sample was characterized before and after plasma exposure using atomic force microscopy. As a result of the plasma exposure, the surface roughness of the sample was reduced

These results demonstrate FPS device is capable of producing plasma conditions of relevance to plasma processing. Further, these results demonstrate that argon plasmas maybe used to smooth the surfaces of thin films. This may lead to an improvement in the adhesion and a reduction in the contact resistance of electrical contacts that are applied to the films. Our long term goal is to apply the results of this work towards the development of plasma processing techniques for $Cd_{1-x}Zn_xTe$ (CZT).

Acknowledgments

The authors of this paper would like to acknowledge the support of the members of the Fisk University Center for Photonic Materials and Devices without whom this work would have been impossible. The CPMD is supported by NASA Grant No. NCC-8- 133.

[1] Z. Lu, M. T. Schmidt, R. M. Osgood, Jr., W. M. Holber, D. V. Podlesnik, J. Vat. Sci. Tech. At 9 1040 (1991).

[2] J. Asmussen, J. Vat. Sci. Tech. A, 7883 (1989).

[3] J. Mantese, I. G. Brown, N. W. Cheung, and G. A. Collins, MRS Bulletin, 21, No. 8, 52 (1996).

[4] G. Collins and D. J. Rej, MRS Bulletin, 8 26(1996).

[5] for further information, refer to the following WWW page: http://www.fisk.edu/~etjr/psl.html

[6] see for example, P. C. Stangeby and G. M. McCracken, Nucl. Fusion, 30, 1225 (1990).

[7] Francis F. Chen, *Introductions of Plasma Physics and Controlled Fusion, Vol. 2, 2nd Ed.,* (Plenum Press, New York, 1984). Chap. 8, p. 290.

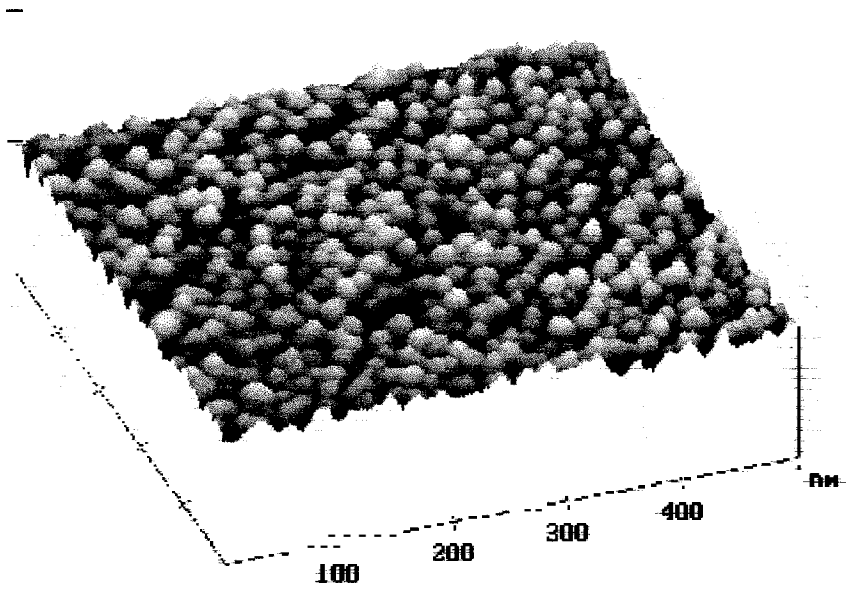


Figure 2(a): AFM image of the surface of the Au/Si sample prior to plasma exposure. The mean surface roughness is 0.65 nm.

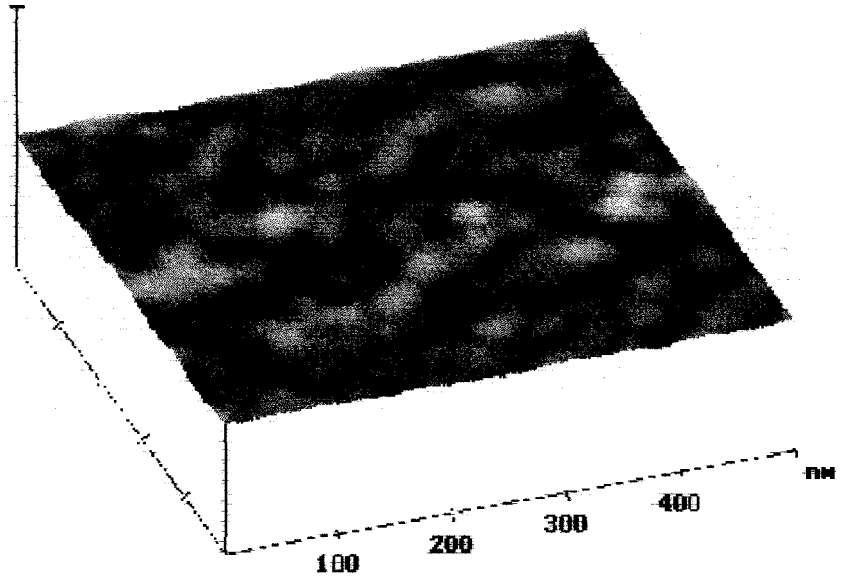


Figure 2(b): AFM image of the surface of the Au/Si sample after to plasma exposure. The mean surface roughness is 0.23 nm.

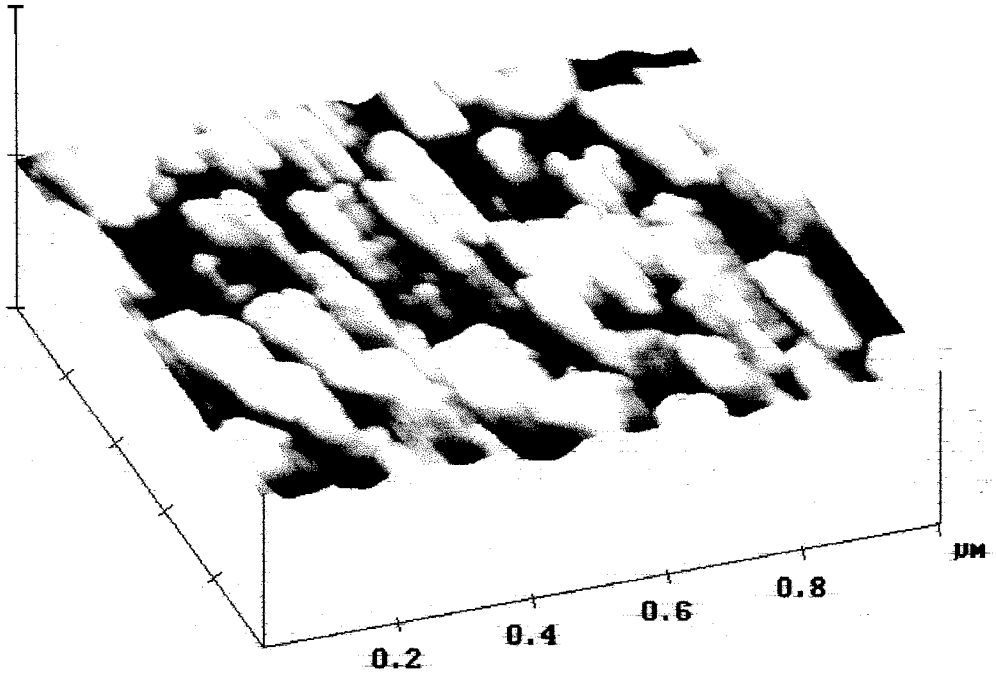


Figure 3(a): AFM image of the surface of the ZnSe:Er/Si sample prior to plasma exposure. The mean surface roughness is 4.5 nm.

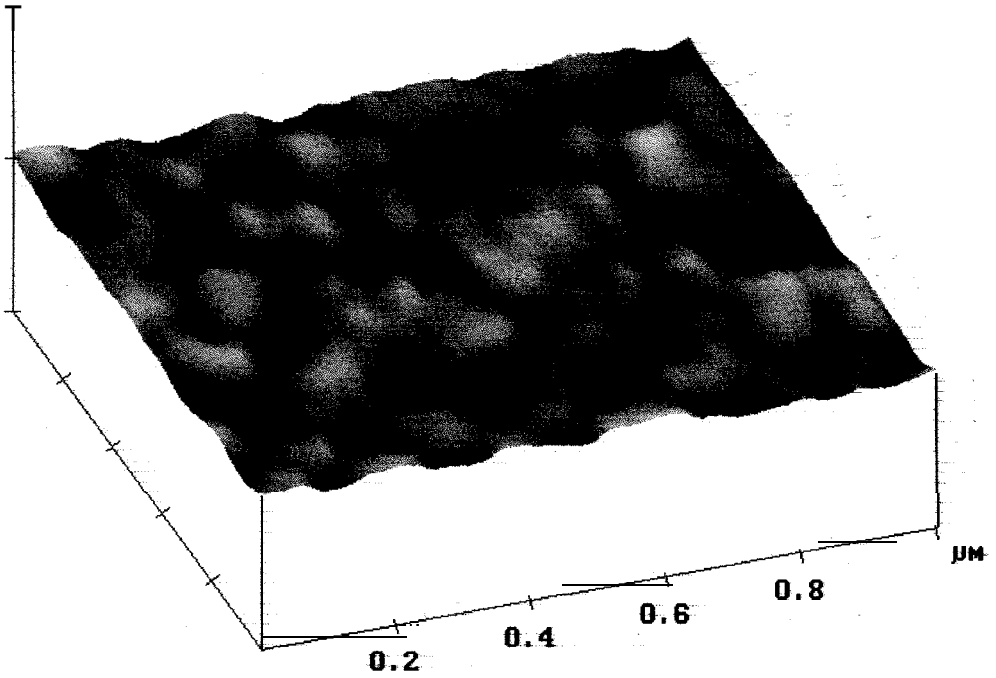


Figure 3(b): AFM image of the surface of the ZnSe:Er/Si sample after to plasma exposure. The mean surface roughness is 0.17 nm.

SYNTHESES OF OCTASUBSTITITED METAL PHTHALOCYANINES FOR NONLINEAR OPTICS

Huaisong Guo[a], Cheryl Townsend', Mohan Sanghadasa[b], Robert L. S. Amai[a], Ronald D. Clark[a] and Benjamin Penn[c]

'Department of Physical Sciences, New Mexico Highlands University, Las Vegas, New Mexico 87701.
[b]Department of Optical Physics, University of Alabama in Huntsville, Huntsville, Alabama 35899.
[C]Space Sciences Laboratory, NASA Marshall Space Flight Center, Huntsville, Alabama 35812.

ABSTRACT

Many organic materials can be used as nonlinear optical media. Phthalocyanines are of special interest because they show an unusually large third order nonlinear response, they are thermally and photochemically stable and they can be formed into oriented thin films (Langmuir-Blodgett films)[1]. They also can be easily complexed by a large variety of metals, which place them at the interface between organics and organometallics, and allows for fine tuning of the macro cycle electronic properties by the coordinated metal and substituent groups. A series of 1,4,8,11,15, 18,22,25-octaalkoxy metal-free and metal phthalocyanines and 2,3,9,10,16, 17,23,24-octaalkoxy metal phthalocyanines has been synthesized. Their nonlinear optical properties have been measured. The physical properties of all the phthalocyanines synthesized in this work are subject to both acid and solvent effects.

INTRODUCTION

The first synthesis of phthalocyanine was reported in 1907 by Braun and Tcherniac when they heated o-cyanobenzamide at a high temperature to form an unsubstituted metal-free phthalocyanine.[2] Traditionally, phthalocyanines have been used as dyes and pigments in industry. Recently, Phthalocyanines have been of particular interest in many fields such as nonlinear optics,[3,4] Langmuir-Blodgett films,[5] liquid crystals,[6] optical data storage,[7] gas sensors,[8] electrophotography,[9] photosensitizers,[10] rectifying devices, [11] low dimensional metals,[12] and electrochromism.[13]

Several phthalocyanines have been synthesized and tested for third-order nonlinearity. For example, Ho, Ju, and Hetherington[14] have studied unsubstituted gallium and aluminum phthalocyanines. Shirk, et. al. 15 have studied metal-free and platinum derivatives of tetrakis-cumylphenoxy phthalocyanines.

Matsuda, et. al. [16] have studied alkylthiol derivatives of copper phthalocyanine with alkyl chain lengths of four to twelve carbon atoms. As expected, the compounds with the longer chains were soluble in organic solvents, while the short chain compounds were insoluble. Alkylthio groups were observed to increase the THG except when the chain was excessively long. The largest value of $\chi^{(3)}$ was observed at wavelengths of 1.5 and 1.9 μ for the chain length of ten carbons or at wavelength of 2.1 μ for eight carbons. Matsuda and co-workers also observed that molecules containing a permanent dipole caused by an axial ligand tended to have larger THG.

In an unrelated study, Shelnutt and co-workers investigated the use of selected porphyrin films as catalysts. 17 These authors did extensive theoretical modeling in addition to experimental work on film formation. They have done some very interesting studies with these compounds using long chain dialkyl amines with the porphyrin molecules in films. Although the spectroscopic properties of these compounds were determined, their nonlinear optical properties have not been studied. The Langmuir-Blodgett films made with these compounds should be similar to those of the phthalocyanines.

In all of these studies, there has been no systematic investigation of the effect of substituents on the nonlinear optical properties of either the bulk phthalocyanines or the films. In the present work, twenty phthalocyanines with both short and long chain substituents were synthesized. Their nonlinear optical properties have been measured. Nonlinear optical properties for eight of the phthalocyanines have been

reported. [18] Complete nonlinear optical result including these twenty phthalocyanines will be published.[19]

SYNTHESIS

1,4,8,11,15,1 8,22,25 -octasubstituted metal-free and metal phthalocyanines and 2,3,9,10,16,17,23,24-octasubstituted metal phthalocyanines were synthesized by modification of the procedures[20,21] shown in Figure 1. Route 1 was used for 1,4,8,11,15, 18,22,25-octaalkoxy metal phthalocyanines and route 2 and 3 were used for 2,3,9,10,16,1 7,23,24-octaalkoxy metal phthalocyanines. The 20 phthalocyanines that have been synthesized are shown in Table 1.

TABLE 1. The List of Phthalocyanines Synthesized.

#	Phthalocyanine
1	1,4,8,11,15, 18,22,25-octabutoxy metal-free phthalocyanine
2	1,4,8,11,15, 18,22,25-octaoctyloxy metal-free phthalocyanine
3	1,4,8,11,15, 18,22,25-octadecyloxy metal-free phthalocyanine
4	1,4,8,11,15, 18,22,25-octabutoxy copper phthalocyanine
5	1,4,8,11,15, 18,22,25-octaoctyloxy copper phthalocyanine
6	1,4,8,11,15, 18,22,25-octadecyloxy copper phthalocyanine
7	1,4,8,11,15, 18,22,25-octahexadecyloxy copper phthalocyanine
8	1,4,8,11,15, 18,22,25-octabutoxy zinc phthalocyanine
9	1,4,8,11,15, 18,22,25-octadecyloxy zinc phthalocyanine
10	1,4,8,11,15, 18,22,25-octabutoxy nickel phthalocyanine
11	1,4,8,11,15, 18,22,25-octaoctyloxy nickel phthalocyanine
12	1,4,8,11,15, 18,22,25-octadecyloxy nickel phthalocyanine
13	1,4,8,11,15, 18,22,25-octahexadecyloxy nickel phthalocyanine
14	1,4,8,11,15, 18,22,25-octabutoxy cobalt phthalocyanine
15	1,4,8,11,15, 18,22,25-octabutoxy palladium phthalocyanine
16	1,4,8,11,15, 18,22,25-octadecyloxy palladium phthalocyanine
17	2,3,9,10,16, 17,23,24-octadodecyloxy nickel phthalocyanine
18	2,3,9,10,16,17,23 ,24-octadodecyloxy copper phthalocyanine
19	2,3,9,10,16, 17,23,24-octadodecyloxy zinc phthalocyanine
20	2,3,9,10,16, 17,23,24-octadodecyloxy nickel phthalocyanine

ACID AND SOLVENT EFFECTS ON PHTHALOCYANINE SPECTRA

The spectroscopic properties of twenty octaalkoxy metal-free and metal phthalocyanines synthesized in this work were all subject to both acid and solvent effects. For example:

a) 1.4.8,11.15. 18,22,25-Octadecyloxy Copper Phthalocyanine. This compound was a sticky semi-solid at room temperature, An ethanol or acetone solution of this compound was green. Its UV-VIS spectrum in ethanol is shown in Figure 2. When several drops of acetic acid were added to the green solution (to give a concentration of acetic acid in the solution of 3. 00/0), the color changed to brown, and the UV-VIS spectrum did not match the typical copper Phthalocyanine spectrum (green). When several drops of the brown solution were added to ethanol, the color remained brown. However, when several drops of the brown solution were added to acetone, the color changed back to green, and the UV-VIS spectrum of the green solution was that of the typical copper phthalocyanine. This change was reversible. When chloroacetic acid was added to the green solution of the phthalocyanine dissolved in ethanol (to give a concentration of chloroacetic acid in the solution of 1.80/0), the color changed to brown. The UV-VIS spectrum of the brown solution is shown in Figure 3. This change was also reversible in acetone. Trichloroacetic acid produced an irreversible change. The UV-VIS spectrum obtained by adding a drop of the green acetone solution of the phthalocyanine to water was different from the typical UV-VIS spectrum in ethanol. A pH study using aqueous buffer solutions showed

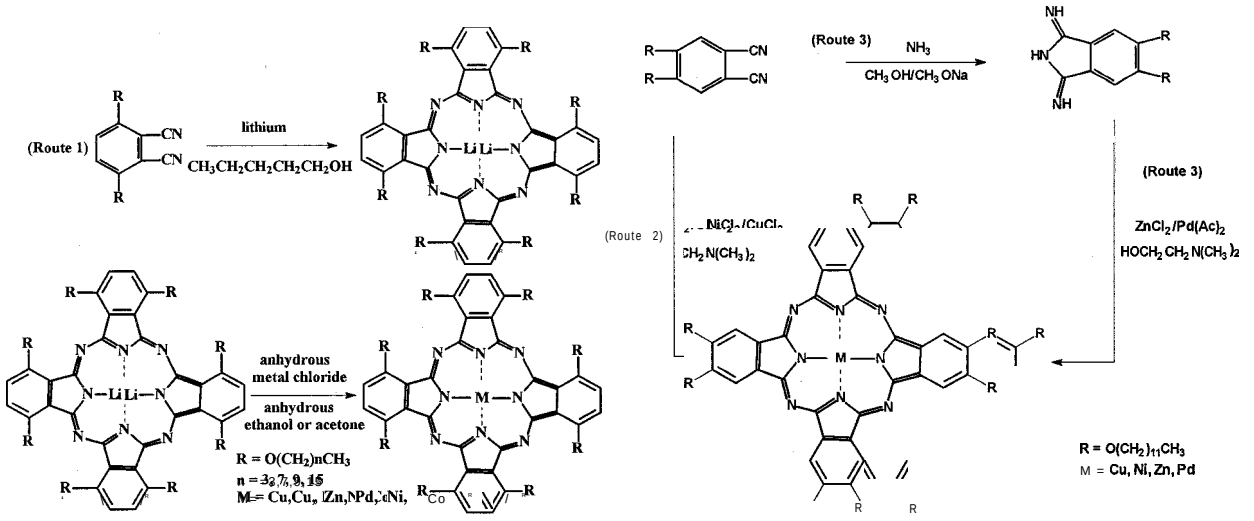


FIGURE 1. Synthetic Routes.

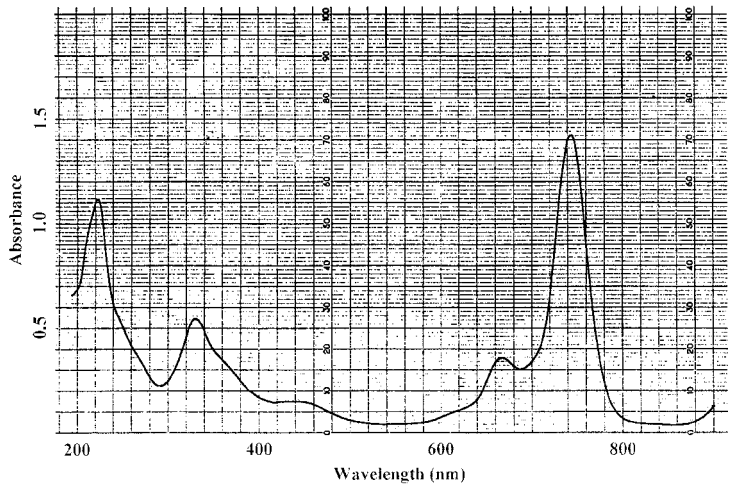


FIGURE 2. UV-VIS Spectrum Of 1,4,8,11,15,18,22,25-Octadecyloxy Copper Phthalocyanine In Ethanol

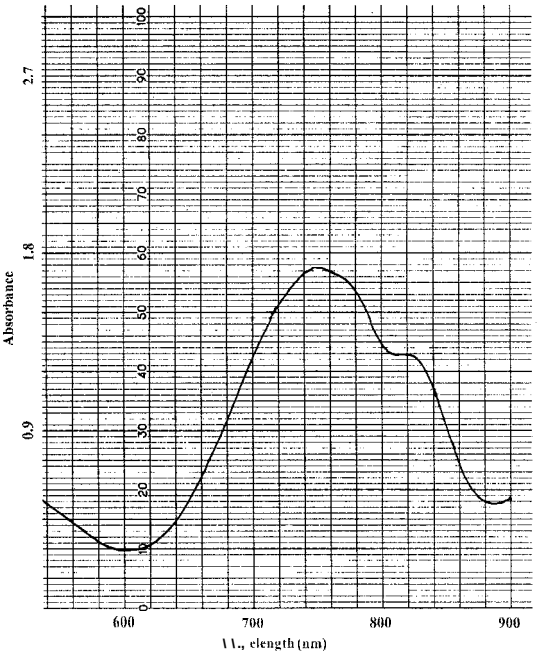


FIGURE 3. UV-VIS Spectrum Of 1,4,8,11,15,18,22,25-Octadecyloxy Copper Phthalocyanine In Ethanol (With 1.8 % Chloroacetic Acid)

that the color of this substance was green when pH > 6 and blue black when pH < 6.

b) 1.4.8,11,15, 18,22,25-Octadecyloxy Palladium Phthalocyanine. This compound was a viscous liquid at room temperature. This compound exhibited a normal UV-VIS spectrum (green) in acetic acid and chloroacetic acid, but it changed color to brown in trichloroacetic acid. When this compound was dissolved in chloroform, the color was green, and the UV-VIS spectrum was that of the typical palladium phthalocyanine spectrum. However, when this compound was dissolved in methylene chloride, the color was pink, and the UV-VIS spectrum changed. When several drops of this pink solution was added to ethanol, the color remained pink. However, when several drops of this pink solution was added to acetone, the color changed back to green, and the UV-VIS spectrum of this solution was that of the typical palladium phthalocyanine. This change was reversible.

C) 1.4.8.11.15. 18,22,25-Octadecyloxy Zinc Phthalocyanine. This compound was a green sticky semi-solid. An ethanol solution of this compound was green. When several drops of acetic acid were added to this green solution, the color changed to brown, and peaks in the visible region of the UV-VIS spectrum changed. When this compound was dissolved in chloroform, the color was yellow, and when this compound was dissolved in methylene chloride, the color was blue. In both of these cases, the peaks in the visible region significantly changed and were shifted to longer wavelength.

D) 2.3.9.10.16. 17,23,24-Octadodecyloxy Zinc Phthalocyanine. This compound was a green solid. A chloroform solution of this compound was green. When several drops of acetic acid were added to this green solution, the color remained green, and the peaks in the visible region changed very slightly. When several drops of hydrochloric acid were added to the original green solution, the peaks in the visible region changed even more than in the case of the acetic acid addition, even though the color remained green. When this compound was dissolved in methylene chloride, the peaks in the visible region changed. However, this change was slight, and the color remained green.

In summary, the twenty phthalocyanines synthesized are all subject to acid and solvent effects. However, the phenomena for 2,3,9,10,16, 17,23,24-octaalkoxy metal phthalocyanines are not as pronounced as those for 1,4,8,11,15, 18,22,25-octaalkoxy metal phthalocyanines. Also, 2,3,9,10,16, 17,23,24-octaalkoxy metal phthalocyanines tend to be more stable toward acid and solvent changes. In view of their particularly interesting properties, they may find special applications such as sensors, pH indicator and light limiting devices. Further investigation of acid and solvent effects is currently in progress.

EXPERIMENTAL

All chemicals used in this study were purchased from Aldrich Chemical Company or Fisher Scientific Products unless otherwise indicated and were used without further purification. UV-VIS spectra were determined using a PERKIN-ELMER Model 553 Fast Scan UV-VIS Spectrophotometer or HEWLETT PACKARD Model HP 8453 W-Visible Spectrophotometer. NMR spectra were obtained using a BRUKER MODEL AC -300 MHZ Spectrometer. UNISOL (Norell Inc.) was used as the NMR solvent unless otherwise indicated.

Representative Synthetic Procedures:

A. Preparation of 1,4,8,11,15,18,22,25-Octabutoxy Metal-Free Phthalocyanine. 3,6-Dibutoxy- 1,2-dicyanobenzene (1.7 g, 6.3 mmol) and 1-pentanol (15 ml) were placed in a 125 ml Erlenmeyer flask with a ground glass joint. Lithium metal (1.0 g, 144.1 mmol) was added in small portions to prevent overheating. After all the lithium was dissolved, a condenser was added and the mixture was heated to reflux for one hour using a hot plate to form a solution of 1,4,8,11,15, 18,22,25-octabutoxy dilithium phthalocyanine. The reaction product was deep green, and its UV-VIS spectrum showed a characteristic visible peak at 717 nm. After cooling, the reaction mixture containing 1,4,8,11,15, 18,22,25-octabutoxy dilithium phthalocyanine was dissolved in 80 ml of acetone, the solution was filtered and 20 ml of acetic acid was added. The color of the solution changed from deep green to brown green. The solvents were removed by rotary evaporation. The residue was extracted first with ethyl acetate-water and then with chloroform-water. The impurities remained in the water layer, and the product was in the ethyl acetate and chloroform layer. Finally, after the solvent was evaporated, the crude product was chromatographed two times on a silica gel column (using chloroform as a solvent to prepare the column and chloroform-ethyl acetate (10:1) as the eluant) to give a brown green solid 1,4,8,11,15,1 8,22,25 -octabutoxy metal-free phthalocyanine. Yield: 0.61 g (35.4 %). Melting point: 125.0-126.0 'C.

The UV-VIS spectrum of the product m chloroform showed absorption at 774 nm, 753 nm, 677 nm, 733 nm, 242 nm.

^{13}C NMR spectrum (solvent: d-chloroform): Peaks were observed at δ (ppm) 14.00, 19.29, 31.42, 71.21, 117.18, 126.49, 151.06, 151.06.

B. Preparation Of 1,4,8,1 1,15,18,22,25-Octabutoxy Copper Phthalocyanine. Reacting 3,6-Dibutoxy-1,2-dicyanobenzene (2.0 g, 7.4 mmol), 1 -pentanol (20 ml) and Lithium (1.2 g, 172.9 mmol) to forma solution of 1,4,8,11,15, 18,22,25-octabutoxy dilithium phthalocyanine. The reaction mixture containing 1,4,8,11,15, -18,22,25-octabutoxy dilithium phthalocyanine was dissolved in 20 ml of anhydrous ethanol. A solution of 0.5 g (3.7 mmol) of anhydrous copper(II) chloride dissolved in 20 ml of anhydrous ethanol was added . The mixture was refluxed for one hour to form a suspension of 1,4,8,11,15, 18,22,25-octabutoxy copper phthalocyanine. The reaction mixture was dissolved in ethyl acetate (80 ml), and the solution was filtered, after which the solvents were removed by rotary evaporation. The crude product was extracted first with ethyl acetate-water and then with chloroform-water. The impurities remained in the water layer, and the product was in the ethyl acetate and chloroform layer. After the solvent was evaporated, the crude product was chromatographed two times on a silica gel column (using chloroform as a solvent to prepare the column and chloroform-ethyl acetate (2: 1) as the eluant) to give a black green solid 1,4,8,11,15, 18,22,25-octabutoxy copper phthalocyanine (0.93 g). Yield: 43.6%. Melting point: 79.0 -81.0 'C.

The UV-VIS spectrum of the product in ethanol showed absorption at 744 nm (ε: 10.52 x 104), 668 nm (ε: 2.63 x 104), 329 nm (ε: 4.02 x 104), 223 (ε: 8.26 x 104).

Quantitative determination of Cu by atomic absorption spectroscopy gave 5.7 % (required: 5,5 %).

^{13}C NMR spectrum: Peaks were observed at δ (ppm) 12.85, 17.93, 29.75, 68.40, 118.30, 120.20, 148.77, 166.30.

C. **Preparation Of 1,4,8,1 1,15,18,22,25-Octabutoxy Palladium Phthalocyanine.** Reacting 3,6-Dibutoxy-1,2-dicyanobenzene (1.5 g, 5.5 mmol),1-pentanol (15 ml) and Lithium (1.0 g, 144.1 mmol) to form a solution of 1,4,8,11,15, 18,22,25-octabutoxy dilithium phthalocyanine. The reaction mixture containing 1,4,8,11,15, -18,22,25-octabutoxy dilithium phthalocyanine was dissolved in 20 ml of anhydrous ethanol. A solution of anhydrous palladium chloride dissolved in 35 ml of anhydrous acetone was added. This solution was prepared by Soxhlet extraction using 1.3 g (7.3 mmol) of anhydrous palladium chloride and 35 ml of anhydrous acetone. The extraction was continued for forty five hours to give a dark brown solution of the palladium chloride dissolved in acetone. At the end of the Soxhlet extraction, the palladium chloride in the porous paper thimble was not all dissolved.

The solution of anhydrous palladium chloride dissolved in anhydrous acetone was added to the solution of dilithium phthalocyanine, and the mixture was refluxed for fifteen hours to form a suspension of 1,4,8,11,15, 18,22,25-octabutoxy palladium phthalocyanine. The reaction mixture was dissolved in acetone (80 ml), the solution was filtered, and the solvents were removed by rotary evaporation. The crude product was extracted first with ethyl acetate-water and then with chloroform-water. The impurities remained in the water layer, and the product was in the ethyl acetate and chloroform layer. After the solvent was evaporated, the crude product was chromatographed three times on a silica gel column (using chloroform as a solvent to prepare the column and chloroform as the eluant) to give a purple green liquid 1,4,8,11,15,18,22,25-octabutoxy palladium phthalocyanine (0.64 g). Yield: 38.9 %. Melting point: Below -11 'C.

The UV-VIS spectrum of the product in chloroform showed absorption at 741 nm (ε: 2.88 x 104), 660 nm (ε: 0.70 x 104), 328 nm (ε: 1.75 x 104), 252 nm (s: 2.68 x 104).

^{13}C NMR spectrum: Peaks were observed at δ (ppm) 13.03, 18.20, 30.34, 70.03, 116.27, 125.55, 141.15, 149.65.

D. Preparation Of 2,3,9,10,16,17,23,24-Octadodecyloxy Nickel Phthalocyanine. 4,5-Didodecyloxy-1,2-dicyanobenzene (2.0 g, 4.0 mmol), N,N-dimethylethanolamine (10 ml), nickel chloride (0.20 g, 1.5 mmol) and 1,8-diazabicyclo [5,4,0] undec-7-ene (DBU) (7 drops) were placed in a 25 ml angled 2 necked round bottom flask. A condenser was connected to the top of the flask, a heating mantle was used to heat the flask, and a Variac transformer was used to adjust the voltage. The mixture was heated to reflux under nitrogen for 16 hours. After cooling, the mixture was poured into a mixture of H_2O/MeOH (1: 1) and filtered. The collected material was washed with hot ethanol several times to remove some polar impurities. Then, more polar impurities in the material were removed by Soxhlet extraction using ethanol as an extracting solvent. Finally, the material remaining in the porous thimble of the Soxhlet extractor chamber was dissolved in chloroform

using the Soxhlet extracting process. After evaporation of the solvent, a green solid (2,3,9,10,16,17,23,24-octadodecyloxy nickel phthalocyanine) was formed. Yield 0.62 g (30.1 %). Melting point: 114.0 -115.0 'C.

The UV-VIS spectrum of the product in chloroform showed absorption at 672 nm, 606 nm, 410 nm, 312 nm, 291 nm.

"C NMR spectrum (d-chloroform): Peaks were observed at δ (ppm) 14.09,22.64,25.87,28.85, 29.31, 29.31,29.31,29.31,29.31,29.31, 29.59,31.88,69.55, 106.32, 125.54, 153.88, 168.29.

ACKNOWLEDGMENTS

This work was supported in part by NASA Grant No. NAG8-1057 and No. NAG8-1288 and in part by NASA-IRA Grant No. NAGW-4078.

Thanks are given to Tiffany Myers for making 3,6-dioctyloxy-1,2-dicyanobenzene and 3,6-didecyloxy-1,2-dicyanobenzene and to Angela Romero for making 4,5-didodecyloxy-1,2-dicyanobenzene.

REFERENCES

1. a) Yan, W.; Zhou, Y.; Wang, X.; Chen, W.; Xi, S. *J.Chem. Sot., Chem.Commum., 1992,873.* b)Casstevens, M. K.; Samoc, M.; Pfleger, J.; Prasad, P. N. *J.Chem.Phys. 1990,92 (3), 2019.*
2. Braun, A.; Tcherniac, J. *Ber. Deut.Chem.Ges. 1907,40,2709.*
3. Casstevens, M. K.; Samoc, M.; Pfleger, J.; Prasad, P. N. *J.Chem.Phys. 1990,92,2019.*
4. Hosoda, M.; Wada, T.; Garito, A. F.; Sasabe, H. *Jpn.J.Appl.Phys. 1992,31,* L249.
5. Roberts, G. G.; Petty, M. C.; Baker, S.; Fowler, M. T.; Thomas, N. *J.Thin Solid Films 1985,132,* 113.
6. Simon, J.; Sirlin, C. *Pure Appl.Chem. 1989,61,* 1625.
7. Kuder, J. E. *J.Imag.Sci. 1988,32, 51.*
8. Collins, R. A.; Mohamed, K. A. *J.Phys. D 1988,21, 154.*
9. Loutfy, R. O.; Her, A. M.; Hsiao, C. K.; Baranyi, G.; Kazmaier, P. *Pure Appl.Chem. 1988,60,* 1047.
10. Kate, M.; Nishioka, Y.; Kaifu, K.; Kawamura, K.; Ohno, S. *Appl.Phys.Lett. 1985,46, 196,*
11, Abe, K.; Sate, H.; Kimura, T.; Ohkatsu, Y.; Kusano, T. *Makromol.Chem. 1989, 190,2693.*
12. Marks, T. *J. Science 1985,227, 881.*
13. Corker, G.; Grant, B.; Clecak, C. *J.Electrochem.Sot. 1979,126,* 1339.
14. Ho, Z. Z.; Ju, C. Y.; Hetherington, W. M. *J.Appl.Phys. 1987, 62(2), 716.*
15. Shirk, J. S.; Lindle, J. R.; Bartoli, F. J.; Hoffman, C. A.; Kafafi, Z. H.; Snow, A. W. *Appl.Phys.Lett. 1989,55, 1287.*
16. Matsuda, H.; Ohada, S.; Masaki, A.; Nakanishi, H.; Suds, Y.; Shigehara, K.; Y. Yamada. *SPIE Vol 1337 Nonlinear Optical Properties of Organic Materials III,* 1990,105.
17. Shelnutt, J. A.; Medforth, C. J.; Berber, M. D.; Barkigia, K. M.; Smith, K. M. *J. Am. Chem. Sot. 1991, 113,4077.*
18. Sanghadasa, M.; Wu, B.; Clark, R. D.; Guo, H.; Penn, B. G. "Third Order Nonlinear Optical Properties of Metal Phthalocyanines with Butoxy and Decyloxy Substituents". Nonlinear Optical Properties of Organic Materials X, Proc.SPIE 1997,3147.
19. Sanghadasa, M.; Wu, B.; Clark, R. D.; Guo, H.; Penn. B. G., in preparation.
20. (a) Barrett, P. A.; Frye, D. A.; Linstead, R. P. *J.Chem. Sot. 1938,* 1157. (b) McKeown, N. B.; Chambrier, I.; Cook, M. J. *J.Chem. Sot, Perkin Trans. 11990,* 1169.
21. Haisch, P.; Hanack, M. *Synthesis, 1995, 1251.*

Raman and Photoluminescence Spectroscopy of Nano-crystalline $PbTiO_3$ Sensor Materials with Different Doping Ions

R. S. Katiyar and Meng Jinfang

Department of Physics, University of Puerto Rico, San Juan, Puerto Rico 00931-3343, USA

ABSTRACT

Raman spectra & photoluminescence studies in $PbTiO_3$, have been carried out, as a function of particle size, temperature, pressure and dopants. There appears respectively a distinct temperature-induced soft mode phase transition in each sample whose Curie temperature can be determined from the mean-field theory. The detailed Curie temperature shift in the modified $PbTiO_3$ ceramics by Ba, Sr, La and Zr, has been investigated as a function of particle size. Pressure-induced phase transitions display an obvious diffuse behavior. Room temperature photoluminescence for nanocrystalline $Ba_{1-x}Pb_xTiO_3$ have been observed. These studies favor preparations of high efficiency $PbTiO_3$ sensors.

INTRODUCTION

It is well known that pyroelectric infrared sensors show good sensing characteristics. Differing from photon sensors, they can be used at room temperature and their IR response does not depend on the wavelength of the IR radiation used. By combining these promising pyroelectric effects with sophisticated silicon IC technology, a realization of silicon monolithic IR sensor at near ambient temperature could be expected for a wide variety of applications. Nowadays, various kinds of pyroelectric materials like $LiTaO_3$ or TGS single crystals are commercially available. Among these materials, $PbTiO_3$ is a good candidate for IR detector because it has a large spontaneous polarization $P_s(75\mu C/cm^2)$ and a small dielectric constant (~ 100) along the polarization axis 1.$PbTiO_3$ is a perovskite type of ferroelectric isomorphous with $BaTiO_3$ and its tetragonality(c/a=1.06) is also very large [2]. But the Curie temperature T_c of $PbTiO_3$ single crystal is quite high, around 490 'C. For better sensitivity of detectors, the Curie temperature must be close to the operating temperature, In this sense, if the Curie temperature of $PbTiO_3$ could be reduced to about 160 "C by ionic substitution, its pyroelectric efficiency at room temperature can be strongly influenced. Obviously, the decrease of the Curie temperature T_c of $PbTiO_3$ can be realized by the substitution of of Pb or Ti ions. On the other hand, effects of particle size on physical properties and the Curie temperature T_c of materials have been widely received an increasing interest and attention[3-9]. For example, investigations for $PbTiO_3$ show that its Curie temperature can be lowered to room temperature as particle size is decreased to 12.6-20 nm. In this paper, we have made a detailed study for the modified $PbTiO_3$ by Ba, Sr, La and Zr as a function of particle size using Raman scattering spectroscopy. At the same time, $Pb_{0.8}Ba_{0.2}TiO_3$ and Pb0.8Sr0.2TiO3 with grain size of 60 nm are obtained in order to effects of

grain size on phase transitions. Photoluminescence for nanocrystalline is also important subject due to the usage of light emission devices. We perform photoluminescence study for nanocrystalline Pb 1 -xBaxTiO3 system at room temperature.

II. EXPERIMENTAL

Nanocrystalline $Pb_{0.8}Ba_{0.2}TiO_3$(PBT20), $Pb_{0.8}Sr_{0.2}TiO_3$(PST20), $Pb_{0.8}La_{0.2}TiO_3$(PLT20) and $PbTi_{0.75}Zr_{0.25}TiO_3$(PZT25) with different particle sizes were separately prepared, using a sol-gel process. Raman spectra and Photoluminescence spectra were collected using a Spex-1402 double-grating monochromator. A Mao-Bell type diamond anvil was employed to produce high pressure. A 16:3:1 methanol-ethanol-water mixture. The fluorescence of a small ruby chip was used for the pressure calibration.

III. RESULTS AND **DISCUSIONS**

Raman spectra of PBT20, PST20, PLT20 and PZT25 with particle sizes of 60 nm, 37 nm, 54 nm and 44 nm are respectively shown in Fig. 1a - d. These spectra reveal the tetragonal structure of nanocrystalline materials at room temperature, similar to one observed for bulk $PbTiO_3$(PT) 10 . From Fig. 1, it is found that the low frequency phonon mode of every sample, the E(TO) soft mode, shows a decrease in frequency, i.e. "softening", and widening in linewidth, on increasing temperature, as observed in pure bulk $PbTiO_3$. It should be noted that, because of a strong stray light in these materials, especially in PBT20, it is very difficult to measure the lowest frequency of the soft mode close to T_c. This is merely done by extrapolation of the observed data. Thus, it is reasonably considered that PBT20, PST20, PLT20 and PZT25 also undergo a similar temperature-induced soft mode as in pure $PbTiO_3$.

A detailed particle size dependence of the Curie temperature in the above mentioned materials has been performed and illustrated in Fig. 2a-d. One may notice that the transition temperature from cubic (CP) to tetragonal (TP) is shifted towards a lower temperature with decreasing particle size in nanocrystallines. A careful observation reveals that the particle size corresponding to a Curie temperature of 20 °C in PBT20 and PST20 is smaller than that in PLT20 and PZT25. The particle size dependence of Curie temperature in PZT25 is also found to have a slow change above the particle size of 40 nm, which differs from that in PBT20, PST20 and PLT20. This may originate from many factors, for example, different chemical properties of Ba, Sr, La and Zr as well as their different ion radii(Ba^{2+} -0.143 nm, Sr^{2+} ~ 0.127 nm, La^{3+} ~ 0.122 nm, Zr^{4+} - 0.079 nm) will lead to various distortions of TiO_6 octahedra thereby changing the short-range forces responsible for the phase transitions. In addition, the particle size dependence of T_c may result from different substitution forms of Ba, Sr and La ions for Pb ions at A sites, and Zr for Ti ions at B sites in ABO_3 perovskite oxides.

Fig. 3a-b displays Raman spectra for $Pb_{0.8}Sr_{0.2}TiO_3$ and $Pb_{0.8}Ba_{0.2}TiO_3$with grain size of 60 nm. The results reveals that the soft mode has an obvious "softening" with increasing pressure. And its intensity is reduced rapidly. The coupling of $E(TO_3)$ and $A_1(TO_3)$ is also observed. It is interesting to find that all phonon modes exhibit very weak intensity at high pressure and approach disappearance. Thus, it is thought that a pressure-induced phase transition occurs respectively at 7.84 GPa for $Pb_{0.8}Ba_{0.2}TiO_3$ and 7.12 GPa for $Pb_{0.8}Sr_{0.2}TiO_3$.

Obviously, their phase pressure are lower that that of bulk $PbTiO_3$, which results from effects of dopants and grain size on the phase transition pressure. However, the surface stress caused by small grain size is a main reason of the decrease of phase transition pressures.

Photoluminescence spectra for $Ba_{1-x}Pb_xTiO_3$ with an average grain size of 60 nm, and with x = 0.0, 0.1, 1.0, are indicated in Fig. 3c. At x = O (BT), a very broad structureless band appears at -2.12 eV and ~ 2.0 eV for the excitation lines of 488.0 and 514.5 nm. The dependence of the emission energy upon the Pb content is shown in Fig. 3d. The results indicate that the emission energy has a maximum blue shift ΔE -0.18 eV and 0.03 eV for the excitation lines of 514.5 and 488.0 nm, respectively. Photoluminescence for BPT at room temperature results may be well explained by such a model, in which, electrons in the valence band of $BaTiO_3$, instead of getting excited to the conduction band directly, are first excited to these localized sensitizing centers in the band-gap. Because of the existence of a strong relaxation of excitons, the electron polarons interact with holes to form an intermediate state, self-trapped excitons (STEs), either immediately or after being trapped for a certain time.

IV. CONCLUSIONS

High temperature Raman spectra for PBT20, PST20, PLT20 and PZT25 with various particle sizes have been measured. The Curie temperature is found to shift towards a lower temperature with decreasing particle size and it also depends upon the substitution forms of doping ions for the ions at A and B sites in ABO_3 perovskite materials. The phase transition pressures for PBT20 and PST20 with grain size of 60 nm are found to decrease with respect to the bulk PbTiO3. The surface stress from small particles plays an important role in lowering Curie temperatures and the phase transition pressures. The visible emission for the nanocrystalline PBT system at room temperature can result from effects of surface states in the band gap.

ACKNOWLEDGMENTS

The work was supported by the NASA-NCCW-0088 and NSF-OSR-9452893.

REFERENCE

1. V. G. Gaurilyachenko, R. I. Spinko, et al., *Sov. Phys. Solid State 12, 1203(1970).*
2. G. Shirane, R. Pepinsky and B. C. Frazer, *Act Crystallogr. 9,* 13 1(1956).
3. K. K. Deb, *Ferroelectrics, 82, 45(1988).*
4. Soma Chattopadhyay, Pushan Ayyub, V. R. Palkar et al., *Phys. Rev.* **B52,** 13 177(1995).
5. Jinfanf Meng, Guangtian Zou et al., *J. Phys.: Condens Matter, 6, 6543(1 994).*
6. R. Bachmann and K. Baerner, *Solid State Commun. 68(9),865(1988).*
7. K. Uchino, E. Sadanaga, T. Oonish, T. Morohashi, H. Yamamura, *Transactions. 8, 107(1990).*
8. T. Yamamoto, K. Urabe and H. Banno, *Japanese Journal of Applied Physics, 32, 4272(1993).*
9. Stefan Schlag and Hans-Friedrich Eicke, *Solid State Commun. 91, 883(1994).*
10. R. J. Nelmes and W. F. Kuhs, *Solid State Commun. 54, 721(1985).*

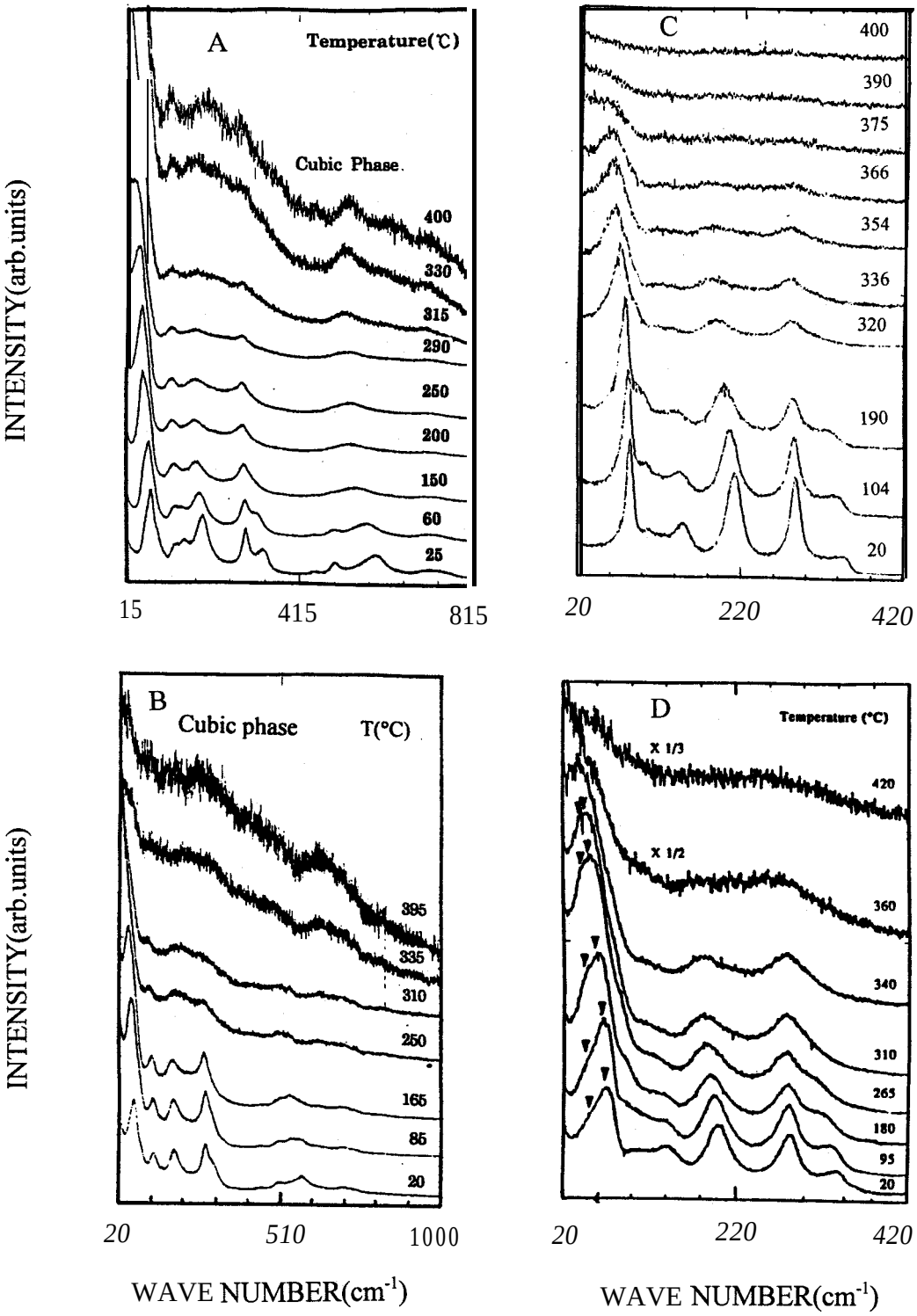


Fig. 1 Raman spectra of nanocrystalline materials at various temperatures
A. $Pb_{0.8}Ba_{0.2}TiO_3$ B. $Pb_{0.8}Sr_{0.2}TiO_3$
C. $Pb_{0.8}La_{0.2}TiO_3$ D. $PbTi_{0.75}Zr_{0.25}O_3$

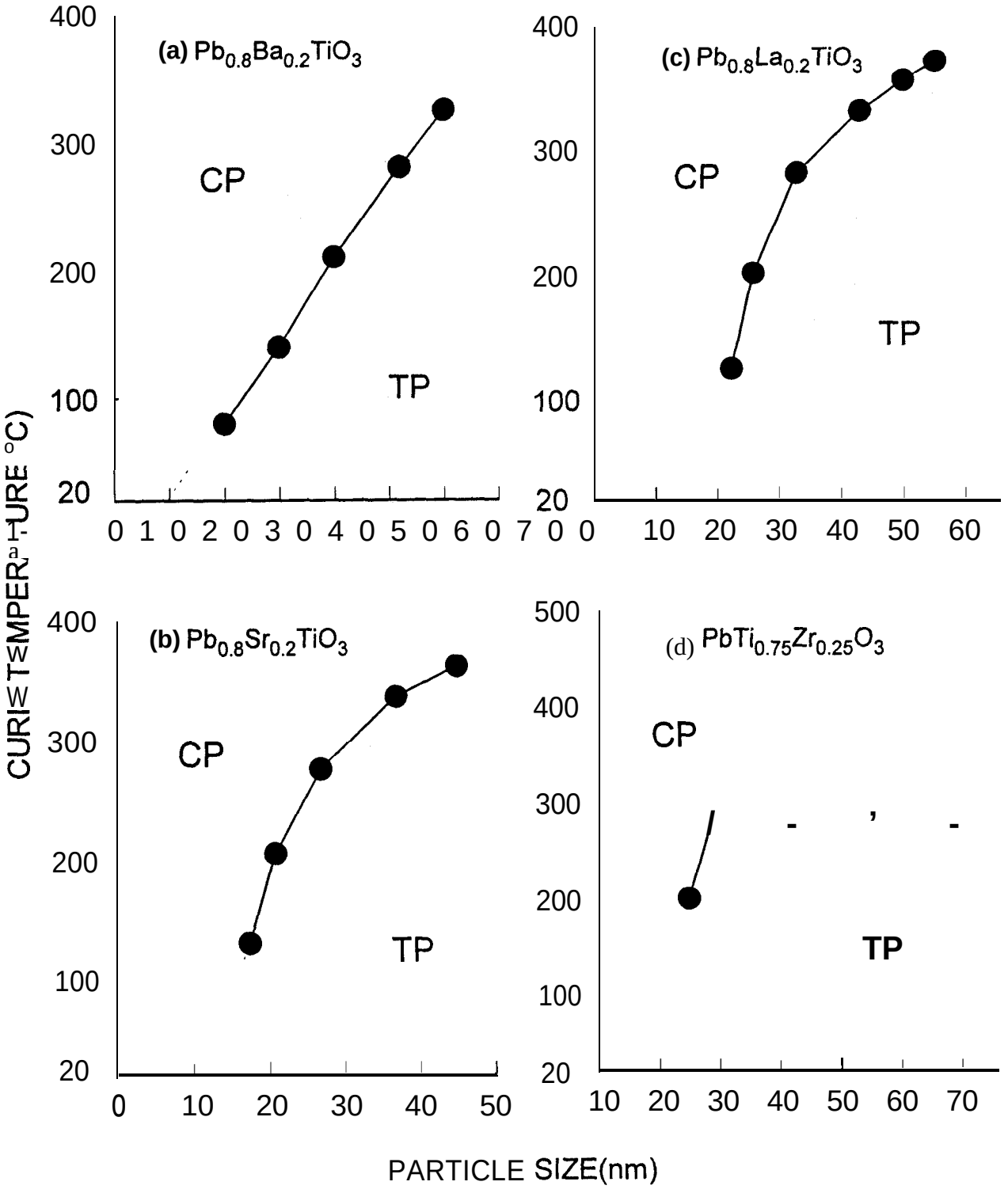
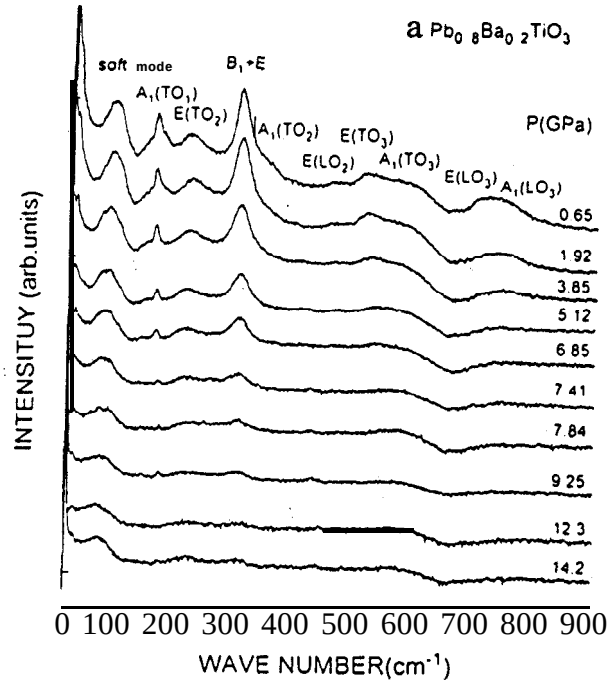
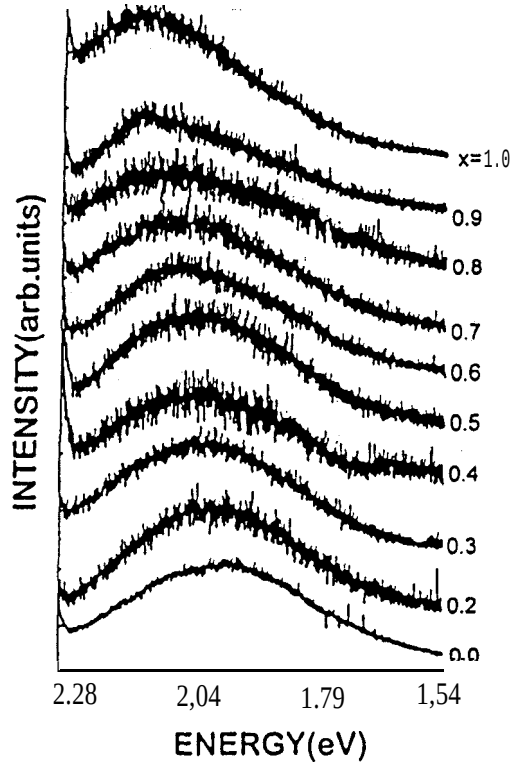
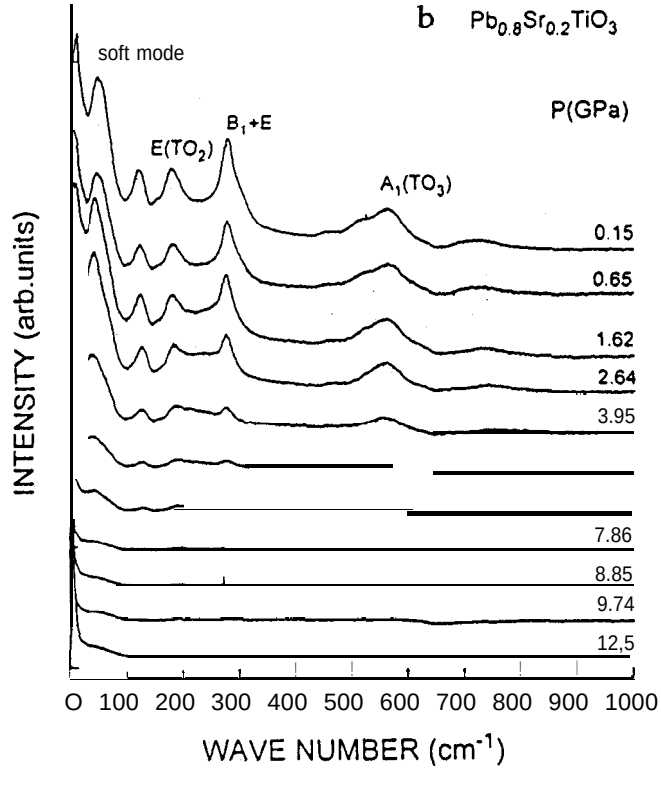


Fig.2 The particle size dependence of Curie temperature for nanocrystalline







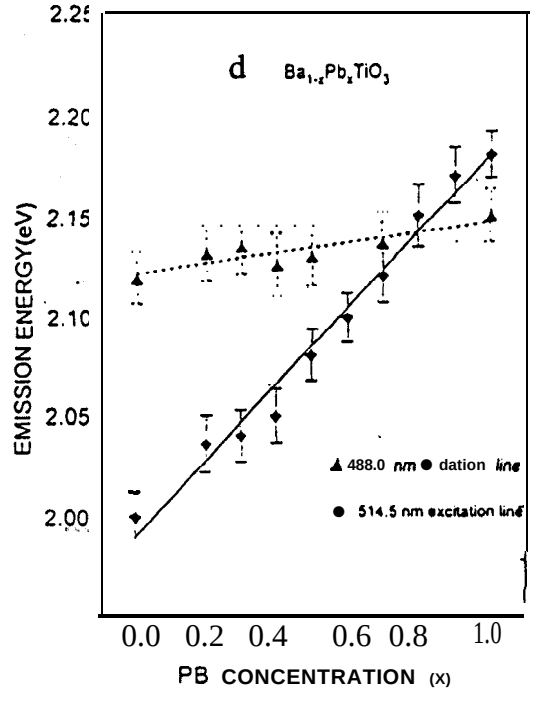


Fig.3a-b High pressure Raman spectra
a. $Pb_{0.8}Ba_{0.2}TiO_3$ (60 nm)
b. $Pb_{0.8}Sr_{0.2}TiO_3$ (60 nm)

Fig. 3c Photoluminescence for $Ba_{1-x}Pb_xTiO_3$
Fig. 3d The excitation line dependence of photoluminescence

Optical Properties of Phosphorus Implanted into Fused Silica

D. Denmark, T. Taylor, R. Mu, A. Ueda, Y. S. Tung, M.H. Wu and D.O. Henderson
Chemical Physics Laboratory, Department of Physics, Fisk University, Nashville, TN 37208
C.W. White, R. Zuhr and A. Meldrum
Solid State Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6057

Abstract

A series optical characterization of phosphorus ion implanted fused silica annealed at various temperatures with three different ion doses has been conducted. It is clearly demonstrated the thermal annealing effect on silica defects due to P ion implantation, formation of the P_2 molecules and the final assembly of the P semiconductor quantum dots in each ion dose. The study of the P implanted silica has not only provide a unique opportunity to understand ion - material interaction, thermal diffusion and other fundamental physics and chemistry, it also serve as a guidance for semiconductor quantum dot fabrication.

Introduction

The study of nanometer-sized semiconductor particles (called quantum dots) has drawn great attention both in fundamental and applied science arena. One of the most striking characteristics of these quantum dots is the quantum confinement effect when the photoinduced electron (e) - hole (h) pairs (excitons) are spatially confined within the particles. This effect will modify the bulk electronic band structure which leads to the alteration of the exciton properties of the quantum dots. The degree of the quantum confinement can be classified into three different regimes[1,2] by comparing the physical size of the quantum dot R with the bulk exciton Bohr radius a_B. The weak confinement regime occurs when $R >> a_B$, where the excitons in the particles experience moderate physical confinement. Therefore, the Coulomb interaction between the electron (e) and hole (h) is dominant. Strong confinement is considered when $R << a_B$. In this case, the Coulomb term turns out to be very small and can be ignored or treated as a perturbation. The electrons and holes can be treated as confined independent particles. The simple particles in a box concept can be readily applied. When $R \sim a_B$, the system is considered as the intermediate confinement, where both Coulomb and quantum confinement of the excitons are considered. Therefore, the study of how the electronic properties a) change as a function of crystal size and b) can be altered by the confining host can lead to new device development with different electronic and optical characteristics.

There have been a number of methods developed to fabricate quantum dots in various host materials and in solution phases, such as inorganic synthesis[3] sol-gel synthesis,[4, 5] impregnation into porous host,[6] vapor deposition,[7] and ion implantation[8] Although ion implantation can be considered as one of the most clean and flexible techniques to obtain the quantum dots in isolating host, the fabrication of the quantum dots has been limited to single element systems, such as Ge, Si, P, etc. Only recently, we have shown that sequential ion implantation followed by post-thermal annealing is a promising and effective technique to fabricate II - VI and III - V quantum dots in the dielectric hosts.[9]

On the other hand, some of the most challenging problems associated with binary and ternary quantum dot fabrication in dielectric host materials are i) to keep the chemical stoichiometry balanced; ii) to tailor a desired or well-defined surfaces and interfaces; and iii) to control the right particle size and size distribution. The solutions to these problems often rely on a fundamental understanding of

the physical and chemical properties of: *a)* each elemental material which constitute these compound semiconductor materials in the hosts; *b)* the host materials; and c) host-guest interactions at atomic and molecular level to clusters and nanoparticles. Phosphorus is a very important element not only due to its crucial role in forming valuable III-V compound semiconductors such as InP, GaP All? *etc.,* but also due to the fact that phosphorus itself is an interesting semiconductor material. In addition, it can form molecules and molecular clusters. Therefore, it is a very good candidate to understand a material evolution from single atom to molecules, clusters and semiconductor quantum dot formation. The phosphorus based glasses have also shown very intriguing optical response.[10]

Experimental

P ions with an energy of 70 keV were implanted into Dow Corning 7940 optical grade silica substrate. The implantation experiment was conducted at room temperature. Three different doses of P ions were implanted. They are $1x10^{16}$, $3x10^{16}$ and $6x10^{16}$ ions/cm^2 respectively. SRIM simulation suggests that the peak the ion concentration is located at 1100 Å with a Gaussian profile. The width of the implanted profile is about 100 nm.

In order to study the thermal annealing effect on phosphorus implanted silica, a step-wise (every 50ºC) thermal annealing process was conducted under 95% Ar and 5% H_2 environment. The starting temperature was 150ºC and atypical annealing time was 15 min. Upon each annealing, a set of UV-Vis and infrared spectra were collected to examine the changes of the electronic and vibrational properties of the samples with respect to the previous sets of spectra.

Optical measurements were made with Hitachi 3501 UV-Vis-NIR spectrophotometer in the range of 3200-185 nm in transmission mode. Atypical spectral resolution was 1 nm. Infrared spectra were obtained with a Bomem DA3 spectrometer with resolution of 4 cm-l. All the vibrational spectra were collected at room temperature with pressure p < 1 torr with 300 scans.

Results and Discussion

Figure 1 shows the UV-Vis transmission spectra of as-implanted samples with the doses of $1x10^{16}$, $3x10^{16}$ and $6x10^{16}$ ions/cm^2 together with the spectrum of a virgin silica substrate. Clearly, upon P ion implantation, the sample starts to absorb photons at the energy shorter than 400 nm (3.1 eV). The absorption onset moves to longer wavelength (i.e. redshift) as the ion dose increases. In addition, a band at about 245 nm (5.06 eV) is clearly observed for all four dosed samples. At high ion dose, the strong broad range absorption below 400 nm nearly wash out the band .Interestingly, the peak height of the band at 245 nm does not show any clearly trend of relating the ion dose. Figure 2 gives an illustrative example of the thermal annealing effect on optical spectra of the P implanted sample with the ion dose of $3x10^{16}$ ions/cm^2. There are three clear temperature ranges at which the optical transmission undergoes disrupt changes: *1)* At RT < T < 550ºC, optical transmission below 400 nm increases and the onset absorption blueshifts. The absorption intensity at 245 nm decreases and the band is disappeared at 550ºC; 2) at 650< T < 950ºC, a band at -228 nm (-5.4 eV) develops; 3) at T > 950ºC, abroad band absorption in the *W* region starts to increase and the onset absorption redshifts along with increase of temperature. At 1100ºC, the onset absorption is at -530 nm (2.3 eV). Figure 3 shows the difference spectra of the optical transmission with respect to the spectrum obtained at 550ºC annealing. The spectrum change in the three temperature ranges as described above can be clearly demonstrated. A band at 245 nm is annealed out at 550ºC and a new band developed at T > 650ºC. Abroad absorption band is formed in the *W* region and the onset of the band redshifts as temperature increases. Figure 4 illustrates the thermally annealed P implanted samples at 11 00ºC

with three different ion doses. As the ion dose increases, the absorption onset moves to the longer wavelength. At the highest dose, the absorption onset is near 550 nm (2.25 eV).

It is known that ion implantation in silica substrates 'can introduce various defects when the energetic ions are slowed down by the host materials. The study of these defects via EPR and other techniques has been reviewed by Weeks. [11] The band observed at 245 nm has been assigned to oxygen molecule ions[11] (B_2 centers) or recently to oxygen-deficient centers (ODC) z. e. oxygen vacancies[12] as the result of ion implantation. Therefore, the presently observed band at ~245 nm in fugure 1 must be due to the B_2 centers created during the ion implantation. As illustrated in figures 2 and 3, this band can be annealed out at 550°C, which is also consistent with the results reported in literature. It is also known that ion implantation can create various defects in glasses as discussed by Weeks[11]. The broad absorption in the *W* region for the as-implanted samples are due to these defects. As the annealing temperature increases, these defects are gradually annealed out. Therefore, the gradual transmission improvement observed in our temperature dependent annealing is attributed to the decreases of the defect centers. When the annealing temperature is above 600°C, as shown in figures 2 &3, a new band is developed. It is argued that this band maybe due to P_2 molecule formation in silica. Hosono *et al.* [13] have studied P ion implanted silica at relatively high dose. When the dose > $2x10^{17}$ ions/cm^2, P ions can form nanoparticles in silica without thermal annealing. When the sample was heated up to 600°C, the band is observed at 5 eV, which they have attributed this band to P_2 molecules. The formation of the P_2 molecules is the result of evaporation of the amorphous red phosphorus nanoparticles. At lower dose as the present investigation, absorption spectra of the P_2 molecules are not observed in the as-implanted sample suggesting that the ion concentration is not high enough for P ions to form sufficient number of P_2 molecules in the as-implanted silica to detect by the optical spectrophotometer. At high temperature, the thermal energy is high enough to mobilize the P ions in silica so that P_2 molecules can be formed. It is also argued that these P_2 molecules or clusters are trapped in the silica host. And the thermal energy at T <950°C is not high enough to allow P_2 to diffuse in the appreciable rate so that a large number of nanoparticles can be formed. On the other hand, the observed peak position at 228 nm (5.4 eV) has relatively higher energy than what Hosono[13] and Herzberg[14] reported at 5.03 eV. It is also known that P atom has absorption at -218 nm (5.7 eV). However, the high temperature thermal annealing should not favor P atom formation by considering phosphorus is a relatively reactive element. Therefore, the blueshift in absorption energy may be associated with the trapping sites. When the temperature T >950°C, a broad band absorption, which is referred to the absorption band edge of the phosphorus quantum dots later on, is observed indicating the formation of the phosphorus semiconductor quantum dots. It is reported that the phosphorus quantum dots are in the amorphous red phosphorus state since the crystalline red phosphorus requires very slow annealing and it is also confirmed an amorphous red phosphorus nanoparticles are indeed formed for the high dose (>$2x10^{17}$ ions/cm^2) P implanted silica. 13

As illustrated in figure 4, the onset phosphorus absorption band edge shows a systematic redshift as the ion dose increases suggesting that the bigger quantum dots are being formed. On the other hand, the onset absorption edge of the phosphorus formed in the silica host is consistently blueshifted with respect to the bulk amorphous red phosphorus (E_g -2.1 eV, λ = 590 rim), which is due to the quantum confinement of the phosphorus quantum dots confined in a dielectric host proposed by Brus and others.[1,2] Both TEM and XRD will be conducted to further confirm the quantum dot formation in silica host as well as their particle size with different dose.

Conclusion

Relatively low dose P ion implanted silica with three different doese has been investigated via optical spectroscopy. It is demonstrated that the ion induced defects in silica can be effectively annealed out at 550°C. At the temperature above 600°C, a new peak is observed in the P implanted silica which is assigned to P_2 molecules trapped in silica host. The further annealing at T > 950°C suggests phosphorus nanoparticle quantum dots formation. The blueshift and the onset of the absorption band edge with the ion dose is consistent with quantum confinement theory.

References

1. A. I. Ekimov, A. L. Efros, A. A. Onushchenko, Solid State Commun. 56, 921 (1985).
2. L. E. Brus, J. Chem. Phys. 80,4403 (1984).
3. For example, Y. Wang and N. Nerron, J. Phys. Chem. 91, 5005 (1987); O. I. Micic, Z. Li, G. Mills, J. C. Sullivan, and D. Meisel, J. Phys. Chem. 91, 6221 (1987); C. J. Sandroff, S. P. Kelty, and D. M. Hwang, J. Chem. Phys. 85, 5337 (1986).
4. O. I. Micic, C. J. Curtis, K. M. Jones, J. R. Sprague, and A. J. Nozik, J. Phys. Chem. 98,4966 (1994).
5. R. L. Wells, S. R. Aubuchon, S. S. Kher, and M. S. Lube, Chem. Mater. 7, 793 (1995).
6. J. Warnock, D. D. Awschalom, Phys. Rev. B 32, 5529 (1985); D. O. Henderson, R. Mu, A. Ueda, A. Burger, K. T. Chen, and D. O. Frazier, Mat. Res. Sot. Symp. Proc. 366,283 (1995).
7. For example, M. Fujii, S. Hayashi, and K. Yamakoto, Appl. Phys. Lett 57,2692 (1990); Y. Sasaki and C. Horie, Phys. Rev. B 48,2009 (1993).
8. T. Shimizu-Iwayama, K. Fujita, S. Nakao, K. Saitoh, 'T. Fujita, and N. Itoh, J. Appl. Phys. 75, 7779 (1994); H. Atwater, K. V. Shcheglov, S. S. Wong, K. J. Vahala, R. C. Flagan, M. L. Brongersma, and A. Polman, Mat. Res. Sot. Symp. Proc. 316,409 (1994).
9. C. W. White, J. D. Budai, J. *G.* Zhu, S. P. Withrow, R. A. Zuhr, Y. Chen, D. M. Hembree, Jr., R. H. Magruder, D. O. Henderson, Mat. Res. Sot. Symp. Proc. 358, 175 (1995); C. W. White, J. D. Budai, J. G. Zhu, S. P. withrow, D. M. Hembree, Jr., D. O. Henderson, A. Ueda, Y. S. Tung, R. Mu, and R. H. Magruder, J. Appl. Phys (under review) (1995).
10. H. Hosono, T. Satake, M. Hosoe and Y. Abe, J. Am Ceram. Sot. 68, 7 (1985); ibid 64, 206 (1981).
11. R. A. Weeks, *Glasses and Amorphous Materials,* ed. J. Zarzicki, Chap. 6, pp332-373, Wiley New York 1992.
12. G. Pacchioni and G. Ierano, Phys. Rev. Lett. 79, 753 (1997); A.V. Amossov and A.O. Rybaltovsky, J. Non-tryst. Solids 179, 75 (1994).
13. H. Hosono, Y. Suzuki and Y. Abe, J. Non-tryst. Solids 142, 287 (1992).
14. G. Herzberg, *Molecular Spectra and Molecular structure, I. Spectra of Diatomic Molecules,* Chap. VI, pp-343, Nostrand, New York (1953).

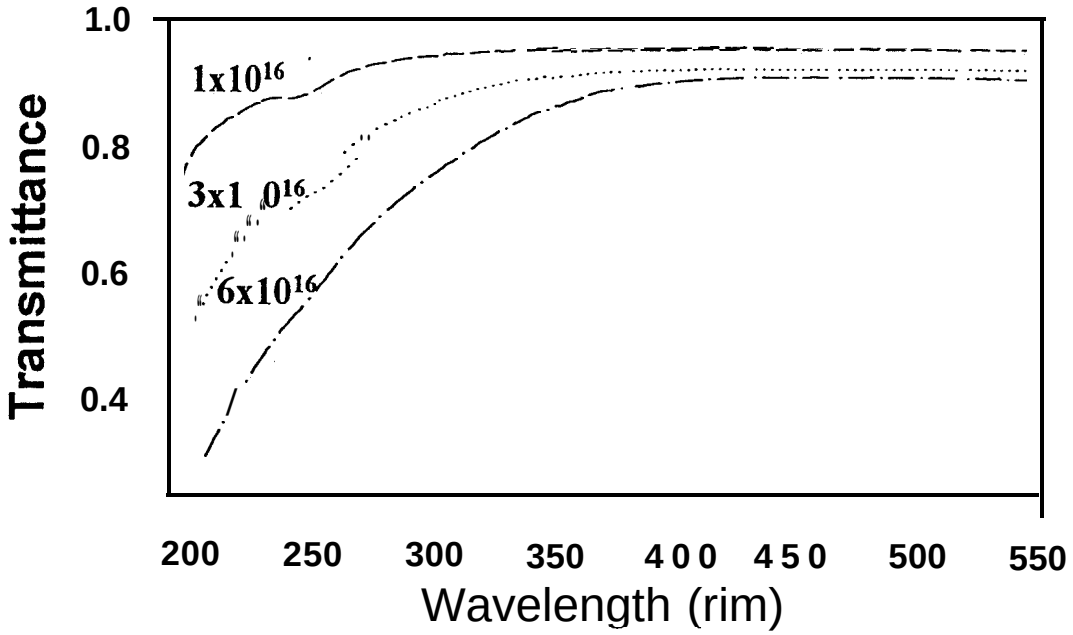


Figure 1. Optical spectra of phosphorus implanted fused silica with three different doses as indicated in the figure before any thermal treatment

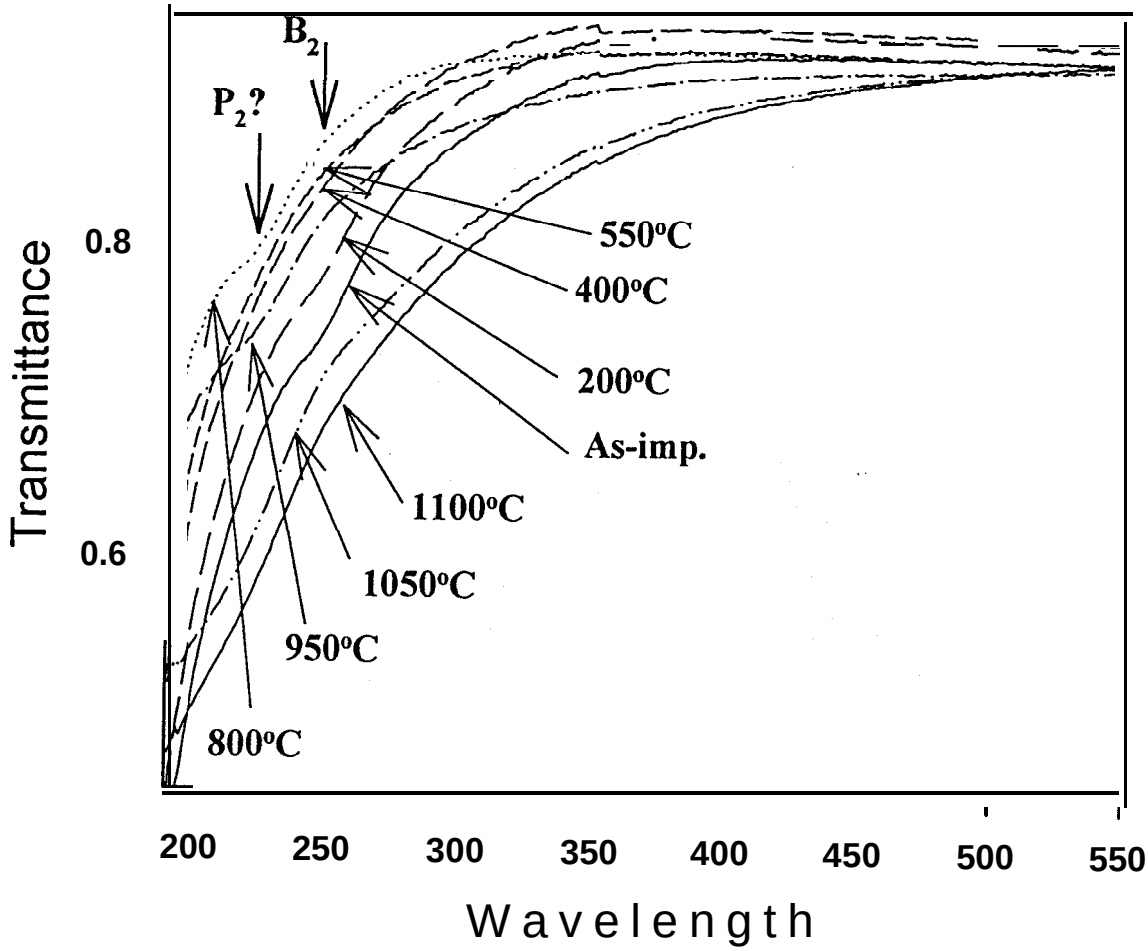


Figure 2. A set of illustrative optical transmission spectra of P implanted silica with $3x10^{16}$ ions/cm^2 annealed at the temperature up to 1100°C.

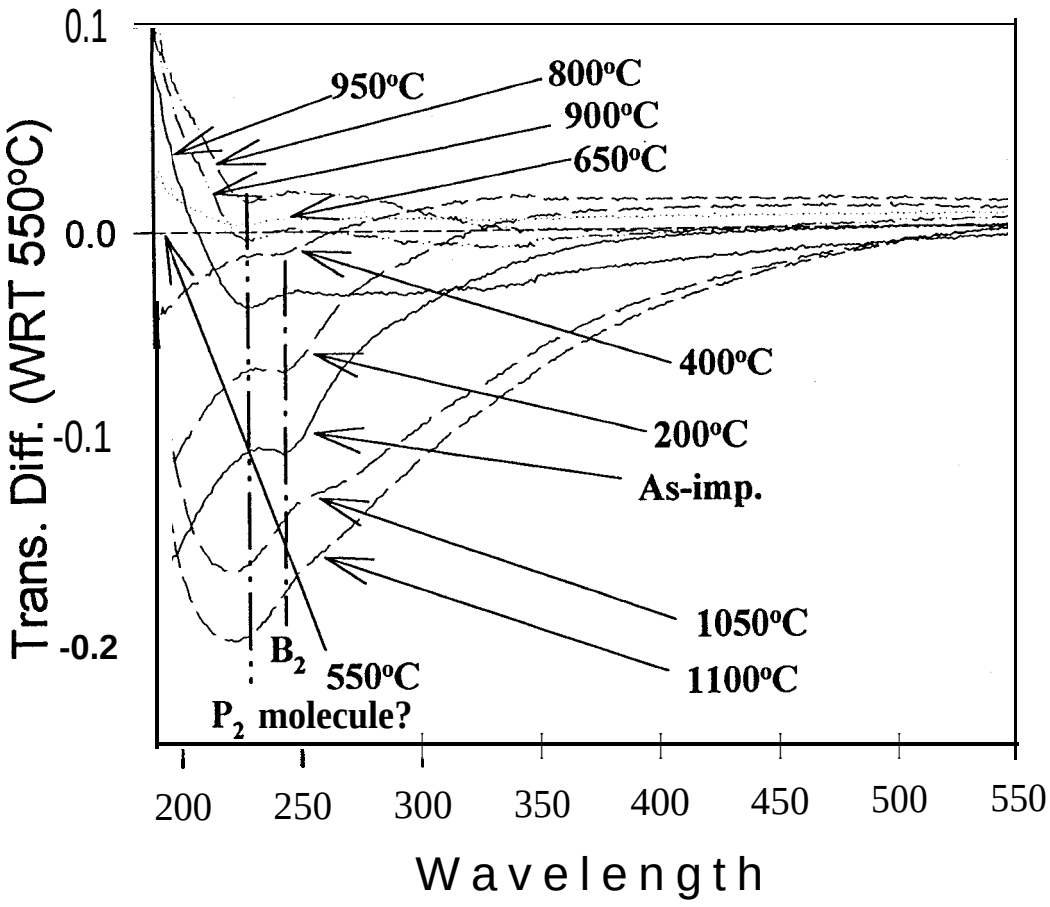


Figure 3. The selected difference transmission spectra of P implanted silica annealed at various temperatures. The ion dose of this sample is $3x10^{16}$ ions/cm^2 The difference spectra were generated by subtracting the spectrum at 500°C from the spectra at other temperatures.

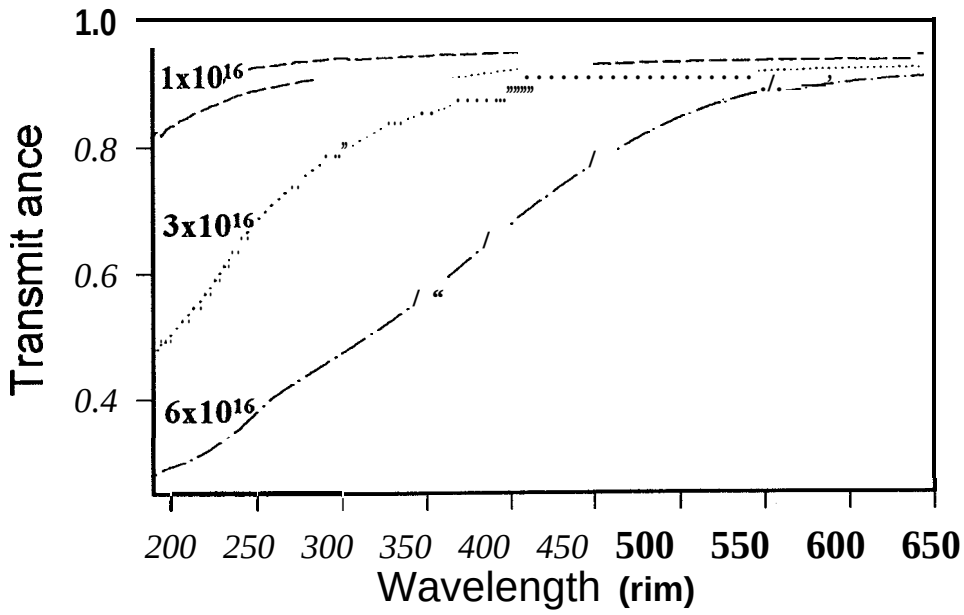


Figure 4. Optical spectra of P ion implanted silica after annealed at 1100°C with three different norminal doses as indicated in the figure.

Raman and Infrared Spectroscopy of Yttrium Aluminum Borate Glasses and Glass-ceramics

J. Bradley, M. Brooks, T, Crenshaw, A. Morris, K. Chattopadhyay, S. Morgan*
Center for Photonic Materials and Devices
Department of Physics
Fisk University
Nashville, TN 37208

Abstract

Raman spectra of glasses and glass-ceramics in the Y_2O_3-Al_2O_3-B_2O_3 system are reported. Glasses with B_2O_3 contents ranging from 40 to 60 mole percent were prepared by melting 20 g of the appropriate oxide or carbonate powders in alumina crucibles at 1400 C for 45 minutes. Subsequent heat treatments of the glasses at temperatures ranging from 600 to 800 C were performed in order to induce nucleation and crystallization. It was found that Na_2CO_3 added to the melt served as a nucleating agent and resulted in uniform bulk crystallization. The Raman spectra of the glasses are interpreted primarily in terms of vibrations of boron - oxygen structural groups. Comparison of the Raman spectra of the glass-ceramic samples with spectra of aluminate and borate crystalline materials reveal that these glasses crystallize primarily as yttrium aluminum borate, $YAl_3(BO_3)_4$.

Introduction

Glass-ceramics are ceramic materials produced by controlled crystallization of bulk glasses. This is accomplished by means of appropriate heat treatment which induces nucleation on a large number of sites and results in a fine-grained material with excellent mechanical strength.

Yttrium aluminum borate glasses were chosen for this study because of the interesting optical properties of potential crystalline phases such as yttrium aluminum borate, $YAl_3(BO_3)_4$, and yttrium aluminum garnet, $Y_3Al_5O_{12}$. Yttrium aluminum borate (YAB), for example, has a relatively large second order nonlinear optical coefficient ($d_{11} = 3.9 \times d_{36}$ (KDP)), and has been studied as a potential solid-state laser material. [1] Both YAB and yttrium aluminum garnet (YAG) are useful hosts for optically active rare-earth ions, which can substitute easily on the yttrium sites. These materials, of which Nd:YAG is the best known example, have found numerous applications in solid-state lasers. Nd:YAG is also of interest to NASA as a potential material for selective emitter applications in thermo-photo-voltaic energy systems. [2] The possibility of producing a polycrystalline, mechanically robust material with Nd:YAG as one crystalline phase is particularly interesting due to the possibility of texturing the surface of such a material.

In this paper, we report the results of experiments on the crystallization of yttrium aluminum borate bulk glasses to form glass-ceramic materials. These experiments include (a) the investigation of the glass forming region of the Y_2O_3-Al_2O_3-B_2O_3 system, (b) the effectiveness of Na, Ti, and P as nucleating agents for this glass system, and (c) the determination of the crystalline phases present in the resulting glass-ceramics.

Experiment

Glasses were made by mixing appropriate amounts of reagent grade Y_2O_3, Al_2O_3 and B_2O_3 in 20 g batches. Each batch was placed in an alumina crucible and melted in a Lindbergh box furnace at 1400 C for a period of 45 minutes. The molten glass was quenched by pouring onto a copper plate and pressing from the top with a second copperplate. After quenching, the glass was quickly placed into a second furnace and annealed at 450 C for approximately 1 hour. This resulted in a roughly circular glass disk about 3 cm in diameter and 2-3 mm thick. Samples were cut from the glass disks with a low-speed diamond saw and polished using conventional techniques. Ti, Na, and P (as TiO_2, Na_2CO_3 and $NH_4H_2PO_4$) were added to selected compositions in order to investigate their use as nucleating agents in the glass-ceramic process. Ceramming experiments were performed by heating the bulk glass samples to temperatures ranging from 600 to 800 C and holding for 24 hours. Compositions of the glasses are shown in Table I.

Table I. Summary of compositions (in mole percent) used for synthesis of yttrium aluminoborate glasses, glass-ceramics, and polycrystalline powders.

Sample	Y_2O_3	Al_2O_3	B_2O_3	Na_2CO_3	Other	Results
1	15	25	60			Glass, not glass-ceramic
2	15	25	50	10		Glass, glass-ceramic
3	15	25	58		2 [TiO_2]	Glass, not glass-ceramic
4	12	20	45	23		Glass, not glass-ceramic
5	15	25	50		10 [$(NH_4)H_2PO_4$]	Glass, not glass-ceramic
6	10	30	40			Synthesis of $YAl_3(BO_3)_4$
7	15	25	60			Glass, SiO_2 crucible
8	10	25	50	10	5 [Nd_2O_3]	Glass
9	10	30	60			Glass
10	5	35	60			Failed to form glass
11	20	20	60			Failed to form glass
12	25	15	60			Failed to form glass
13	22.5	37.5	40			Failed to form glass
14	25	15	60			Failed to form glass
15	5	35	60			Failed to form glass

Raman measurements were made using a Spex 1403 double monochromator with a thermoelectrically cooled photomultiplier detector and photon counting electronics. Excitation was provided by an argon ion laser operating at 514.5 nm with a power of 200 mW.

For comparison purposes, a sample of yttrium aluminum borate, $YAl_3(BO_3)_4$, was synthesized by solid state reaction using the procedure reported by Jung. [3] Y_2O_3, Al_2O_3 and B_2O_3 powders in the molar ratio 10:30:40 were thoroughly mixed and pressed into a 2.5 cm diameter by 3 mm pellet under a pressure of 200 kgf/cm^2. This pellet was then heated to 700 C and held for 20 hours. The pellet was then ground, pressed and heated to 1150 C for 30 hours to complete the reaction.

Results

Results of the glass making and glass-ceramic transformation experiments are summarized in Table I. These results indicate, that for glasses containing less than 60 mol% B_2O_3, glass formation is confined to very narrow composition range around $15Y_2O_3{\cdot}25Al_2O_3{\cdot}60B_2O_3$. Within the glass forming region, glasses were clear and colorless with no evidence of devitrification. After exposure to atmosphere over a period of several months, these glasses are durable, showing no visible evidence of surface degradation. Glasses without the addition of nucleating agents, or to which Ti or P had been added, did not form glass-ceramics. Only those glasses containing Na as a nucleating agent resulted in the glass-ceramic formation.

A Raman spectrum of the $15Y_2O_3{\cdot}25Al_2O_3{\cdot}60B_2O_3$ glass is shown in Figure 1. All of the glass samples showed similar Raman spectra, which can be explained by the relatively small variations in composition. Because of the broad, overlapping nature of the Raman bands, and the negligible compositional dependence of the spectra over the narrow glass-forming range, it is difficult to make a complete assignment of the glass spectra. Some conclusions, however, can be drawn from the Raman spectra. The appearance of a band near 770 cm-' is characteristic of a symmetric vibration of a six-membered borate ring with one BO_4 unit, [4] A band near 620 cm-l, which can also be observed in the spectra of these glasses, has been attributed to ring-type metaborate groups. [5]

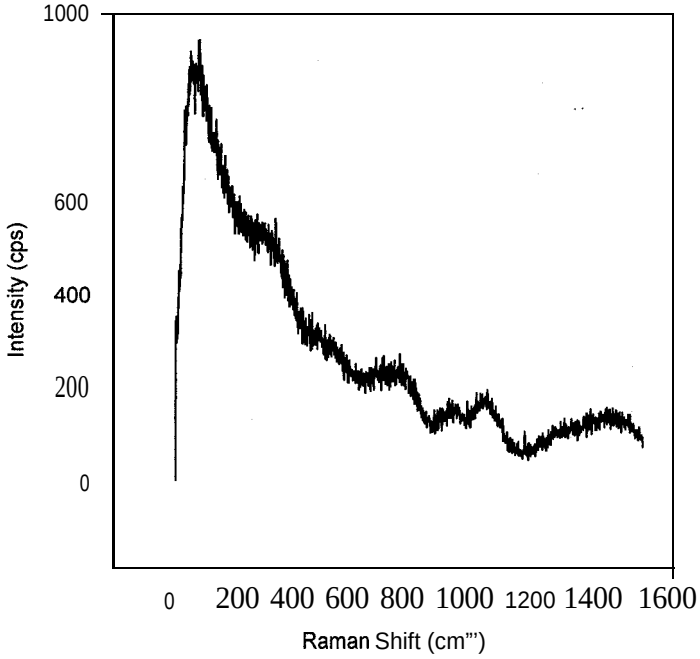


Figure 1 Raman spectrum of $15Y_2O_3{\cdot}25Al_2O_3{\cdot}60B_2O_3$ glass.

The Raman spectra of an yttrium aluminum borate glass-ceramic and $YAl_3(BO_3)_4$ powder are shown in Figure 2. It is apparent from these spectra that the principal phase present in the glass-ceramic is $YAl_3(BO_3)_4$. Some additional peaks can be observed in the glass-ceramic spectrum, such as the peak at approximately 430 cm-l. These peaks indicate the presence of additional crystalline phases in the glass-ceramic. The identity of these phases has not been determined at this time. The unidentified peaks, however, do not correspond to any peaks in the spectra of the starting materials, shown in Figure 3, or in the spectrum of YAG.

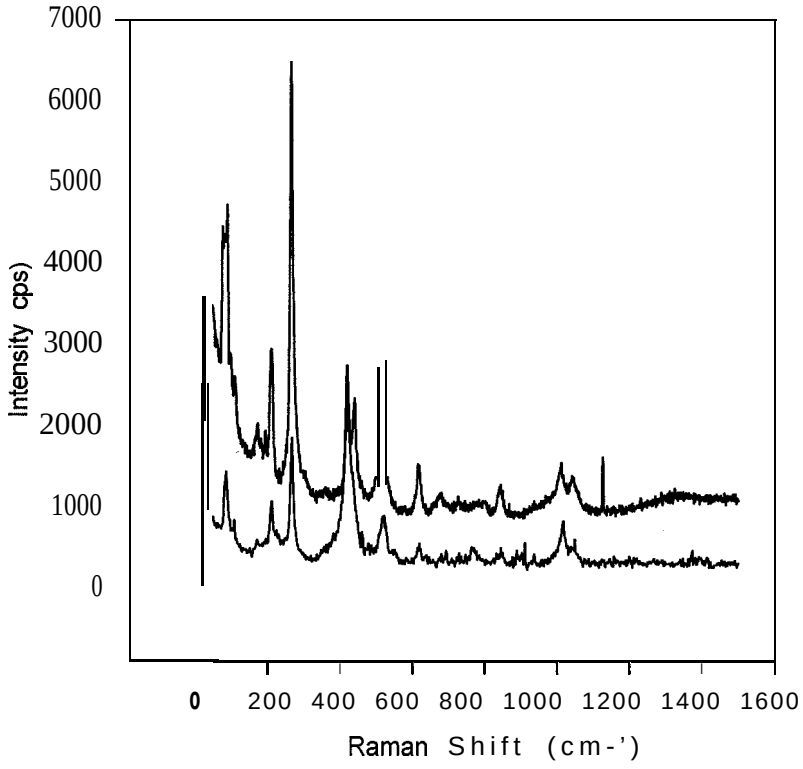


Figure 2 Raman spectra of yttrium aluminum borate glass-ceramic (top) and $YAl_3(BO_3)_4$ powder (bottom).

Summary

Glasses in the yttrium aluminum borate system with B_2O_3 contents ranging from 40 to 60 mole percent have been prepared by melt-quench techniques. Heat treatments of these glasses at temperatures ranging from 600 to 800 C were performed in order to induce nucleation and crystallization. It was found that Na (in the form of Na_2CO_3) added to the melt served as a nucleating agent and resulted in uniform bulk crystallization. The Raman spectra of the bulk glasses indicated the presence of six-membered borate rings with one BO_4 unit, and ring-type metaborate groups. Raman spectra of the glass-ceramics indicate that these materials consist primarily of a

$YAl_3(BO_3)_4$ phase, with the presence of one or more unidentified phases at a much lower concentration.

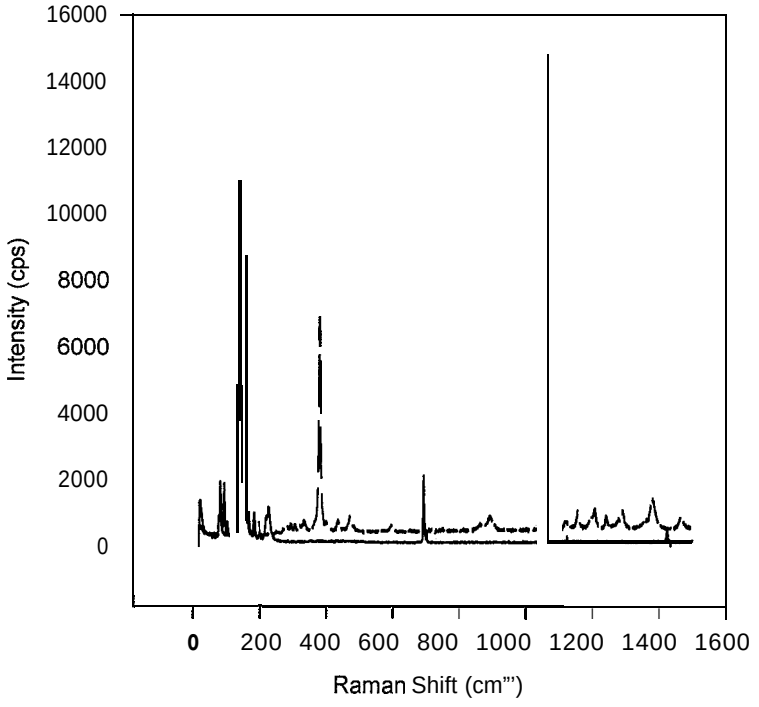


Figure 3 Raman spectra of Al_2O_3 (solid line) and B_2O_3 (dashed line).

Acknowledgments

The authors acknowledge the support of NASA through the Fisk Center for Photonic Materials and Devices, Grant NCC8-133.

References

1. L.M. Dorozhkin, 1.1. Kuratev, N.I. Leonyuk, T.I. Timchenko, A.V. Shestakov, Sov. Tech. Phys. Lett. 7,555 (1981).

2. R.A. Lowe, D.L. Chubb, S.C. Farmer and B.S. Good, Appl. Phys. Lett. 64,3551 (1994).

3. S.T. Jung, J.T. Yoon and S.J. Chung, Mat. Res. Bull. 31, 1021 (1996).

4. W.L. Konijnendijk and H. Verweij, J. Am. Ceram. Sot. 59,459 (1976).

5. W.L. Konijnendijk and J.M. Stevels, J. Non-Cryst. Sol. 18,307 (1975).

'Use of Earth Observing Satellite Data for the Development of "Learning Exercises" for College-Level Science Courses

Armond T. Joyce, Ph.D.
NASA, Education/University Affairs Office
Stennis Space Center, MS 39529

This paper is based on experiences being gained through a project entitled "The Mississippi Community College Pilot Project". The project was labeled "pilot" because it is thought that lessons learned during the implementation of this project may aid similar endeavors in other states. The objective of the project is to provide curriculum enrichment and associated faculty enhancement through the use of earth observations data in biological and physical sciences courses. The premise underlying the objective is that information from earth observations from satellite and aircraft platforms provides an effective means of illustrating and explaining science topics/phenomena in a new and/or different perspective. It is also thought that the use of data acquired from space may also serve to captivate the students interest and/or inquisitiveness about the particular science issue.

The project is implemented through two basic approaches. One approach is to make faculty aware of existing curriculum support material derived from earth observation data. This activity was based on a survey of teaching faculty that indicated a shortage of time available to them to search for and then procure such material. Part of the problem is that faculty who are not familiar with remote sensing technology or earth observations programs do not necessarily know what to look for or where to look. Although searches of web sites can be a very effective means of finding material, most faculty did not have internet-access or the necessary hardware and software for finding and downloading material during the course of this project. Therefore, providing an annotated list of available materials together with telephone numbers, postal addresses, and web sites for procuring the materials was thought to be necessary and time-effective. However, in the course of the project, it was also determined that such a list needs continual update due to new material, new web sites, material being expended, as well as changing postal addresses, telephones, email and web site addresses. The curriculum support material included on the listing included hardcopy printed material (phamplets, brochures, etc.) lithograph 8x10 color reproductions, posters, 35mm. slides, videos, and CD's. Most of the material (outside of that available on web sites) was provided through NASA or NASA-funded projects, but some material was also provided through other federal agencies. Once such material was in the hands of Community College faculty, it was left to them to determine the relevance of the material to their respective courses, and how to use it.

The other approach, which is the major focus in this paper, was to work with faculty to develop computer-implemented "learning exercises". The "learning exercises" developed in this project embodied the following guidelines:

1. Develop a basic understanding of the science topic.
2. Minimize text, and give emphasis to use of graphs, images, animation, and video clips with audio.
3. Give highest priority to earth observation images that best illustrate the science topic, but strive to use some images in the local geographic areas familiar to the students.
4. Include self-administered multiple-choice quiz's embedded in or at the end of the "learning exercise".

5. Organize the material with multi-media software to be a self-learning activity requiring the equivalent time of one or two lab periods to complete.
6. Assume a higher level of hardware technology than exists in the schools at the present time.

An example of a "learning exercise" could be as follows:

1. The topic of "vegetation monitoring" is identified as a topic in the context of the subject of photosynthesis being covered in a biology, environmental science, or plant physiology course.
2. Text and graphics are developed to summarize how photosynthesis results in the formation of chlorophyll and how the presence of chlorophyll influences energy reflected from the leaves.
3. Text and graphics are developed to show the electromagnetic spectrum and illustrate basic principles of remote sensing.
4. Text, graphics, and images are introduced to show the characteristics of a particular remote sensor and how it measures reflected energy.
5. Images derived from the data acquired by the sensor are displayed to show how the seasonal cycle of new leaf flush, leaf development, loss of chlorophyll pigment, and leaf senescence can be monitored throughout the year for large areas.

The overall development of "learning exercises" for this project proceeded through the following steps:

1. Identify and contact instructors of science courses.
2. Provide orientation/training in earth observations through remote sensing technology via distance learning and hands-on training (2-day and 5-day orientations).
3. Work with instructors to identify topics within course curriculum for which earth observations data can be used to illustrate a phenomena, relate to a scientific issue, or be used for applied calculations.
4. Work with instructors, to locate and procure digital earth observation images/data and other relevant source material, e.g., videos.
5. Provide assistance/orientation in use of multimedia software in the development of the learning exercise.
6. Provide "learning exercise" on floppies and/or CD to other faculty for evaluation.
7. Use feedback to finalize the "learning exercise" and post on server and press CD's.

To date, the following "learning exercises" that have been developed, and the courses to which they are likely to be relevant are as follows:

LEARNING EXERCISE	COURSE(s)
Monitoring Seasonal Changes in Vegetation	Biology, Environmental Science, Plant Science
Electromagnetic Radiation and the Spectrum	Physics
Earth's Radiation Budget	Physical Science
Global Carbon Budget and Tropical Deforestation	Geography, Environmental Science, Survey of Forestry
Determining Surface Temperature	Biology, Physical Science, Marine Science
Phytoplankton Concentrations and Ocean Color	Marine Science, Biology
Atmospheric Ozone Depletion	Environmental Science, Physical Science, Chemistry
Detection of plant stress due to insects, disease, and/or drought	Biology, Plant Science, Forest Protection

As "learning exercises" were developed, they were placed on floppies (in compressed form) and distributed to selected faculty for constructive critique. After those listed above were developed, they were placed on CD's which were distributed to a larger number of teaching faculty for a more elaborate evaluation. For this purpose, a "mail-back" evaluation form was developed and distributed with the CD, but evaluators were also encouraged to submit more detailed comments, opinions, critiques in their own words via email. They were also encouraged to involve students in the evaluation. The feedback is subsequently to be used to improve the content and finalize the "learning exercise".

At the beginning of this project, it was planned to offer a five-day remote sensing/image analysis short course to faculty who wished to be involved in the development of "learning exercises". After project activities commenced, it became obvious that many faculty lacked an understanding of earth observations through remote sensing, and that they hesitated to use material derived from remote sensing in their courses for that reason. In consideration of this observation, it was decided to offer a four-hour introduction and a two-day course in addition to the five-day course. The four-hour introduction was conducted through an interactive TV distance learning approach, actually split into two two-hour sessions in order to accommodate the faculty's schedules. The two-day course focused on the basics of remote sensing, the characteristics of the sensors, and the information content of the earth observations data/images. The five-day course covered the same topics as the two-day course, but also included training via hands-on, computer-implemented image processing/analysis.. The two-day and five-day courses were scheduled shortly after the termination of the school year and again shortly before the beginning of the school so as not to be disruptive of either the school semester or the summer activities of the faculty. This training was provided at Jackson State University and at NASA's Stennis Space Center. Although time did not permit the necessary arrangements for this project, it would have been desirable to arrange this training through a university as "for-credit" courses. Such a course would provide more incentive for participation.

A basic concern at the initiation of the project was the need to provide for future access of data. Most sources necessitate familiarity with accessing databases through the INTERNET, and the downloading and storage of image data. Since most faculty were not familiar with these operations, arrangements were made to provide training. This training was provided in two-day short courses conducted by the University of Mississippi. Once that faculty are familiar with these operations, they can use the INTERNET for getting information on similar programs for curriculum support, curriculum development, tutorials, and training opportunities in addition to getting data. In order to facilitate access to such information, the participating faculty are provided with a continually updated listing of relevant web site addresses,

Another problem encountered during the project was the coordination of meetings to introduce the concept and for training with respect to the faculty's schedules and available time. This was partially alleviated by use of distance learning, self-learning through videos and web sites, and making the training available on individual campuses rather than at distant sites. However, the lack of available time for teaching faculty during the school year was still a limiting factor.

The instructor-student evaluation of the "learning exercises" listed in this paper is currently in-progress, after which finalization will take place during the summer of 1998 with intentions of pressing the final CD and placing the "learning exercises" on the web site by the fall of 1998.

GLOBE: An international-Interdisciplinary Prow-am

J. Lopez-Garriga ,R. Camacho, W. Nazario, W. De Jesus, and D. Mora University of Puerto Rico.

Introduction

In a world where science and engineering have become two of the most challenging fields of studies, interdisciplinary education, experimental knowledge, and hands-on experience are components of a successful general education. Students should be given opportunities to develop adequate communication and reasoning skills through hands-on scientific experiences. They must learn how to think critically and logically in order to create and analyze scientific data and arguments. The integration helps students to understand, consolidate, and mature the knowledge and the relevance of their science activities. This in turn will prepare them to choose wisely their carrier and it will enhance their educational opportunities. To achieve these goals various educational initiatives have been created (1-1 O). One of them, the Global Learning and Observations to Benefit the Environment (GLOBE) program, (www.globe.gov) involves students and teachers from over 5,500 schools in more than 60 countries around the world. Pre-college students and teachers work with professors and research scientists to learn more about our planet. The program uses environmental experiments to help develop skills in chemistry, earth sciences, mathematics, data analysis and communication. The general topics discussed in support of the GLOBE protocols measurements involve Atmosphere, Hydrology, Land Cover, and Soils.

The Department of Chemistry of the University of Puerto Rico at Mayaguez, is an active franchise of the GLOBE program. Both programs seek the integration of chemistry demonstrations with activities of the GLOBE program as a way to motivate pre-college students and create an integration between chemistry, environment and communications. The basic concepts are the science shows, science demonstrations, and the GLOBE workshop sessions for teachers. These activities are a key to make the study of chemistry both fun, educational, and related to daily activities and communications. The initiative has two overarching goals. The first is to help teachers incorporate GLOBE activities and the chemistry demonstrations in their science and mathematics courses. The second one is to disseminate the new learning activities to both economically and challenged schools. Through these goals the program has been able to: (1) Implement a cost effective program in chemistry and science, reaching pre-college students and teachers. (2) Continue professional development of K- 12 teachers using hands-on experiments and presentations that model the activities at successive training sessions. (3) Involve grade K-12 teachers in educational initiatives to develop and adapt relevant chemical demonstrations related to the GLOBE activities. (4) Motivate pre-college students to enroll in science courses and to participate more actively in science programs. (5) Improve the general scientific understanding and background knowledge of those students who wish to pursue a science career, as of those who may select another field of studies. (6) To disseminate a bank of reproducible chemistry activities at international level though out the World Wide Web site of the Science on Wheels program and the GLOBE Bulletin.

Activities and Methodology

Workshops for teachers constitute our main link between the junior high and high school teachers and the GLOBE activities. The workshops and follow-up sessions provide teachers the opportunity to apply the learning activities and science skills through hands-on experiences. Experiences which in turn they can use as demonstrations in their science classes or as hands-on

projects for their students. To prepare teachers a professional program that combines intensive workshop and the follow-up implementation has been designed. Typically, thirty teachers participate in each workshop, working in pairs and twelve teachers participate in the follow-up sessions. Workshops last from five to six hour, including discussion periods. The workshops last for a period of a week and the follow-up sessions are during Saturdays.

One of the objectives of the pre-college teachers' workshops is to prepare teachers to include demonstrations, addressing the students' eagerness to see more of these activities incorporated in their science classes. As part of this preparation after a question and answers period, the experiments are performed by the teachers. The teachers learn how to reproduce interesting yet safe demonstrations and experiments for their students with accessible materials (from drug stores, supermarkets and hardware stores), so that the experiments can be repeated by the students at home. The participants receive documentation, equipment and materials for each of the activities practiced during the workshop. In this way they can implement the activities in their own classrooms. An evaluation for each new learning activity is prepared to generate feedback of the effectiveness of the new learning activities in improving conceptual knowledge in specific content areas and skills, such as data analysis, that are aided by the use of technology. This is a way to widen the exposure of teachers, students and parents to scientific phenomena.

During the follow-up sessions teachers practice the learning activities and receive feed back from their peers and workshop instructors. They share and discuss issues and experiences on the actual implementation of the lesson; and additional activities and resources are provided according to the specific needs of the participants. We assist teachers in developing communities of learners that continue to meet on their own to further the curricular strategies. The chemical demonstrations, the GLOBE protocols and the computer communications, provide teachers with activities that can be presented in their own classes. The synergism generates a theoretical background and hands-on experience needed to incorporate chemistry, science, environmental measurements, and data collection/analysis activities in their classes. The knowledge acquired by the teachers translates into activities that will increase the scientific educational background of their students. We have been engaged in scientific demonstrations for teachers over the last four years as part of Science on Wheels. In the literature there are several excellent books and review articles (1- 10) that can certainly enrich the nature of chemistry demonstrations used to explain chemistry related to the GLOBE protocols and measurements. For example, Table 1 shows a list of demonstrations using house-hold chemicals that are been adapted to clarify chemistry concepts behind the GLOBE protocols. For each suggested activity, the purpose, level, key concepts and skills are discussed in away easy to implement in a class room.

The Science on Wheels program and the Puerto Rico GLOBE franchise have established a series of links between pre-college schools and the university. This coherent communication allows an effective dissemination of the learning activities. The most active schools in the application of the GLOBE protocols are serving as Professional Development Mini-centers with leader teachers. They help disseminate the activities to other schools throughout their region. It is important to remember that we have consecutive meetings with the teachers. In each meeting we reinforce the information that they need to perform and understand particular activities. These activities and demonstrations are resources for teachers in economically challenged school districts. We actually have a World Wide Web (http: //netdial.caribe. net/~sonw) site for the dissemination of the activities that we carry out in both the GLOBE and Science on Wheels programs. The dissemination includes communications about new science activities that teachers and students can do together.

Results and Discussion

Last August, at the facilities of the Science on Wheels Outreach program, thirty-two teachers became certified GLOBE instructors. Participating teachers came from each of the seven educational regions of the Island as determined by the Puerto Rico Department of Education. They represented twenty schools from both rural and urban areas. Figure 1 (Items **1,9-12)** shows that the workshop was very successful. The teachers feedback evaluations demonstrated their satisfaction and interest in implementing the GLOBE program. Despite these efforts, Figure **1** as shown in the next page (i.e. items 3 to **8), makes evident the need for activities** that reinforce: (i) ***Science Concepts (mathematics, chemistry, and earth sciences)*** and (ii) ***Computer Skills (data handling and interpretation).*** The need for enhanced technological skills for in-service science teachers is also evident by curricular reform efforts both at the island and national levels. In terms of critical curricular objective that could be strengthened through the GLOBE program are: the new year long scientific research course required by all high school students; the depth of conceptual understanding and interdisciplinary connections promoted through the new science and mathematics standards based on national standards of the NCTM and NRC; and the State

Table 1. Science Concepts and associated GLOBE Protocols

Concept	***Protocol***
pH	**Atmosphere, Hydrology, Soils**
Conductivity	**Hydrology**
Cloud formation	**Atmosphere**
Condensation	**Atmosphere**
Heat capacity of water	**Hydrology, Soils**

pH

Purpose To demonstrate how the pH of solutions can change from basic to acidic or from acidic to basic.
Level Intermediate, Advanced
Key concepts and skills:
Concepts: pH; pH scale; Acids; Bases; **Indicators**
Skills: Testing prediction Observing; Describing and identifying basic and acidic solutions; Generating questions and developing hypotheses.

Conductivity

Purpose: To enable students to understand the conductivity of a substance and to test if a substance has the capacity to transmit electric current.
Level Intermediate, Advanced
Key Concepts and Skills:
Concepts: **Conductivity, Anode, Cathode, Insulator, Circuit**
Skills: Making observations, Analyzing situations and deducting

Condensation

Purpose: To demonstrate that water can be taken from the air and to simulate the process of condensation.
Level All
Key Concepts and Skills
Concepts: Condensation, **Temperature changes, Water phases**
Skills: Observation of the water cycle, Deduction based on observations, Identification of relationships, Analysis of different phases of the water

Cloud Formation

purpose students can comprehend and simulate the formation of clouds.
Level All
Key Concepts and Skills
Concepts: **Clouds, Steam, Condensation, Temperature changes**
Skills: **Understand atmospheric behavior, Compare characteristics of the atmosphere.**

Heat Capacity of Water

Purpose: Demonstrate the effect of the water on heat and the importance of it in the Earth.
Level All
Key Concepts and Skills
Concepts: **Heat, Heat capacity, Properties of water**
Skills: Analyze concepts, Think critically, predicting the water function

Technology Plan that seeks to develop computer and electronic communication literacy among all students. Teachers need more learning activities and hands-on experiences which can help them explain to their students the science concepts and data interpretation behind the GLOBE protocols and science courses. They also need to understand the mathematical link between measurements, data representation, and data analysis.

As part of our GLOBE workshop we presented a set of science demonstrations to the teachers which included topics such as:
Combustion and the fire triangle (illustrated with balloon explosions using helium, hydrogen and oxygen), ***Acid-base reactions and indicators*** (illustrated using basic solutions with different indicators and dry ice as a reagent to change pH and see indicator color changes), ***Phase changes and temperature*** (illustrated by the use of dry ice), and *Cloud formation* (illustrated by the use of liquid nitrogen). The demonstrations and scientific discussion during the presentation helped the teachers clarify various chemical and physical concepts behind the GLOBE protocols. It also created a set of activities very useful for explaining scientific concepts of the protocols to their students.

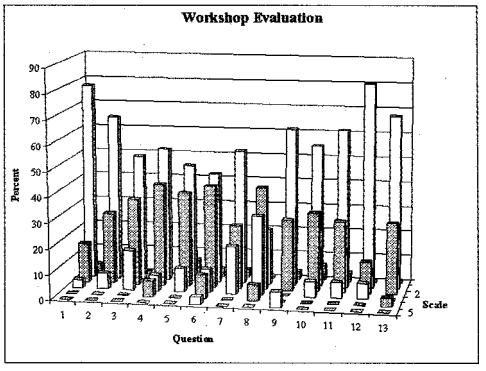


FIGURE ONE

Scale: The evaluation is based on a 5-point scale where one (1) means strongly agree and five (5) means strongly disagree.
Question: 1. The workshop provided an effective introduction to the GLOBE. 2.-5. I feel confident that this workshop has prepared me to conduct the GLOBE Scientific Measurements with my students in the area of 2. Atmosphere 3. Hydrology 4. LandCover /Biology 5. Soils 6. I feel confident that this workshop has prepared me to send and receive GLOBE data via computer technologies. 7. I feel confident that the workshop has prepared me to help students understand and use the visualizations provided by GLOBE on the GLOBE web server. 8. I feel confident that the workshop has prepared me to use the MultiSpec Software with my students. 9. I think the GLOBE Program workshop materials provided a useful resource to help me participate in the program. 10. I think that the workshop introduced the materials sufficiently to enable my use of them when implementing the program. 11. The workshop structure provided me with sufficient opportunities to interact with the GLOBE training team. 12. The workshop structure provided me with sufficient opportunities to interact with other teachers. 13. I feel confident that the workshop has prepared me to involve other teachers and classes in my school in GLOBE.

Contributions and potential applications

Students will acquire higher educational skills in science and mathematics. They will learn to question data, identify cause and effect, and apply the learned skills to new situations. These activities will develop a better understanding of the environment and its changes, through models from data, patterns, or relationships.

Teachers will development innovative activities trough elemental instrumentation, that complements Ecological, Environmental, Earth Sciences and Research courses. Increase the relationship between the students and teachers. Increase learning and teaching methods.

REFERENCES

1. Deavor, J. P.; Deavor, J.W. *J. Chem. Ed. 1995, 72,798.*

2. **Turner, K.** *J. Chem. Ed. 1995, 72,203.*

3. Waldman, A. S.; Schechinguer, L.; Nowick, J. S. *J. Chem. Ed.* **1996, 73,762-764.**

4. Barry, D.M. *J. Chem. Ed .1996, 73,209.*

5. Nolan, W. T.; Gish T.J. *J. Chem. Ed.* **1996, 73,651-653.**

6. Beasley, W. *J. Chem. Ed.* **1996, 73,344-346.**

7. Shakahashiri, B. ***Chemical Demonstrations: A Handbook for Teachers of Chemistry;*** The University of Wisconsin Press: Madison, Wisconsin. 1989; Vol. 1-4.

8. Cash, T., Parker, S. and Taylor, B. 175 ***More Science Experiments to Amuse and Amaze your Friends***; Random House: New York, 1989.

9. Cash, T., Parker, S. and Taylor, B. ***175 More Science Experiments to Amuse and Amaze your Friends***; Random House: New York, 1990.

10. Lopez, J.; Muñoz, Y, Torres, V, Echevarria ,Y; Nazario ,W.; De Jesus, W.; and Camacho, R. *J. Chem. Educ.* **1997,74,1346-1349.**

Development of Multi-Disciplinary Finite Element Method Analysis Courses at California State University, Los Angeles

By John McKinney P.E.
and
Dr. Chivey Wu

Department of Mechanical Engineering
California State University, Los Angeles

The NASA Dryden Flight Research Center (DFRC) Partnership Awards Grant to California State University, Los Angeles (CSULA) has two primary goals that help to achieve NASA objectives. The overall objectives of the NASA Partnership Awards are to create opportunities for joint University NASA/Government sponsored research and related activities. One of the goals of the grant is to have university faculty researchers participate and contribute to the development of NASA technology that supports NASA goals for research and development (R&D) in Aeronautics and Astronautics. The other goal is technology transfer in the other direction, where NASA developed technology is made available to the general public and more specifically, targeted to industries that can profit from utilization of government developed technology.

This years NASA Dryden Partnership Awards grant to CSULA entitled, "Computer Simulation of Multi-Disciplinary Engineering Systems", has two major tasks that satisfy overall NASA objectives. The first task conducts basic and applied research that contributes to technology development at the Dryden Flight Research Center. The second part of the grant provides for dissemination of NASA developed technology, by using the teaching environment created in the CSULA classroom. The second task and how this is accomplished is the topic of this paper.

The NASA STARS (Structural Analysis Routines) computer simulation program is used at the Dryden center to support flight testing of high-performance experimental aircraft and to conduct research and development of new and advanced Aerospace technology. Figure 1 depicts the role STARS plays at Dryden in support of flight projects, aerospace vehicle technology development, computer tools (R&D), and how teaching is a part of it.

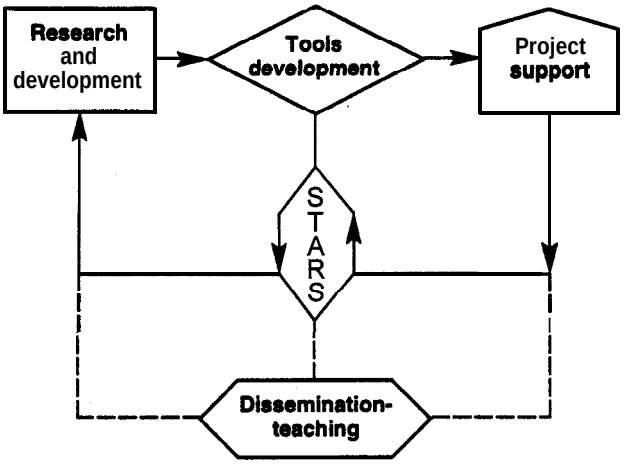


Figure 1 Role that STARS plays at NASA Dryden Flight Research Facility

STARS has been developed at DFRC to analyze high performance research vehicles where interactions of many complex physical phenomena are taking place. It is a single unified computer program that can be used to simulate structures, aerodynamics, heat transfer, controls and their interactions simultaneously. The program is based on the Finite Element Method (FEM) and has incorporated a number of advanced numerical techniques used to accurately simulate aerospace-flying vehicles. In addition to the core solver, there are both pre and post processing modules for creation of vehicle models and the graphical interpretation of their results. The program is classified as a "Multi-Disciplinary Simulation Program", because of its overall integrated approach to problem solving. Figure 2 gives an overview of the STARS computer program and its multi-disciplinary working environment.

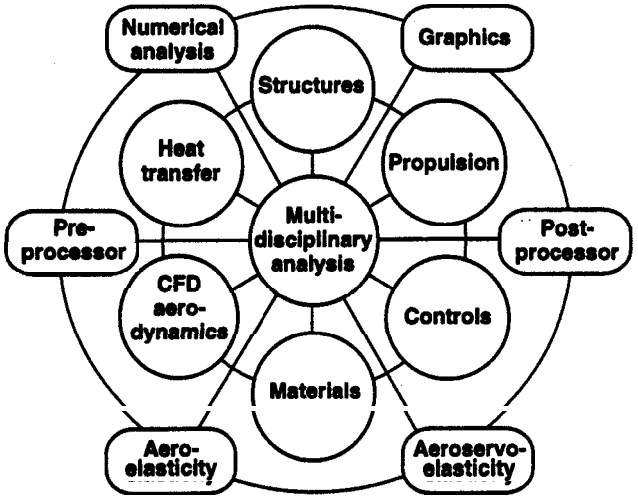


Figure 2 The STARS Multi-Disciplinary Working Environment

To help achieve the goal of disseminating NASA Dryden computer simulation technology, the primary goal of the Partnership Award Grant is to develop, STARS based, multi-disciplinary engineering analysis courses at the CSULA campus. It is felt that developing and teaching courses on the principles and applications of the STARS program would be a highly effective way to transfer the technology not only into the university environment, but also ultimately into the commercial sector. This transfer would be created by having both university students and practicing engineers learn the program and than carry

their knowledge of the technology back to their current and future employers, It is here, in the industrial community, where the investment in government developed technology, hopefully, will find its ultimate payback.

The courses being developed at CSULA are designed to meet educational needs of undergraduate students as well as graduate students and practicing engineers in the local aerospace industries. A first introductory course will introduce the students to the basic principles of the Finite Element Method (FEM) and its implementation on the STARS program. Lectures, examples and computer lab projects covering structures, dynamics, heat transfer, loads, boundary conditions, computer modeling and post-processing of results will be covered in this course. Example problems will be selected to provide explanation and physical understanding of the FEM modeling process. More advanced examples typical of those found in industry and at NASA will also be introduced and attempted by the students as their learning experience grows. A fundamental part of this process will be gained in a "hands on" computer lab set-up especially for this project. Figure 3 shows the course schedule for the first 10-week course.

wk 1-2	Course Introduction and Overview NASA Dryden Research Grant The STARS Computer Program Windows NT Workstations STARS Computer Simulation Project Laboratory
wk 3-4	Overview of Aerospace Structural Analysis Basic Methods, Determinate Structures &Strength of Materials Introduction to the Finite Element Method Computer Project on Rods, Beams and Trusses
wk 5-6	Indeterminate Structures Energy Methods Rod and Beam Element Stiffness Introduction to Matrix Methods The Finite Element Method (FEM) Computer Project on Space Frames&Stress Analysis
wk 7-8	Introduction to FEM Plates and Shells Material Models Applied Loads Mesh Generation Techniques Computer Project on Plates and Pressure Vessels
wk 9-10	Analysis of Aerospace Structures Using STARS Beam Stiffened Shells Combined Loading Computer Project on Wing and Fuselage Analysis

Figure 3 Introductory Multi-Disciplinary FEM Course Outline

The lab is equipped with 3 "state-of-the-art", high-end, Windows-based workstations, specifically configured to run the STARS program. The students working with the user's manual and test cases will initially fill their knowledge of STARS by solving fundamental problems. More advanced models of aircraft structures, spacecraft and other typical problems will be generated using the pre-processor. These problems will explore the more practical aspects of aerospace vehicle design and analysis.

A second more advanced course will address more of the multi-disciplinary analysis capability of the STARS program. Capabilities such as dynamic response, non-linear behavior, fluid-structure interaction, based on computational aerodynamics, and control system interaction will be covered using advanced principles, examples, problems and test cases. It is expected that successful completion of the courses will give the students a working knowledge of the STARS program and understanding of the basic principles on which the program is based. This knowledge will hopefully initiate the transfer of NASA technology to the university and its students, who would in turn bring it into the private sectors and aerospace industries.

As an overview example of how the first course is being organized and conducted, a condensed excerpt taken from the lecture material will be presented and discussed.

A typical plane truss problem shown in figure **4** is to be analyzed by "hand", using finite element theory, and also using the STARS computer program. A comparison and discussion of the results using both techniques will be made.

The truss has been modeled as a collection of "bar" elements, connected between nodes, and having specified material properties and cross-sectional areas.

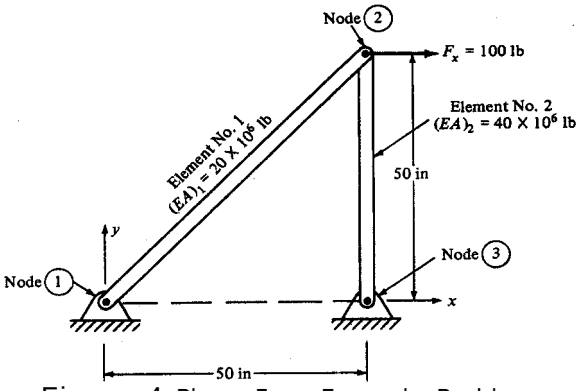
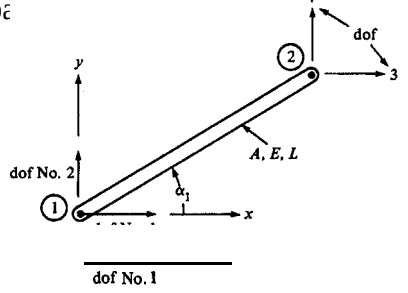


Figure 4 Plane Truss Example Problem

A key aspect of the finite element method utilizes stiffness matrices for different classes of elements. In this particular case the very much simplified bar or truss element is used. Figure 5 shows a typical bar element in a "globa



$$k = \frac{AE}{L}\begin{bmatrix} C^2 & CS & -C^2 & -CS \\ & S^2 & -CS & -S^2 \\ & & C^2 & CS \\ \text{Symmetry} & & & S^2 \end{bmatrix}$$

Figure 5 Local Element Stiffness Matrix

By applying the basic equations of static equilibrium at each nodal point, a system of equations relating the applied forces to the" internal" element forces can be developed. The element forces are a result of their" stiffness" and how much they deform, the deformations being the unknowns of the problem, Figure 6 presents the basic equations of the finite element method.

$$\sum_{e=1}^{N} k^{(e)} = K \qquad \sum_{e=1}^{N} f^{(e)} = F$$

$$F = Kd$$

Figure 6 Basic Equations for Linear Static Analysis Using FEM

For this "classical" example problem there are only two unknowns, the displacements of node 2 in both the x and y directions, In a "real world" problem the number of unknowns can be in the thousands and would only be solvable using the computer, Figure 7 shows assembly of the "global" set of equations after condensation has taken place and solution of the unknowns,

$$[k] = \begin{matrix} 141{,}443 & 141{,}443 \\ [141{,}443 & 941{,}443] \end{matrix}$$

$$\begin{bmatrix} 141{,}443 & 141{,}443 \\ 141{,}443 & 941{,}443 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} 100 \\ 0 \end{Bmatrix}$$

$$141{,}443u_1 + 141{,}443u_2 = 100$$

$$141{,}443u_1 + 941{,}443u_2 = 0$$

$$u_1 = 0.000833 \text{ in}$$

$$u_2 = -0.000125 \text{ in}$$

Figure 7 Condensed System of Equations and Solution

```
   NN     NEL    NMAT  NMICN   NEP        SST   NLGCS  NMANCL  NSTACK  MAXLEL
    4       2       1      4     2          0       0       0       0       0

  NTMP     NPR   NSPIN      NC    NBUN   NLSRC  NCNTRL    NOUT    NEXP     NNA
     0       0       0       L       0       0       0       0       0       0

 IPROB    IEIG    IDRS    IBAN  IPLUMP  IMLUMP   INMM   IINTP   IGLE
     8       0       0       1       1       0       0       0       0

 IPREC   IPLOT  IPRINT INDATA  IERCHK  INCFOR  INZDBC   IDBC  IMATE   NLOOP
     2       0       1       0       1       0       0       1       0       1
GENERATED NODAL COORDINATES

   NODE no.          X-COORD.        Y-COORD.        Z-COORD.   OX   UY   UZ   TX   TY   TZ LGCS DRCS INC
 INT     EXT

   1       1        .0000E+00       .0000E+00       .0000E+00    1    1    1    1    1    1    0    0    0
   2       2        .5000E+02       .5000E+02       .0000E+0     0    0    0    0    1    1    1    1    0    0    0
   3       3        .5000E+02       .0000E+00       .0000E+00    1    1    1    1    1    1    0    0    0
   4       4        .1000E+02       .1000E+02       .1000E+02    1    1    1    1    1    1    0    0    0

    GENERATED XL EMENT DATA

     ELEM ELEM  NODE  NODE  NODE  NODE  NODE  NODE NODE  NODE  IMPP IKPP ITMPP  IPRR  ISPI INC
     TYPE NO .     1     2     3     4     5     6    7     8       /ITHTH

        1    1     1     2     -     4     1          1     0     0     0     1     1     0     0     0     0
        1    2     2           3     4     1     1          0     0     0     1     2     0     0     0     0

    PROPERTIES OF LINE ELEMENTS

     PROP          AREA              JX              IY              IZ              SFY             SFZ
     TYPE

      1       .6667E+00       .0000E+00       .0000E+00       0000E+00       .0000E+00       .0000E+00
      2       .1333E+01       .0000E+00       0000E+00       .0000E+00       .0000E+00       .0000E+00

    MATERIAL PROPERTIES

    ISOTROPIC CASE
     PROP                    E           NU        ALPHA          RHO       YIELD  STRESS  HP
     TYPE

      1              .3000E+08       .3000E+00       .0000E+00       0000E+00       .0000E+00       .0000E+00

    NODAL LOAD SETS

     SET  NO. 1

        NODE
       EXT   INT   DOF          LOAD

         2     2     1       .1000 0E+03
```

Figure 8 STARS Ouput 1 File for the Example Truss Problem

The modeling approach and solution will now be discussed. First an input file of the truss is created using the format specified in the NASA STARS users manual, NASA Technical Memorandum TM 4795.

Successfully reading the input deck results in the creation of three different output files. The first file is an echo back of the input deck, shown in figure 8; its purpose is to verify that the model has been constructed according to the intended modeling approach. This model is composed of two elements, four nodes, with an extra "dummy" node being used to orient the elements, one material and two material properties. The 2nd to 4* lines are solution parameters that control many of the numerical procedures implemented in the program.

The nodal coordinate block specifies the x,y,z coordinates of each node, as well as information about the constraints associated with them. The element data block describes how the elements are connected, what type of elements they are, the material they're made of and what their physical properties are, in this case only the cross-sectional area.

The physical properties block gives the area of the two different elements and the material properties block specifies the material constants used in the elements. In this case the truss is completely made from steel with a modulus of elasticity of 30 million psi. The last block of data is for the specification of the loads acting on the structure.

The other output files give the results of the simulation. The output 3 file is shown in figure 9 and contains the displacement of node 2 as well as the internal forces in each element. We see the "hand" solution and the STARS results are identical.

NODE	X-DISPL.	Y-DISPL.	Z-DISPL.	X-ROTN.	Y-ROTN.	Z-ROTN.
1	.000000E+00	.000000E+00	.000000E+00	.000000E+00	.000000E+00	.000000E+00
2	.832103E-03	-.125031E-03	.000000E+00	.000000E+00	.000000E+00	.000000E+00
3	.000000E+00	.000000E+00	.000000E+00	.000000E+00	.000000E+00	.000000E+00
4	.000000E+00	.000000E+00	.000000E+00	.000000E+00	.000000E+00	.000000E+00

MAXIMUM DEFORMATION IN THE X-DIRECTION

NODE	DISPLACEMENT
2	.832103E-03

MAXIMUM DEFORMATION IN THE Y-DIRECTION

NODE	DISPLACEMENT
2	-.125031E-03

Figure 9 Partial Listing of STARS Output for Example Truss Problem

The Partnership Awards Grant from NASA Dryden to CSULA has created opportunities for both research and teaching, as related to the Multi-Disciplinary STARS computer simulation program. Both undergraduate and graduate courses are being developed and taught in the Mechanical Engineering Department at California State University, Loa Angeles. The course curriculum starts with basic FEM theory and ultimately progresses to the highly technical field of multi-disciplinary computer simulation. Course material is based on lectures, example problems and real world problems taken from both NASA and the Aerospace Industry.

EFFECT OF UNCERTAINTY OF IDENTIFIED PARAMETERS ON STRUCTURAL DAMAGE DETECTION

Mario Davidson, Jiann-Shiun Lew, and S. Sathanthan
NASA Center for Automated Space Science
Tennessee State University
Nashville, TN 37203-3401

ABSTRACT

This paper presents a study of the effect of the uncertainty of identified parameters on the damage detection of structure. The identified parameters, which are obtained by applying system identification algorithms to analyze the time-domain response data, have uncertainty due to noise, environmental change, etc. The damage of some structural elements may have less effect on the identified parameters than noise or environmental change, so some damage situations cannot be distinguished under these circumstances. An algorithm is presented to determine whether the considered damage can be distinguished. A twenty-bay truss structure is used to demonstrate and verify the study.

KEYWORDS: Identified parameter uncertainty, System identification, Structural damage, Finite element analysis.

INTRODUCTION

The damage detection of large structures, such as space station and bridges, is an important and challenging task for the researchers in structural dynamics and system identification communities. In the last decade, various approaches has been proposed [1-3]. All these approaches are based on the changes of the identified parameter between the damaged system and undamaged system.

The identified parameters of a structure vary with the environmental uncertainty and the structural damage [1]. The environmental uncertainty includes temperature change, noise, equipment error, etc. The first step of the damage detection is to identify the status of the structural damage. The uncertainty of the identified parameters due to the environment change may have more significant effect on system than some minor structural damages and these kinds of damages can not be identified. The uncertainty of the identified parameters from environmental uncertainty can be obtained by testing the system under various circumstances. In this paper, the simulation based on the finite element model of a twenty bay truss structure is used for the study. We use many sets of noisy input/output data with various process and measurement noises to simulate the environmental uncertainty. Then we apply the system identification algorithm [4] to process each

set of input/output data to obtain an identified model. The identified parameters of all these identified models are used to generate the bounds of each parameter to represent the uncertainty for that parameter. This process is like Monte Carlo simulation. In this paper, the considered damages are various stiffness losses of the element of this twenty-bay truss structure. An algorithm is presented to check whether the considered damage can be distinguished from the environmental uncertainty.

TRUSS STRUCTURE

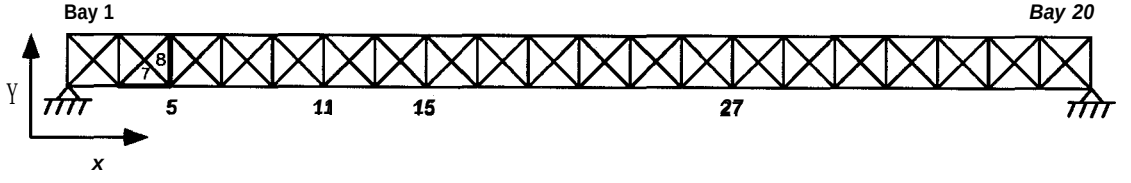


Figure 1: Twenty bay truss.

The finite element model [5] of a 20-bay truss structure as shown in Figure 1 is the system used for the study. This structure is freed to the ground at both ends. The structure has 101 aluminum tube elements with cross-sectional area of 5.97×10^{-5}, and each nondiagonal element is one-meter long. The mass of the fastener, which is used to connect elements, at each node is 0.60374Kg. Two 10Kg added masses are located at nodes 15 and 27, respectively. The natural frequencies of the healthy structure of the first three modes are 3.86, 11.35, and 31.80 Hz, respectively. In this paper, the finite element model of these first three modes is used.

In this study, an actuator located at node 11 is used to excite the vibration in the y-direction and a sensor at node 5 is used to measure the displacement in the y-direction. The time domain response data are simulated by the state space model

$$x(k+1) = Ax(k) + Bu(k) + \mathrm{v(k)} \tag{1}$$

$$y(k) = Cx(k) + \mathrm{w(k)} \tag{2}$$

where [A B C] represents the discrete state space model of the first three modes, u is the actuator input, y is the displacement output. The variables v and w are process and measurement noise, respectively, and are assumed Gaussian, zero mean signals. The input u is also chosen as Gaussian, zero mean signal. For the added noise, the standard deviation of the measurement noise is 2% of the magnitude of the noise free output. The standard deviation of the random process noise is 1% of the magnitude of the corresponding state. The identified parameters are obtained by applying the system identification software, which is developed in NASA Langley Research Center [4], to analyze the input/output data. The exactly model can be identified by using this identification algorithm to process the noise free input/output data. Next the effect of

the added noise on the identified parameters of the healthy structure is studied by using Monte Carlo simulation.

MONTE CARLO SIMULATION

There are one hundred runs of Monte Carlo simulation. Each run uses different random input, process noise, and measurement noise. For each run, the input/output data with 6000 samples at 100 Hz sampling rate are used to obtain the identified parameters as the coefficients of the transfer function

$$g_i(s) = \frac{a_{i1}}{s^2 + \omega_{i1}^2} + \frac{a_{i2}}{s^2 + \omega_{i2}^2} + \frac{a_{i3}}{s^2 + \omega_{i3}^2} \tag{3}$$

where a_{ij} is the amplitude of the jth natural frequency ω_{ij} of the identified transfer function $g_i(s)$ for the ith run. The identified parameters are used to obtain the parameter bounds as

$$a_j^+ = \max\{a_{1j}, a_{2j}, \cdots, a_{100j}\}, \quad \omega_j^+ = \max\{\omega_{1j}, \omega_{2j}, \cdots, \omega_{100j}\} \tag{4}$$

$$a_j^- = \min\{a_{1j}, a_{2j}, \cdots, a_{100j}\}, \quad \omega_j^- = \min\{\omega_{1j}, \omega_{2j}, \cdots, \omega_{100j}\} \tag{5}$$

The results based on the 100 Monte Carlo runs are listed in Table 1

Table 1: Parameter bounds of Monte Carlo simulation.

$(\omega_1^+)^2$	$(\omega_2^+)^2$	$(\omega_3^+)^2$	a_1^+	a_2^+	a_3^+
5.8893×10^2	5.0921×10^3	3.9935×10^4	4.894×10^{-3}	1.674×10^{-2}	4.619×10^{-2}
$(\omega_1^-)^2$	$(\omega_2^-)^2$	$(\omega_3^-)^2$	a_1^-	a_2^-	a_3^-
5.8570×10^2	5.0847×10^3	3.9917×10^4	4.443×10^{-3}	1.545×10^{-2}	3.738×10^{-2}

RESULTS OF STRUCTURAL DAMAGE

Here we consider the damaged systems with various stiffness losses at element 7, or 8. The stiffness losses for each element are from 0% to 100%, Tables 2 and 3 show the results of identified parameters for the noisy cases with various stiffness losses. In Table 2, 'True' represents the original finite element models without noise for the structure with various stiffness losses at element 7, 'Noise' represents the identified results obtained by analyzing the noisy data of the system with various stiffness losses. The variable a_i represents the amplitude of the identified ith natural frequency ω_i. From Tables 1 and 2, the following observations are noted:

1. Table 1 shows that the identified amplitude a_i is much more sensitive to noise than the identified natural frequency ω_i for each mode, since natural frequencies represent the system characteristics.

Table 2: Identified parameters f(beam 7 stiffness loss.

	0%	20%	40%	60%	80%	100%
ω_1^2 Noise	5.8799×10^{2}	5.8530×10^{2}	5.8201×10^{2}	5.7563×10^{2}	5.6674×10^{2}	5.3484×10^{2}
True	5.8759×10^{2}	5.8549×10^{2}	5.8235×10^{2}	5.7710×10^{2}	5.6653×10^{2}	5.3442×10^{2}
ω_2^2 Noise	5.0884×10^{3}	5.0754×10^{3}	5.0541×10^{3}	5.0186×10^{3}	4.9540×10^{3}	4.7493×10^{3}
True	5.0880×10^{3}	5.0745×10^{3}	5.0543×10^{3}	5.0205×10^{3}	4.9528×10^{3}	4.7497×10^{3}
ω_3^2 Noise	3.9925×10^{4}	3.9777×10^{4}	3.8997×10^{4}	3.7179×10^{4}	3.3885×10^{4}	2.6875×10^{4}
True	3.9926×10^{4}	3.9777×10^{4}	3.8998×10^{4}	3.7186×10^{4}	3.3892×10^{4}	2.6875×10^{4}
a_1 Noise	4.719×10^{-3}	4.505×10^{-3}	4.586×10^{-3}	4.390×10^{-3}	4.442×10^{-3}	3.785×10^{-3}
True	4.678×10^{-3}	4.642×10^{-3}	4.588×10^{-3}	4.499×10^{-3}	4.323×10^{-3}	3.808×10^{-3}
a_2 Noise	1.649×10^{-2}	1.556×10^{-2}	1.625×10^{-2}	1.658×10^{-2}	1.691×10^{-2}	1.882×10^{-2}
True	1.609×10^{-2}	1.620×10^{-2}	1.636×10^{-2}	1.662×10^{-2}	1.714×10^{-2}	1.870×10^{-2}
a_3 Noise	4.146×10^{-2}	4.772×10^{-2}	3.258×10^{-2}	2.131×10^{-2}	1.890×10^{-2}	1.427×10^{-2}
True	4.198×10^{-2}	4.747×10^{-2}	3.467×10^{-2}	2.334×10^{-2}	1.839×10^{-2}	1.372×10^{-2}

2. The values of the identified $\omega_1^2, \omega_2^2, \omega_3^2$ of all the damage cases with stiffness losses from 20% to 100% in Table 2 are outside the parameter bounds in Table 1, any of these damages can be distinguished from the noisy effect by using any of the identified natural frequencies.

3. The identified amplitude a_1 of the frost model is outside the interval (a; a_1^+) when the stiffness loss of element 7 is larger than 60%, this implies that damage can be distinguished by using a_1 when the stiffness loss is higher than 60%. For the second mode, the amplitude a_2 can identify the status of damage when the stiffness loss is higher than 80%. The amplitude a_3 of the third mode can distinguish the damage when the stiffness loss is higher than 40%.

By comparing Tables 2 and 3, the stiffness loss of element 8 has much less effect on system than the stiffness loss of element 7. Some observations based on Table 3 are as follows:

1. The identified ω_1^2 can identify the damaged status until the stiffness losses is 100%, ω_2^2 cannot identify the damaged status for the various stiffness losses of beam 8, and ω_3^2 can clearly identify the damaged status when the stiffness losses is 100%. When the stiffness loss is 80%, ω_3^2 is outside the bounds in Table 1, but it is very close to the boundary, the damaged status is not clear.

2. Only a_3 *can* identify the damage status when the stiffness loss reaches 100%.

The damaged status can be identified when any of these variables indicates the damaged situation. For example, the identified ω_3^2 for the element 7 with 20% stiffness loss indicates that the structure is damaged, even the amplitudes al and a_2 cannot identify the damaged status.

Table 3: Identified parameters for beam 8 stiff ess loss.

	0%	20%	40%	60%	80%	100%
ω_1^2 Noise	5.8823×10^{2}	5.8630×10^{2}	5.8649×10^{2}	5.8706×10^{2}	5.8609×10^{2}	5.8381×10^{2}
True	5.8759×10^{2}	5.8742×10^{2}	5.8717×10^{2}	5.8677×10^{2}	5.8699×10^{2}	5.8392×10^{2}
ω_2^2 Noise	5.0886×10^{3}	5.0890×10^{3}	5.0868×10^{3}	5.0878×10^{3}	5.0895×10^{3}	5.0884×10^{3}
True	5.0880×10^{3}	5.0880×10^{3}	5.0879×10^{3}	5.0878×10^{3}	5.0877×10^{3}	5.0873×10^{3}
ω_3^2 Noise	3.9927×10^{4}	3.9922×10^{4}	3.9915×10^{4}	3.9918×10^{4}	3.9916×10^{4}	3.9886×10^{4}
True	3.9926×10^{4}	3.9925×10^{4}	3.9923×10^{4}	3.9920×10^{4}	3.9914×10^{4}	3.9888×10^{4}
a_1 Noise	4.796×10^{-3}	4.654×10^{-3}	4.666×10^{-3}	4.510×10^{-3}	4.542×10^{-3}	4.682×10^{-3}
True	4.678×10^{-3}	4.678×10^{-3}	4.678×10^{-3}	4.677×10^{-3}	4.676×10^{-3}	4.675×10^{-3}
a_2 Noise	1.603×10^{-2}	1.610×10^{-2}	1.624×10^{-2}	1.631×10^{-2}	1.583×10^{-2}	1.602×10^{-2}
True	1.609×10^{-2}	1.609×10^{-2}	1.609×10^{-2}	1.610×10^{-2}	1.610×10^{-2}	1.610×10^{-2}
a_3 Noise	4.311×10^{-2}	3.977×10^{-2}	4.573×10^{-2}	4.058×10^{-2}	4.108×10^{-2}	3.593×10^{-2}
True	4.198×10^{-2}	4.187×10^{-2}	4.170×10^{-2}	4.140×10^{-2}	4.071×10^{-2}	3.777×10^{-2}

CONCLUSIONS

A novel study of the effect of the uncertainty of the identified parameters on the damage detection of structure is presented in this paper. The process noise and measurement noise are used to represent the environmental uncertainty. The results show that the identified natural frequencies are less sensitive to noise than the identified amplitudes. Also the identified natural frequencies provide better information for distinguishing damaged status of structure.

ACKNOWLEDGMENTS

This research is supported in part by NASA Grant NCC5-228. This support is greatly appreciated.

REFERENCES

1. Lew, J.-S., "Using Transfer Function Parameter Changes for Damage Detection of Structures," *AMA Journal,* Vol. 33, No. 11, 1995, pp. 2189-2193.

2. Jauregui, D. V., and Farrar, C. R., "Comparison of Damage Identification Algorithms on Experimental Modal Data from a Bridge," *Proceedings of the 14th International Modal Analysis Conference, 1996,* pp. 1423-1429.

3. Pandey, A. K., Biswas, M., and Samman, M., "Damage Detection from Changes in Curvature Mode Shapes," *Journal of Sound and Vibration,* Vol. 145, No. 2, 1991, pp. 321-332.

4. Juang, J.-N., Phan, M., and Horta, L. G., "User's Guide for System/Observation/Controller Identification Toolbox," NASA TM 107566, 1992.

5. Craig, R. R., *Structural Dynamics: An Introduction to Computer Methods,* Wiley, New York, 1981.

A NEW APPROACH TO OVERCOMING SPATIAL ALIASING IN STRUCTURAL DAMAGE DETECTION

Mark J. Schulz, Ahmad S. Naser, Sunil K. Thyagarajan, and Travluss Mickens
Structural Dynamics and Control Laboratory
Department of Mechanical Engineering
North Carolina A&T State University
Greensboro, NC 27411

P. Frank Pai
Department of Mechanical and Aerospace Engineering
Structural Mechanics and Control Laboratory
University of Missouri - Columbia
Columbia, MO 65211

ABSTRACT

Aircraft, reusable launch vehicles, unmanned aircraft, and other advanced structures are being built using lightweight composite materials/metals with design safety factors as low as 1.25. These advanced structures operate in uncertain and severe environments and are susceptible to damage such as delamination, fiber/matrix damage, hydrothermal strain in composite materials, and fatigue and cracking in metals. To ensure human safety and load-bearing integrity these structures must be inspected to detect and locate often invisible damage and faults before they become catastrophic.

Conventional methods of non-destructive evaluation sometimes miss significant damage and are time consuming and expensive to perform. In contrast, vibrometry or vibration signature techniques are a global method of structural integrity monitoring that potentially can efficiently detect damage on large structures, including damage that is away from sensor locations, and in the interior of structures. However, a barrier problem in damage detection using vibration measurements is the need to measure the vibration response at a large number of points on the structure. Typically, model reduction or expansion procedures such as Guyan reduction or dynamic expansion are attempted to overcome the problem of insufficient measurements or spatial aliasing. These approaches depend on using information from the healthy model, and thus put error into the reduction used to represent the damaged structure.

In this paper, a Frequency Response Function technique is used to detect damage to a fixed-free beam. The technique uses measured frequency response functions from the healthy structure as reference data, and then monitors vibration measurements during the life of the structure to detect damage. In an analytical simulation using a finite-element model of a beam, damage was located using only sparse measurements because the technique uses both rotation and translation measurements from the damaged structure.

1. INTRODUCTION

When using vibration measurements, damage can often be detected using a small number of sensors. However, a large number of degrees-of-freedom (dofs) need to be measured to locate damage[1-3]. Thus, to be most useful, algorithms must be developed to use incomplete measurements or else approximate unmeasured coordinates to locate damage. It has been shown [4] that approximating unmeasured coordinates puts error in the damage diagnosis while using a small number of measurements on a large structure causes spatial aliasing of the damage location due to coarse resolution of the structural spatial response. In this paper a damage detection method is developed based on measuring the Frequency Response Functions (FRFs) of the healthy structure. This approach allows the system FRF equations to be partitioned and inverted to get the system matrix of a reduced measurement set, and then form a damage vector to locate damage. This approach does not use a finite-element model of the structure and no model reduction or expansion are needed.

2. DAMAGE DETECTION THEORY

It is impractical to use a finite-element model to diagnose damage on most structures because the model would not be accurate enough, and geometry changes due to cracks and nonlinear effects would be difficult to include. The method proposed in this paper uses FRF measurements from the healthy structure to identify an input-output model of the structure, where the number of model dofs is equal to the number of accelerometers or sensors on the structure. This approach is suitable for detecting and locating damage, but the level and type of damage is not diagnosed.

The linear structure is assumed to be represented by:

$$M\ddot{x} + C\dot{x} + Kx = f(t) \tag{1}$$

Taking the Fourier transform of (1) gives:

$$A(iw)x(iw) = f(iw) \tag{2}$$

where the system matrix is $A(i\omega) = (K - \omega^2 M + i\omega C)$. Solving for the displacement vector gives:

$$x = Hf \tag{3}$$

where $H = A^{-1}$ is the system FRF matrix and the $(i\omega)$ argument has been dropped for brevity. Equation (3) uses all Dofs of the system to detect damage, which is impractical. Thus the damage detection algorithm must use incomplete measurements or else approximate unmeasured coordinates. Typically, a model reduction procedure such as Guyan reduction or dynamic reduction/expansion is used to eliminate or approximate the unmeasured coordinates based on a model of the structure in the healthy condition. However, the reduction/expansion will almost always include dofs near the damage. This causes an incorrect damage indication because the reduction/expansion based on the healthy model is incorrect for the damaged structure, The approach taken here to the problem of incomplete measurements is to eliminate the unmeasured dofs from the problem, and measure transverse displacements and estimate rotations using finite-differences only at the sensor locations. This will provide a repeatable approximation that

includes the effects of damage and eliminates the error caused by using information from the healthy model to perform a reduction on a damaged structure. In other words, the finite-difference approximation depends only on the geometry of deformation, not the model properties, and this improves the accuracy of damage detection. The procedure for damage detection using sparse measurements is developed by partitioning the system into measured and unmeasured dofs as;

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{4}$$

where x_1 represents the dofs where the response is measured and x_2 represents the dofs where the response is not measured. The procedure assumes that f_1 is known and f_2 = O in the bandwidth where we are looking for damage. This could be accomplished by using actuators at the dofs where measurements are taken. Then the first equation in (4) becomes:

$$x_1 = H_{11} f_1 . \tag{5}$$

The damage location procedure to be developed requires that the *H* matrix be identified for the healthy structure in the bandwidth used to detect damage to the structure. The dimension of x is *nx1* for *n* measurements on the structure, and thus the *nxn H* matrix must be determined experimentally for the healthy structure. The individual *M*, C, and *K* matrices do not need to be known.

To identify the *H* matrix for the healthy structure, a force is applied at each dof one at a time, and the responses are measured and Fourier transformed and the FRFs are computed by solving (5). Each column and hence the full *H* matrix can be identified in this way for the healthy structure. Note that the *H* matrix is symmetric unless gyroscopic forces are present. A practical problem with computing the *H* matrix is that it is difficult to apply moments to obtain the FRFs for rotational dofs, and it is difficult to measure rotations. Once *H is* determined, the A_{11} matrix is calculated as $A_{11} = H_{11}^{-1}$ at each frequency point in the bandwidth that is used to detect damage, and the A_{11} matrices are stored as historical data. Now a damage vector is defined as:

$$d_1 = A^h x_{11}^d - f_{11}^d \tag{6}$$

where *h* and *d* represent the healthy and damaged structure. If damage occurs the d_1 vector will have non-zero forces only at the dofs that are connected to the damaged elements. For example, if damage causes changes to the *i*th row of the stiffness matrix, only the *i*th entry of *d* will be nonzero. To locate damage a damage matrix is formed as:

$$D = \lim_{T_o \to \infty} \frac{2}{T_o} \mathrm{E}(d_1 d_1^*) \tag{7}$$

where E is the expectation. The diagonal entries of *D are* the auto-spectral densities of the damage force. The integral of the diagonal entries of *D is* the rms damage at each of the measured dofs, and is used to locate damage. The RMS damage forces are computed as:

$$d_{rms} = \int_0^\infty diag(D)df \ . \qquad (8)$$

Global changes in the structural properties due to environmental effects such as changes in temperature and pressure will be interpreted as damage in (8) and this problem is being studied using normalization and frequency shifting. The advantage of this approach for damage detection is that after surveying the healthy structure, only the random structural accelerations and excitation forces need to be measured to locate damage. The technique is also simple and suitable for continuous damage monitoring during operation of the structure. Based on (8), damage can be located using only one input point and all response measurements at the sensor locations (i.e. by identifying one column of the reduced FRF matrix).

3. DAMAGE DETECTION SIMULATION

The damage detection technique uses the input-output FRFs measured on the healthy structure, and no model of the structure is used. A FEM model in this example is used only to test the health monitoring system and for guiding the design of the system. The cantilever beam model shown in Figure 1 is used for the simulation. The beam is 44 inches long, 0.25 inches thick, and 3 inches high. The planar model has 18 elements and 36 Dofs, with translation and rotational Dofs at each node. An excitation force is applied perpendicular to the beam at DOF 19 in the model. The response is computed for the frequency range of 1 Hz to 500 Hz. Damage is modeled as a **5°/0** reduction in the elemental stiffness matrix individually at each element in the model. Damping is modeled using a frequency dependent damping ratio method. This overcomes the problem of using modal damping which couples all the Dofs in the model and causes the damage to be smeared over the whole structure, or the problem of using proportional damping that causes the damping to be too large at high frequencies.

The RMS damage vector (8) is used to locate damage using sparse measurements. Damage (5% reduction in stiffness) is now put into all elements of the beam individually, starting with element 1 at the support. The two dofs at every fifth node, that is dofs 9,10, 19,20, and 29,30 are used as the reduced model for the damage calculation. The FRFs in the" healthy condition are computed using the full model and the FRFs at the reduced dofs are used as historical data. The RMS damage vector is calculated for each case and is shown in Figure 2. Note that for all 18 cases, damage is correctly bounded between the two closest measurement nodes, and the DOFs at nodes not nearest the damage have a zero damage force. Thus the RMS damage vector (8) can correctly detect damage, and also locate damage between the two nearest measurement points, The first six natural frequencies of the beam are not very sensitive to small damage, and changed less than 1 % due to 5% damages at any of the elements. Thus changes in the lower natural frequencies are not large enough to indicate small damages. The auto-spectral densities of the damage force were plotted for all elements and have peaks at the resonant frequencies of the beam. This simulation has shown that with sparse measurements the damage location is bounded between two closest nodes at which measurements are taken, but the exact location of damage cannot be determined. When the simulation was performed using only one DOF per node, either translation or rotation, the damage location was not always correct. Thus both translation and rotational measurements are needed at each node to locate damage. Note that the rotational Dofs have a larger damage force than translation Dofs.

Damage was also put into all elements of the beam individually, starting with element 1 at the support, and all Dofs of the beam are used as measurements in the damage algorithm. The damage is correctly located for all cases. The damage force is maximum for the element near the support, and decreases as the damage is moved toward the free end. The damage force is greater when fill measurements are used. Thus the damage force for partial measurements is an indicator of damage severity, but is not the exact damage force. Since the damage vector method uses both translation and rotational FRFs, this requires a new technique to accurately measure rotations. Approaches are being investigated to overcome the difficulties of making these measurements.

4. CONCLUSION

A method for locating damage using sparse measurements has been presented and demonstrated using a FEM of a cantilever beam. The technique is being tested in the laboratory and if a method to apply moments and to make accurate rotational measurements can be developed, the technique may be one solution to the long standing problem of not having sufficient spatial measurements to locate damage. A subtle point in using this technique is that the excitation force can only be applied at the dofs being measured. Otherwise the damage location will be incorrect. For the case where there is unmeasured natural excitation over the entire structure, artificial forces would have to be applied at one or more of the measured dofs at a higher frequency range that the natural excitation. Alternatively, the artificial excitation must be uncorrelated with the natural loading and the technique extended to use the correlation between the artificial excitation and the response.

ACKNOWLEDGEMENT

This work was partially sponsored by the NASA Center for Aerospace Research at North Carolina A&T State University, under grant NCC- 1-255. This support is gratefully acknowledged.

REFERENCES

1. James, G.H. III, *"Development of Structural Health Monitoring Techniques Using Dynamic Testing, "* SANDIA REPORT: SAND96-0810, UC-706, April 1996.

2. Doebling, S. W., C.R. Farrar, M.B. Prime, and D.W. Shevitz. *"Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in their Vibration Characteristics: A Literature Review, "* Los Alamos National Laboratory Report LA-13070-VA5, 1996.

3. James, G. H., D.C. Zimmerman, C.R. Farrar, and S.W. Doebling, *"Current Horizon for Structural Damage Detection Course,"* Short Course Notes, Orlando, Fla., Jan. 30-Feb. 1, 1997.

4. Schulz, M. J., Pai, P.F., and Abdelnaser, A. S., *"Frequency Response Function Assignment Technique for Structural Damage Identification,"* IMAC-XIV Conference, Dearborn, Michigan February 12-15, 1996.

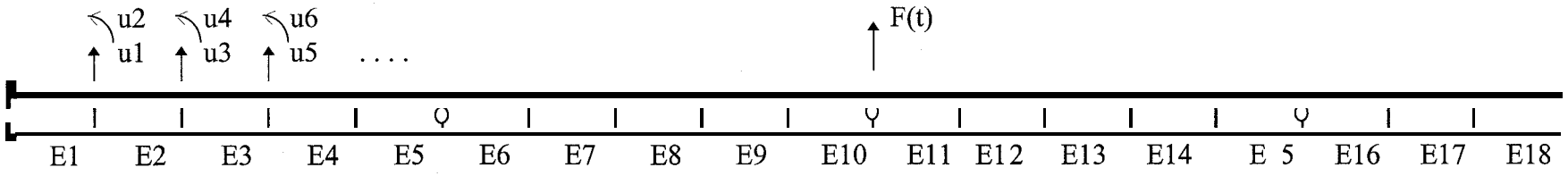


Figure 1. Fixed-free beam element numbers and nodal DOFs for simulation (O= sensor locations)

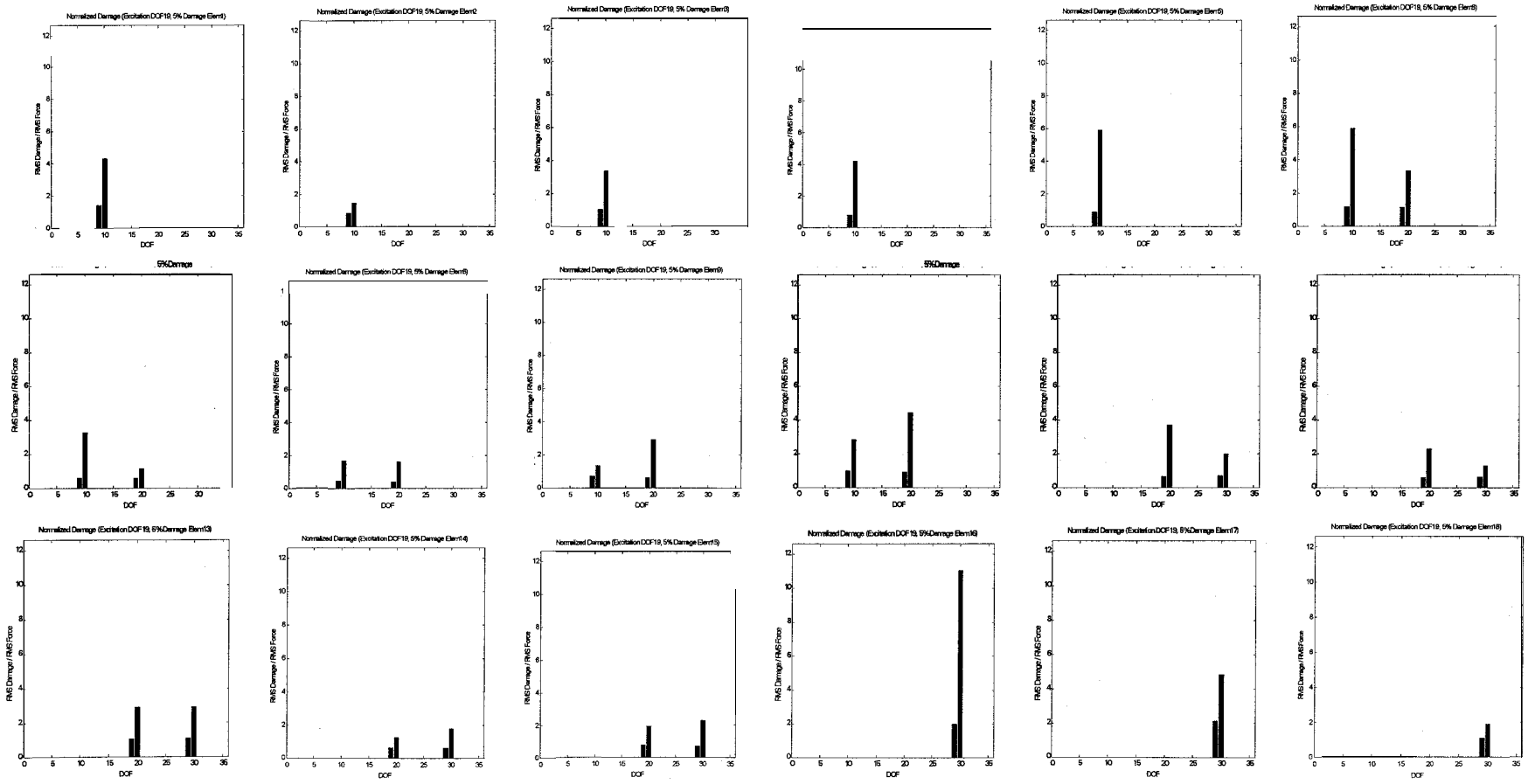


Figure 2. Peak damage indicator for 5% damage to each element individually (and 1-500Hz freq. range)

DAMAGE DETECTION AND IMPACT TESTING ON LAMINATED AND SANDWICH COMPOSITE PANELS

Derke R. Hughes[1], William J. Craft[2], Mark J. Schulz[3], Ahmad S. Naser[4], and William N. Martin[5]
Structural Mechanics and Control Research Laboratory
Department of Mechanical Engineering
North Carolina A&T State University
Greensboro, NC 27411

Abstract

This research investigates "health monitoring" of sandwich shell composites to determine if the Transmittance Functions (TF) are effective in determining the present of damage. The health monitoring test was conducted on the sandwich plates before and after low velocity impacts using the health monitoring technique given in [1]. TFs are a NDE technique that utilizes the ratios of cross-spectrums to auto-spectrums between two response points on the sandwich composites.

The test for transmittance was conducted on the same density foam core throughout the experiment. The test specimens were 17.8 cm by 25.4 cm in dimension. The external sheets (face sheets) were created from graphite / epoxy laminate with dimension of 1.58 mm thick. The polymethacrylide (Rohacell) foam core was 12.7 mm thick. These samples experienced a transformation in the TF that was considered the low velocity impact damage. The low velocity damage was observed in the TFs for the sandwich composites shown in Table 1.

Introduction

While testing the sandwich panels for maximum threshold loads the densities of the sandwich plates were also changed during the experiment. In order to obtain the maximum loads the test specimens were able to support, the drop heights were increased to raise the energy levels and velocities. This procedure was performed to determine the impact resistance of the sandwich foam cores [1]. Before the experiments, the common NDE method for determining the extent of impact damage is C-SCAN. While this method was utilized during the low velocity experiment, the emerging field of "health monitoring" has provided another means for NDE. Transmittance Functions (TF) for non-invasive inspection was considered an excellent option to evaluated the panels beyond a surface assessment. The TFs use the Frequency Response Functions (FRF) as a signature of the entire sandwich plate. Since TFs incorporate the FRFs, the TFs assess the overall damage throughout the sandwich structure[2].

[1]Ph.D. Student, [2]Professor and Department Head, [3]Assistant Professor, [4]Research Professor, [5]Graduate Student

Experimental Procedure

Low Velocity Impact Test

A substantial amount of work on impact damage on laminated composites was performed at Wright-Patterson Flight Dynamic Laboratory. The laboratory contains a Dynatup precision drop weight system. The machine is capable of varying the amount of energy which will cause initial failure or catastrophic damage within a particular specimen. The limitations for the energy levels are caused by the fixed sliding arm height and the amount of weight applied to the chassis of the penetrator. These limitations do not interfere with either determination of the amount of energy required to initiate failure or a detection of the maximum energy absorbed by the tested specimens.

Health Monitoring Vibration Test

In health monitoring, the objective is to determine the transmittance functions of the a healthy structure versus a damage structure. The transmittance function (TF) is defined as the complex ratio between Fourier transforms of response points on the structure and for damage detection. The research uses the ratio of the cross-spectrums to the auto-spectrums. The algorithm for damage detection is shown in the schematic in Figure 1. In Figure 1, the finite Fourier Transform of the structural acceleration vector records the length To which gives

$$\mathbf{a}(\omega, To) = -\omega^2 \mathbf{H}(\omega)[\mathbf{f}_{piezo}(\omega, T_o) + \mathbf{f}_{nat}(\omega, T_o)1$$

where **a** is the *nx1* acceleration vector, $\mathbf{H} = (\mathbf{K} - \omega^2\mathbf{M} + i\omega\mathbf{D})^{-1}$ is the *nxn* receptance frequency response function matrix. $i = \sqrt{-1}$ $\omega = 2\pi f$ is frequency in rad/sec. **M, D,** and **K** are the mass, damping and stiffness matrices of the structure, $\mathbf{f}_{excitation}$ is the loading vector. T_{rs} are the transmittance function.

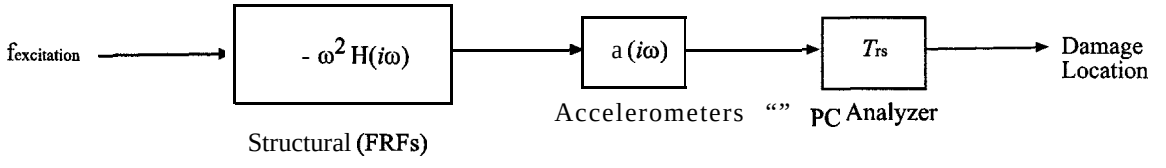


Figure 1. Schematic of the General Transmittance Function Monitoring System

After some derivation, the transmittance function can be defined as follows:

$$T_{rs} = \frac{G_{rs}}{G_{ss}} = \frac{\mathbf{h}_r^* \mathbf{h}_s}{\mathbf{h}_s^* \mathbf{h}_s}$$

which is the ratio of the cross-spectrum of the response at points r ands to the autospectrum of the response at reference points. Note, the transmittance functions are also a function of the columns of the frequency response function matrix.

Results

The primary objective of health monitoring is determining damage and where the damage is located. During the impact test, the Dynatup machine's energy levels were changed as a result of adjusting the drop height. From health monitoring test, the healthy data was obtained through an examination of the transmittance signal before the low velocity impact test. After the low velocity impacts, the damage readings were acquired from the sandwich panels. In Figure 2, the location of the impact area and accelerometers is illustrated.

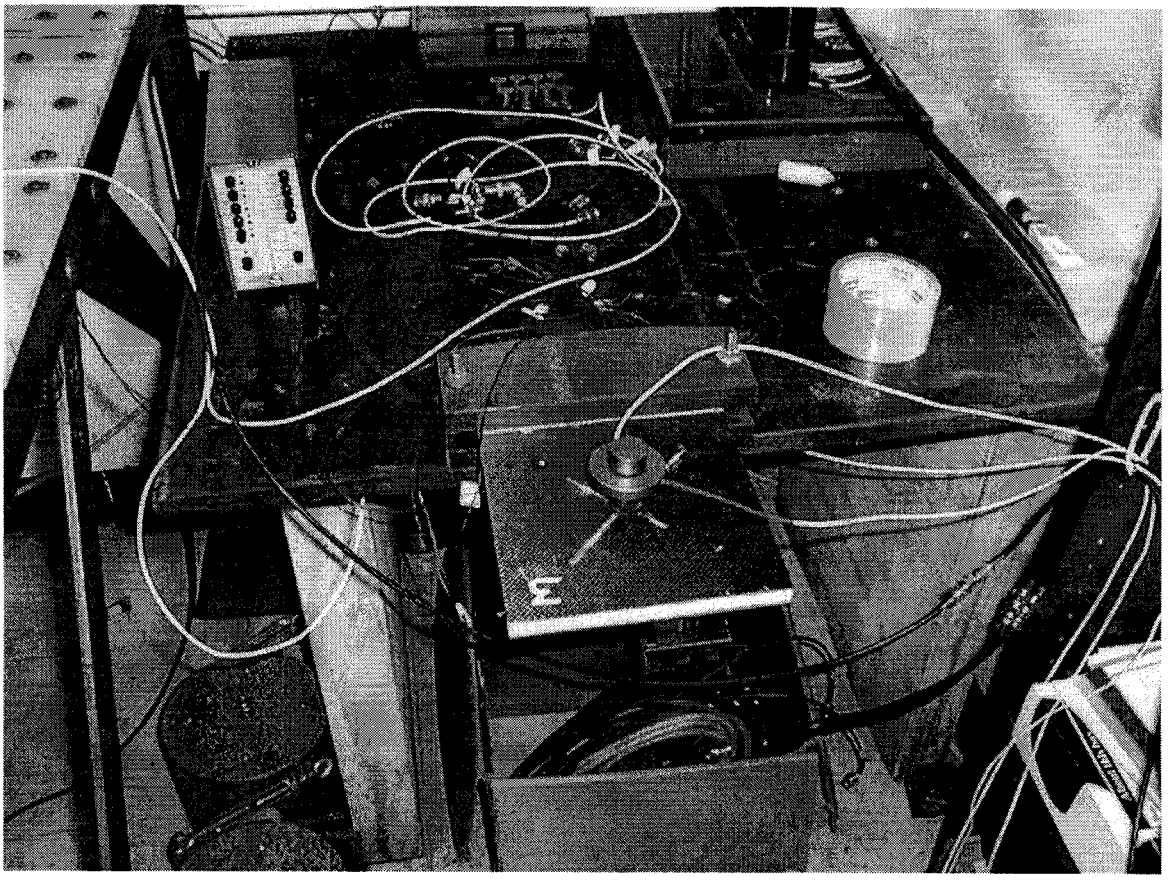


Figure 2. Illustration of the Health Monitoring Test on Sandwich Panel

After using actuator, the transmittance functions were calculated for the healthy and damaged results. Once the accelerometers were mounting on the back of the specimens where the low velocity impacts occurred, the sandwich plates were tested from the top by a random signal from the actuator. The first accelerometer was used as the reference to the other accelerometers. Also, the healthy and damage transmittance functions results form a matrix with respects to the measurement and the excitation node or accelerometer. The healthy and damage matrices T^h and T^d, respectively, were computed in order to find the overall damage matrix.

$$D = \int_{\infty} | T^h - T^d | df \; / \; \int_{\infty} | T^h | df$$

The D represents the normalized damage matrix. The results from the sandwich panels are shown below in Table 1.

Table 1. The D matrix for the tested Sandwich Plates

Specimen #	Drop Height (in)	D12	D13	D14	Max, Load (lb) from Drop Height
1	3	1.6084	**2.4501**	**2.0211**	**574**
2	**6**	**0.4274**	**0.6187**	**0.5552**	**639.6**
3	**9**	**0.4149**	**0,515**	**0.4733**	**663.5**
4	**12**	**0.3984**	**0.6181**	**0.6119**	**673.9**
5	**15**	**0.5136**	**0.6688**	**0.787**	**725.9**
6	**2** in/rein	**0.5338**	**0.6904**	**0.5238**	**590**
7	**5 in/rein**	**0.5933**	**0,6662**	**0.5465**	**612**

The damage matrices results for each tested panel are shown with respect to accelerometer one to the other reference accelerometers. For instance, the D 13 refers to the measurement obtained from the first node in reference to the third node.

Graphs are plotted for the damage to damage transmittance functions in order to illustrate the repeatability of the transmittance functions. The graphs of the second and three specimen are presented in the following Figures 3 and 4. The solid line indicates healthy and dash-line damaged. In Figures 3 and 4, the graphs illustrate the magnitude of the error of the transmittance results, and Table 2 presents the data for specimen 2 and 3.

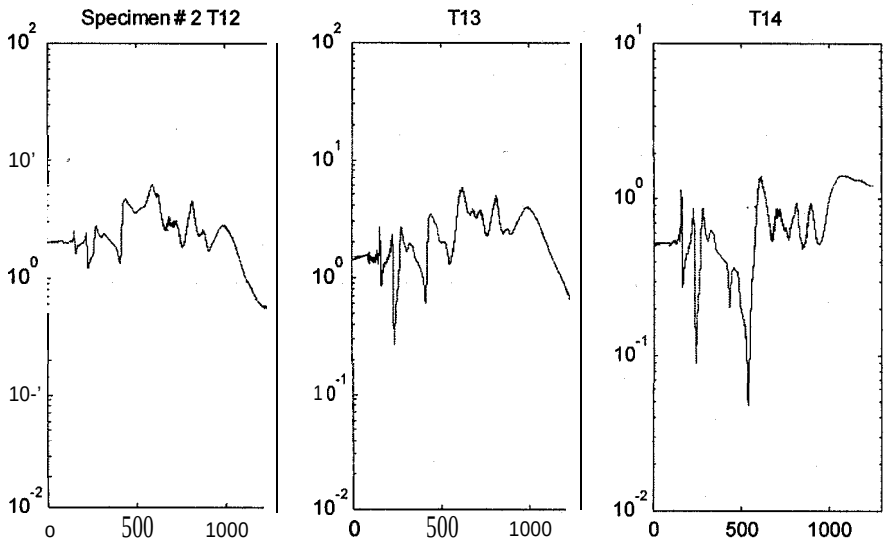


Figure *3.* Specimen 2 Damage to Damage Transmittance

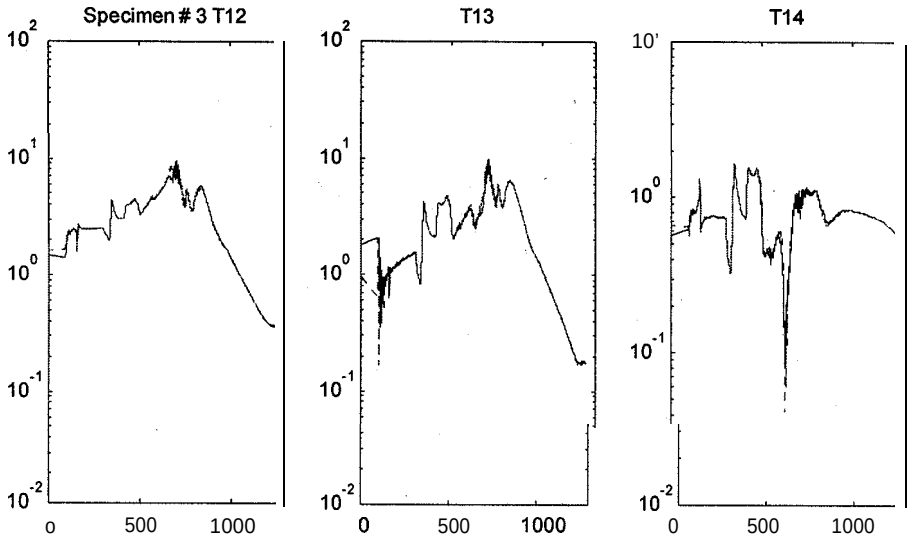


Figure *3.* Specimen 3 Damage to Damage Transmittance

Table 2. The Error of Damage to Damage Transmittances Functions

Specimen 2	Trial 1	D12: 0.0215	D13: 0.2152	D14: 0.0286
	2	D12: 0.0167	D13: 0.0188	D14: 0.162
Specimen 3	Trial 1	D12: 0.0562	D13: 0.1109	D14: 0.0467
	2	D12: 0.0535	D13: 0.0686	D14: 0.0400

The increased damage of D13 is just inconsistency due to the damage such as delamination and debonding. In Specimen 2 and 3, trial 1 and trial2 demonstrate the error introduce into the transmittance by the actuators low natural frequencies which creates the resonant error between 0-300 frequency range.

Conclusion

The error test must be conducted in order to quantify the amount of discrepancy which exist between the healthy and damage values. When comparing the transmittance values of Table 1 and 2, an order of magnitude difference exist between the damage to healthy values and the damage to damage results. Improvements could be made on the number of samples during the experiment.

Additionally, the number of specimen sampled should increase to properly verify the accuracy of the graphical results. The D matrix values for the specimen indicate damage in the sandwich plates, which is also illustrated in Table 1 values.

Acknowledgment

This work was partially sponsored by the NASA Center of Aerospace Research at North Carolina A&T State University, under grant NCC-1-255. This support is gratefully acknowledged.

Reference

1. Hughes, D. R., " Low Velocity Impact Resistance of Organic Foam Core Sandwich Composites," M.S. Thesis, North Carolina A&T State University, North Carolina, 1997.
2. Schulz, M. J., Pai, P. F., and Naser, A. S., " Structural Damage Detection Using Transmittance Functions," IMAC-XV Conference, Feb., 1997.

Fuzzy Versions of Epistemic and Deontic Logic'

Ramasamy S. Gounder and Albert C. Esterline
The NASA Center for Autonomous Control Engineering
Department of Computer Science
North Carolina A&T State University
Greensboro, NC 27411

1. INTRODUCTION

Epistemic and deontic logics are modal logics, respectively, of knowledge and of the normative concepts of *obligation, permission,* and *prohibition.* Epistemic logic is useful in formalizing systems of communicating processes [4] and knowledge and belief in AI. Deontic logic is useful in computer science wherever we must distinguish between actual and ideal behavior [9], as in fault tolerance and database integrity constraints. We here discuss fuzzy versions of these logics. In the crisp versions, various axioms correspond to various properties of the structures used in defining the semantics of the logics. Thus, any axiomatic theory will be characterized not only by its axioms but also by the set of properties holding of the corresponding semantic structures. Fuzzy logic does not proceed with axiomatic systems, but fuzzy versions of the semantic properties exist and can be shown to correspond to some of the axioms for the crisp systems in special ways that support dependency networks among assertions in a modal domain. This in turn allows one to implement truth maintenance systems. For the technical development of epistemic logic, see [4], and for that of deontic logic, see [1]. For previous work on fuzzy modal logic, see [10], [8], and [6]; we follow [7]. To our knowledge, we are the first to address fuzzy epistemic and fuzzy deontic logic explicitly and to consider the different systems and semantic properties available.

In section 2, we give the syntax and semantics of epistemic logic and discuss the correspondence between axioms of epistemic logic and properties of semantic structures. The same topics are covered for deontic logic in section 3. Section 4 presents fuzzy epistemic and fuzzy deontic logic and discusses the relationship between axioms and semantic properties for these logics. Section 5 discusses how our results can be exploited in truth maintenance systems.

2. SYNTAX AND SEMANTICS OF EPISTEMIC LOGIC

For the syntax of epistemic logic we start with primitive (or atomic) propositions in a set Φ and form more complicated formulas by closing under negation, conjunction, and the modal operators $K_1,\ldots,K_n$, where there are n agents named 1, 2,..., n. (Where φ is a formula, $K_i\varphi$ means that agent i *knows* φ. We use the standard symbols $\wedge$, $\vee$, $\neg$, and $\Rightarrow$ for conjunction, disjunction, negation, and implication, respectively".) Thus, if φ and ψ are formulas, then so are $\neg\varphi$, $(\varphi\wedge\psi)$, and $K_i\varphi$, for i = 1 . . . n. Other propositional connectives are defined in the usual way.

The semantics of a language gives the conditions under which the sentences of the language are true or false. We use Kripke structures as formal models for this purpose. A Kripke structure M for n agents over a set Φ of primitive propositions is a tuple $(S, \pi, \kappa_1,\ldots, \kappa_n)$, where S is a set of states or possible worlds, π is an interpretation function that associates with each state in S a truth assignment to the primitive propositions of set Φ, and κ_i, $1\le i\le n$, is a binary relation on S, that is, a set of pairs of elements of S. We shall call each κ, an *accessibility relation.* A binary relation κ is (a) *reflexive* if, for all $s\in S$, we have $(s,s)\bullet L$ (b) *symmetric* if, for all s, $t\in S$, we have $(s, t)\in\kappa$ if and only if $(t,s)\in\kappa$, and (c) *transitive* if, for all s, t, $u\in S$, we have that, if $(s, t)\in\kappa$ and $(t,u)\in\kappa$, then $(s, u)\in\kappa$. A relation κ that enjoys all three of these properties is called an *equivalence relation.*

Where M is a Kripke structure, s is a possible world in M, and φ is any formula, wc write $(M, s)\models\varphi$ to express that φ is true ats in M. Wc define such a notion of truth recursively as follows.

$(M, s)\models p$ iff $\pi(s)(p)$ = true, for $p\in\Phi$
$(M,s)\models\psi\wedge\varphi$ iff $(M, s)\models\psi$ and $(M,s)\models\varphi$
$(M, s)\models\neg\psi$ iff $(M, s)\not\models\psi$

[1] The first author has been partially support by grant NAG 5-4102, "Formal Foundations of Agents," form NASA/GSFC. The second author has been supported by the same grant and by grant NAG 2-1150, "Motion Planning in a Society of Intelligent Mobile Agents," from NASA/ARC.

$(M,s) \models K_i\varphi$ iff $(M,t) \models \varphi$ for all worlds t such that $(s, t) \in \kappa_i$.

In any modal logic, we say that a formula ψ is *valid in* a Kripke structure M if ψ is true at all possible worlds in M; ψ is *valid for a family* F of Kripke structures if ψ is valid in all structures in F. Families of Kripke structures are defined by the properties their accessibility relations satisfy, such as the three above. (In a given epistemic Kripke structure, we assume that all the relations satisfy the same properties.)

Modal logics are generally presented as axiomatic systems. The set of theses of the system are all the formulas that follow from the axioms by zero or more applications of the rules of inference. The axioms and rules of inference for a modal logic system are generally chosen so that its set of theses is exactly the set of formulas valid for that family of structures. For epistemic logics (and analogously, most other modal logics), we have the following rules of inference, which preserve validity:

Modus Ponens: From φ and $\varphi \Rightarrow \psi$ infer ψ.
Modal Generalization: From φ infer $K_i\varphi$

All axiomatic systems of epistemic logic include the following two axiom schemas – any instance of them is valid in any Kripke structure.

All tautologies of propositional calculus	**(Al)**
$(K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi, i = 1, \ldots . n$	**(K)**

The following axioms, however, are valid for only certain families of Kripke structures, that is, only where certain accessibility-relation properties are satisfied.

$K_i\varphi \Rightarrow \varphi$	**(T)**
$K_i\varphi \Rightarrow K_iK_i\varphi$	**(4)**
$\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$	**(5)**
$\varphi \Rightarrow K_i\neg K_i\neg\varphi$	**(S)**

T is valid in any reflexive structure, 4 is valid in any transitive structure, S is valid in any symmetric structure, and 5 is valid in any Euclidean structure. A binary relation κ on S is Euclidean if, for all $s,t,u \in S$, if $(s,t) \in \kappa$ and $(s,u) \in \kappa$, then $(t, u) \in \kappa$. Thus, the system called T, which has the usual rules of inference and axioms **Al, K,** and **T**, has as theses exactly those formulas valid for the family of reflexive structures. The theses of S4, which adds axiom 4, are those formulas valid in reflexive and transitive structures. And the theses of S5, which further adds axiom 5, are those formulas valid in reflexive, transitive, and Euclidean structures. But any binary relation that is reflexive and Euclidean is also symmetric; correspondingly, axiom S is derivable as a theses in S5. Thus the theses of S5 are the formulas valid in structures whose the accessibility relations are equivalence relations.

3. DEONTIC LOGIC

The syntax of deontic logic is just like that of epistemic except that

- K_i is replaced by O,
- no subscripts appear on modal operators (obligation, etc., apply equally to all), and
- we introduce additional modal operators P and F.

$O\varphi$ means that if φ is obligatory, $P\varphi$ means that φ is permitted, and $F\varphi$ means that φ is forbidden. In fact, one normally defines $P\varphi$ as $\neg O\neg\varphi$ and $F\varphi$ as $O\neg F$. For the semantics of deontic logic, we use Kripke structures each with only one accessibility relation, κ. The semantics for O is like that for K_i, and for P we have

$(M,s) \models P\varphi$ iff $(M,t) \models \varphi$ for some world t such that $(s,t) \in \kappa_i$

Again there are different axiomatic systems that we get by choosing different sets of axioms, and various properties of the accessibility relations correspond to various axioms. As before, the rules Modus Ponens and Modal Generalization with O in place of K are used in all systems, as are axioms **Al** and **K.** Henceforth, when referring to the deontic versions of these axioms, we shall precede the name with an "O'. The following are some of the available axioms.

$O(O\varphi \Rightarrow \varphi)$	**(OT')**

$O\varphi \Rightarrow OO\varphi$ **(04)**

$PO\varphi \Rightarrow O\varphi$ **(05)**

$O(PO\varphi \Rightarrow \varphi)$ **(OS′)**

04 and 05 are the deontic versions of **4** and 5 and correspond, respectively, to the transitive and Euclidean properties of κ. **OT′** replaces **OT**, $O\varphi \Rightarrow \varphi$, because we do not require that, if some proposition is obligatory, then it must be true. Correspondingly, we eliminate reflexive accessibility relations. The property of κ corresponding to **OT′** is that of being *almost reflexive (see* [1]). OS' replaces OS, $PO\varphi \Rightarrow \varphi$, and we correspondingly eliminate the symmetric property – a relation that is transitive and symmetric must also be reflexive (which we do not want

The system OT′ includes the usual two inference rules and the axioms **OA1** and **OK**; it also includes axiom **OT′**. System OS4 includes in addition axiom 04, and system OS5 adds axiom 05 to this. An additional axiom, $O\varphi \Rightarrow P\varphi$, states that what is obligatory is also permitted; the corresponding property of accessibility relations is *seriality (see [1]).* If we add this axiom to system OT′ (respectively, OS4 or OS5), we get a system called OT'^{+} (respectively, $OS4^{+}$ or $OS5^{+}$).

4. FUZZY EPISTEMIC AND DEONTIC LOGIC

Fuzzy modal logics give fuzzy versions of both π (the evaluation function) and κ_i (the accessibility relations). Now π, given a possible world and an atomic proposition, gives a value in the range [0,1]. Accessibility relations are now denoted by δ (or, when we have an indexed family, $\delta_i, 1 \leq i \leq n$) and δ is a function that assigns a value in [0,1] to a pair of worlds in *S*. Rather than using the |= notation , we write $V_s(\varphi)$ for the degree of truth of an arbitrary proposition ω at possible world s. We define V_s recursively. The base case and the cases for propositional connective are as follows, where A is a triangular norm and V is a triangular co-norm (see below).

$V_s(p) = n(s)(p)$, where p is an atomic proposition
$V_s(\neg\varphi) = 1 - V_s(\varphi)$
$V_s(\varphi \wedge \psi) = A\ \{V_s(\varphi), V_s(\psi)\}$
$V_s(\varphi \vee \psi) = \mathbf{v}\ \{V_s(\varphi), V_s(\psi)\}$

For illustration, we shall consistently use the minimum function for the triangular norm and the maximum function for the triangular co-norm.

For a strong modality □ (e.g., O or K_i), we take $V_s(\Box\varphi)$ to be the minimum over all worlds s' ● S of $max\{\neg\delta(s,s'), V.(q)\}$, where $\neg\delta(s,s')$ is defined as $1 - \delta(s,s')$. The equation, then, is $V_s(\Box\varphi) = \Delta_{s' \in S} \nabla \{\neg\delta(s,s'), V_{s'}(\varphi)\}$. For a weak modality O (e.g., *P* or $\neg K_i \neg$), we use $\Diamond\varphi \equiv \neg\Box\neg\varphi$ to get $V_s(\Diamond\varphi) = \nabla_{s' \in S} \Delta\{\delta(s,s'), V_{s'}(\varphi)\}$.

Because fuzzy logics work with numerical measures, axiomatic systems are not appropriate for presenting fuzzy logics. So, to capture properties of knowledge or properties of our normative concepts in fuzzy modal logics, we must in the first instance rely on the properties of the accessibility relations. Although other definitions are possible, we usc the following for the three most common properties:

δ is reflexive if $\delta(s,s) = 1$ for all $s \in S$.
δ is symmetric if $\delta(s,t) = \delta(t,s)$ for all $s,t \in S$.
δ is transitive if $\delta(s,u) \geq min\{\delta(s,t), \delta(t,u)\}$ for all $s,t,u \in S$.

Still, we would like to retain a relation between various properties of the accessibility relations and the corresponding formulas – the formulas relate more directly to our usual notions. We can no longer speak of formulas being valid for a family of structures. Instead, given a family of structures, we want some guarantee that the formulas in question are somehow distinguished. The two axioms that hold for any family of structures **(Al** and K) essentially fall back to non-modal fuzzy logic. The interesting cases are the formulas that correspond to the various accessibility-relation properties. Of the seven such formulas we have discussed, five **(T,** 4, 5, S, and 00) are conditionals, $\alpha \Rightarrow \beta$. The remaining two (**OT′** and OS') are "strong conditionals", that is, of the form (using the general strong modal operator □ in place of O) $\Box(\alpha \Rightarrow \beta)$. For **T** and 4, we can show that, as long as the corresponding property of the accessibility relation holds, at any worlds we are guaranteed that $V_s(\alpha) \leq V_s(\beta)$. The significance of these results comes out when we think of a schema $\alpha \Rightarrow \beta$ as a rule (admittedly dependent on Modus Ponens) that sanctions α as support for β. It runs out that, for 5, S, and 00, even when the corresponding property of the accessibility relation holds, we cannot in general guarantee that $V"(a) \leq Vs(\beta)$ at any world s. We have, however, found for each of these formulas conditions under which, when the property holds, so does the guarantee.

The conditions tend to be quite general, so corresponding support rules would be available in most interesting cases. Similar results hold for the strong conditional forms.

Spacc restrictions allow us to prove the guarantee $V_s(\alpha) \leq V_s(\beta)$ for only one, relatively simple case. Consider T, $\Box\varphi \Rightarrow \varphi$, and assume that δ is reflexive. We must show that, for anys $\in S$ and arbitrary S, π, and δ except that δ is reflexive, we have $V_s(\Box\varphi) \leq V_s(\varphi)$, that is,

$$\Delta_{s' \in S} \nabla \{\neg\delta(s,s'), V_{s'}(\varphi)\} \leq V_s(\varphi) \qquad (*)$$

For, since δ is reflexive, we have $\delta(s,s) = 1$ hence $\neg\delta(s,s) = O$. Hence $V\{\neg\delta(s,s), V_s(\varphi)\} = V_s(\varphi)$. Since the minimum operation ranges overs as well, the left-hand side of (*) cannot be greater than $V_s(\varphi)$ – Q.E.D.

5. Truth Maintenance Systems

Truth maintenance systems [5] are a general facility used in problem solving systems, in conjunction with inference engines, to manage logical relationships between statements. The inference engine is responsible for making inferences within the problem domain and the troth maintenance system is concerned with organizing the inferences which are consistent with each other. Every inference is communicated with the TMS by the inference engine. The consequent is the node of the inference engine datum that is infered. The antecedents are the nodes of the data used as antecedents to the inference rule. So a TMS maintains a dependency network with nodes representing *sentences* and *justifications.* A TMS is used for nonmonotonic reasoning where if we add new facts, we can still reason by removing an assumption which causes contradiction. This is called dependency driven backtracking.

There are four types of TMSs:

1. Justification-Based Truth Maintenance System (JTMS): This is a simple TMS where one can examine the consequences of the current set of assumptions. The meaning of sentences is not known.
2. Non-Monotonic Justification-based Truth Maintenance System (NMJTMS): This is much like a JTMS except that it supports non-monotonic inferences.
3. Assumption-based Truth Maintenance System (ATMS): This allows us to maintain and reason with a number of simultaneous, possibly incompatible current sets of assumption.
4. Logic-based Truth Maintenance System (LTMS): The LTMS incorporates negation explicitly and therefore can represent any propositional calculus formula.

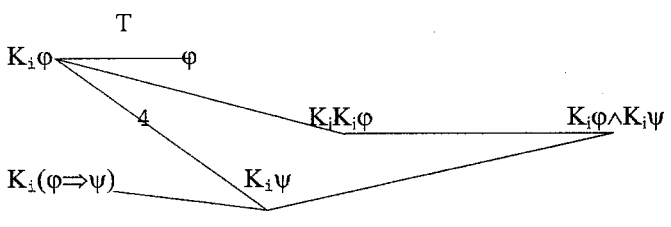
Truth maintenance systems were introduced for reasoning that uses crisp-logic. Recent research has introduced truth maintenance systems that work with multivalued logics [2]. We introduce truth maintenance systems that work with fuzzy epistemic and deontic logic since uncertainty is one of the main concerns of AI.

For example, if proposition Xp is derived from propositions X_2 and X_1, then Xp depends on X_2 and X_1. X_1 and X_2 are supporters of Xp. It is said that Xp is justified by X_1 and X_2. [X_1 X_2 Xp] is called a *justification.* If a proposition is supported by more than one justification then its justification can be expressed as a *disjunction.* Many propositions derived from problem solver create a network of propositions and justifications called *dependency network.*

Axioms of epistemic and deontic logics are used as rules for reasoning." A justification is actually an instance of axiom within the TMS. Thus, a justification can be thought of having a truth value in the interval [0, 1] based on the truth values of the assumptions. Fuzzy logic methods are used to evaluate *conjunction, disjunction* and *implication.* Certain Axioms of epistemic and deontic logics are proved to be valid using fuzzy evaluation techniques. In particular, the truth value of the consequent is shown to be greater than or equal to the truth value of the antecedent in the conditional axioms we use for justifications.

Fuzzy epistemic LTMS

The following represents dependencies that would exist in a fuzzy epistemic LTMS. For example, given that $K_i\varphi$ has a high truth value, formula **T** would ensure that φ also has a high truth value (in fact, at least as high as $K_i\varphi$). The other relations show similar support. Note that 4 requires that both $K_i\varphi$ and $K_i(\varphi \Rightarrow \psi)$ have high truth values to ensure that $K_i\psi$ has a high truth value.



Such an epistemic LTMS can be used for distributed systems where we can have many agents with common knowledge about what each one knows [4]. If we come up with a contradiction of some statement, we can backtrack and remove the assumption causing the contradiction. We can fix a cutoff value for our derived fact and, if the value is less than the cutoff value, we can retract the fact. Our fuzzy epistemic LTMS can be used for communication protocols.

Fuzzy deontic LTMS

Deontic logic is used where there is a conflict between ideal and actual behavior of a system [9]. Applications of deontic logic include policies of an organization. We can put constraints on salaries, ages, deadlines, etc. Our obligation rules (constraints) can be used for specifying those constraints. By using our fuzzy deontic LTMS, we can allow some amount of violation without any penalty. For example, a student can be allowed to return a book to the library within three days after the due date. So this is not a crisp date. These ranges of violation can be given in noncrisp values. If someone violates a rule by more than a certain extent, he or she can be penalized to that extent.
To consider some xamples, let

φ = Person p receives a driver's license.
ψ = Person p is 18 or older.
p = Person p is an employee of company c.
σ = Person p is over 80 years old.
T = Person p is under 20 years old.
μ = Company c gives its employees a bonus.
v = The employees of company c arrive at work not more than ten minutes late.

Then the following are interesting deontic statements:

$O(\varphi \Rightarrow \psi)$ which implies $O\varphi \Rightarrow O\psi$
$p \Rightarrow O(\neg\sigma \wedge \neg T)$
$\mu \Rightarrow O\nu$

These obligations need not be crisp, but can be in a permissible range which is given in the fuzzy truth value in the interval [0, 1]. One obligation may lead to another obligation. So we can derive other obligations from the antecedents which can be nodes of TMS and implications are used as rules for deriving consequent which are also denoted by nodes. Thus, TMS keeps all the facts regarding our reasoning system and removes some obligations if contradiction a occurs.

6. *CONCLUSION*

We have reviewed various systems of epistemic and deontic logic and shown how to characterize fuzzy versions of these logics. We have also reported our results that show that a special relationship often exists even in the fuzzy versions between certain axioms and certain properties of the accessibility relations. For conditional axioms $\alpha \Rightarrow \beta$ (the most important group), this is that, when the relevant property holds, the "degree of truth of the consequent is at least as great as the degree of truth of the antecedent. This allows us to consider an instance of α as supporting a corresponding instance of β. This is the foundation of a family of truth maintenance systems we are building to support inference engines in modal domains. This work complements earlier work on (non-modal) fuzzy spatial logic [3]. Other logics that could be handled in similar fashion are logics of time (usually modal) and logics of motion. The general drift is to produce fuzzy versions of logics that have rich structures and express well-defined features of domains important to computer science and AI.

REFERENCES

1. Åquist, L., "Deontic Logic", in Gobbay, D. And Guenther, F. (eds.), *Handbook of Philosophical Logic;* Vol. II, Dordrecht, The Netherlands, 1984, pp. 605-714.

2. Castro, J.L. and Zurita, J.M., "A Multivalued Logic ATMS," *International Journal of Intelligent Systems* **11** (1996), pp. 185-195.

3. Esterline, A., Dozier, G., and Homaifar, A., "Fuzzy Spatial Logic," *Proc. 1997 Int. Fuzzy Systems Association Conf.*, Prague, 1997.

4. Fagin, Ronald., Joseph Y. Halpern., Yoram Moses, and Moshe Y. Vardi, *Reasoning About knowledge.* Massachusetts: the MIT Press, 1995.

5. Forbus, K.D. and de Kleer, J., *Building Problem Solvers,* Cambridge, MA: The MIT Press, 1993.

6. Hajek, P., and Harmoncova, D. "A Comparative Fuzzy Modal Logic," in Klement and Slany (eds.), *Fuzzy Logic in Al,* Lecture Notes in AI 695, Springer-Verlag, 1993, pp. 27-34

7. Nie, J.-Y, and Brisebois, M.. "Using fuzzy Modal Logic for Inferential Information Retrieval". Department d'Informatique et Recherch Operationnelle, University de Montreal, Montreal, 1996.

8. Schotch, R.K., "Fuzzy Modal Logic," Int. Symp. Multi-Valued Logic, IEEE, 1975, pp. 176-182.

9. Wieringa, R.J. and Meyer, J.-J. Ch., "Applications of Deontic Logic in Computer Science: A Concise Overview," in J.-J. Ch. Meyer and R.J. Wieringa (eds.), *Deontic Logic in Computer Science: Normative System Specification*, Wiley, 1993, pp 17-40.

10. Ying, M.-S. "On Standard Models of Fuzzy Modal Logics," *Fuzzy Sets and Systems 26 (1 988),* pp. 357-363.

Integration of Artificial Intelligence, Muscles, and Neural Systems to Replicate Anthropomorphic Systems with Space Applications

Mark A. Johnson
NASA Center for Autonomous Control Engineering (ACE)
University of New Mexico
Email: johnsonm@unm.edu

An overview of the tasks required to integrate artificial intelligence (AI) and synthetic components which mimic anthropomorphic mechanisms is presented. A description is given of the various AI tools (Artificial Neural Networks (ANNs), Fuzzy logic (FL), Genetic Algorithms (GA), . . . which are needed and how they are applied to the problem of controlling both the individual components and a complete anthropomorphic system. The methods used to put together individual components of the anthropomorphic system along with their characterization and integration are developed. Finally, the baseline parameters with which the completed system is to be tested and evaluated are discussed along with identification of potential space applications.

1. Introduction. The NASA/ACE Intelligent Biomedical Engineering Laboratory (IBMEL) is preparing to study and develop the various technologies involved in the design and construction of a synthetic anthropomorphic arm. The mechanism should be able to replicate human arm performance throughout the normal range of motion. One key component is the development and construction of synthetic actuators from the contractile polymer material developed by Dr. Mohsen Shahninpoor in the Artificial Muscle Research Laboratory at the University of New Mexico (l). Another key component will be the development of "the neural sensing and activation system for individual muscles and muscle groups. In addition, bone structure statics and dynamics will be studied with the goal of furthering the work accomplished by Dr Thompson (2). Finally, the control system structure and methodologies will be researched, studied and tested. This is the arena in which the various Artificial tools will be applied to determine the appropriate structure and mix needed together with classical methods to emulate anthropomorphic systems control. All of these areas will then be integrated into operational subsystems for testing and evaluation and later into a complete anthropomorphic arm replicating the shoulder and elbow joints in performing upper and lower arm motions.

Performing these various developments and then integrating them into a complete operational structure will provided a means for evaluating the ability to use these components either individually or as an integrated whole to replicate systems used today in conventional actuator, sensor, joint and linkage, and control systems technologies. These will provide a basis for identifying their usability within current medical, commercial, military, and space applications.

The focus of the following paper is on the tasks to be performed to realize an integrated synthetic anthropomorphic system. Essentially the following paragraphs present a simple work breakdown structure from which' the tasks required to fulfill each parallel effort along with their integration, testing and evaluation. First, an overview of the work breakdown structure is presented followed by a brief discussion of each of the individual efforts to be accomplished in component and system design and development. Next, the testing and evaluation to be accomplished is given. Finally, a summary is presented with projected potential space environment applications.

2. **Work Breakdown Structure Overview.** A work breakdown structure is a tool used to visualize the tasks needing to be performed and their hierarchy for a given project. The work breakdown structure (WBS) for this endeavor is shown in Figure 1. It is comprised of individual efforts in the areas of muscle development, skeletal development, local and distributed hierarchical element and system control development, and systems integration. Muscle activation and sensing is a separate line of research which is a subset of the muscle development tasks. Within each task shown there are research, design, development, and test and evaluation activities. Administrative functions are not identified within the WB S shown although they will be required.

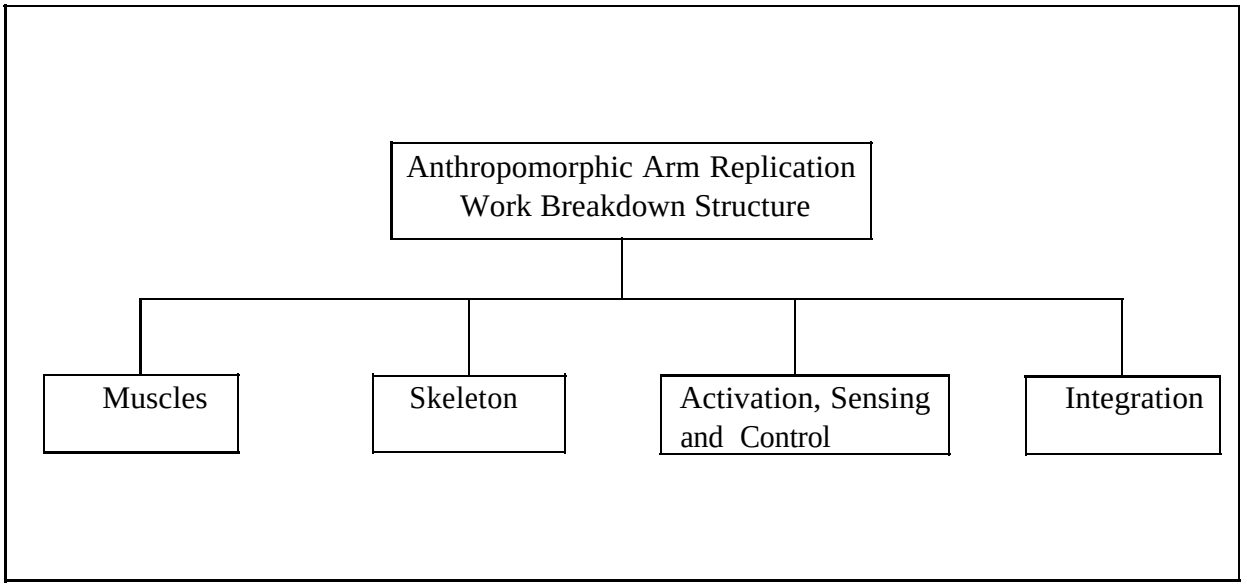


Figure 1. Anthropomorphic Arm Replication Work Breakdown Structure

From a personnel perspective each top level effort will be led by a doctoral student supported by other graduate and undergraduate students. Each doctoral student will be mentored by their appropriate committees.

3. **Design and Development.** Most articulated robotic manipulators are designed using the human arm as a model. A case in point is the vertically articulated PUMA-560. For the proposed project, at a minimum a synthetic skeletal frame with the same stress and strain characteristics of the human arm bones needs to be either design and developed or obtained. A crucial task will be the development and fabrication of the linear muscle actuator(s) which will emulate the human arm muscle actuator(s). Another crucial task will be the development and fabrication of the neuronal communication and computational mechanisms emulating the human arm neural system. The final main component is the arm control mechanisms. These need to control individual muscles and fictional muscle groups in given task performances.

As far a design is concerned every effort will be made to follow human arm design parameters for the shoulder, upper arm, elbow, forearm, wrist and hands. The designs will attempt to replicate skeletal, musculature, and neuronal structures. Skeletal design and development will build upon previous work accomplished by Dr Thompson at the University of New Mexico on characterization of the wrist (2). Since the muscle material developed by Dr. Shahninpoor provides an output voltage for a given deflection, this property will be investigated as a sensory feedback mechanism.

The principle local muscle communications, feedback, and control mechanism will be a three dimensional artificial neural network with spatial and temporal resolution characteristics previously developed by the author (3,4). Top level command and control mechanisms will be designed at the systems functional level in order to control gross and fine system motions. These mechanisms will be developed through investigations of classical methods as well as neural network, fuzzy logic, genetic algorithms and programming, and dynamic programming. The focus will be to determine the appropriate mix and integration of these techniques to satisfy motion controllability, stability, and sensitivity criteria. An attempt will be made to incorporate the work done by Dr. Eddie Tunstel for command and control of mobile robots (5). Behavior-based distributed planning and control may be looked at for specific tasks; however, it is not a component of the project. A majority of the work associated with the project involves performing numerous experiments and will therefore result in several experimental theses and dissertations.

4. **Test and Evaluation.** Testing will focus on characterizing the individual components in engineering terms that would enable an engineer to determine their usability as part of a system or component design. Skeletal components and systems will be tested to ensure they meet or exceed human bone and skeletal frame characteristics. Individual muscle components will be tested to determine tensile strength, shear modulus, temperature affects, excitation methodologies, excitation levels versus performance, return signal acquisition, sensitivity, stability, and controllability. Similar tests will be performed on complete individual muscles to determine the same characteristics of the individual muscles comprising the human arm for comparison and performance predictions.

Control of muscle components, individual muscles, and functional muscle groups will be tested to determine controller design(s) and to train the neural sensory/stimulation system and higher level controllers the unmodeled characteristics of the muscle element or system. Performance versus composition parameters will also be tested to determine their scalability mode (e.g. linear, nonlinear). Tests will cover unloaded system characteristics as well as

performance under positive and negative loading.

Adaptation to non a priori motions and situation will be tested with comparisons to published adaptation capabilities of current robotic systems. Compliance tests will also be performed to determine the potential for compliant operations. In addition, use of the artificial muscles to actuate a PUMA 560 will be attempted to determine their applicability to existing robotic systems. Finally, when the entire structure is assembled and integrated pronation and supination movement tests will be performed covering the entire range space of operation as well as performance characteristics.

Evaluation of test results will focus on comparing and contrasting performance of the artificial components and system with anatomically developed models as well as other existing robotic components and systems. Evaluation will look at the skeletal, muscle, neural communication and control system, system operating range, power characteristics, efficiency, manufacturability, and potential for real world and space applications.

Evaluating muscles will comprise of looking at individual components, complete muscles, and muscle groups throughout their entire range of operation. Their performance will be compared and contrasted with anthropomorphic systems characteristics. Specifically, a look will be taken at their systems' instantaneous velocity and torque capabilities as well as compliance, controllability stability, and sensitivity. In addition, the synthetic system performance will be compared with that of a PUMA-560 vertically articulated robot for similar motions under adaptive and classical control regimens (3).

Neural communication and control systems will be evaluated for their ability to perform the necessary excitation and feedback as well as control functions. Comparisons with anthropomorphic systems performing similar functions will also be evaluated. The operating range of the synthetic system will be completely exercised and evaluated versus that of the human system it replicates. This includes range space, acceleration, velocity, torque, compliance, repeatability, precision, accuracy.

An evaluation of the ability to use the synthetic muscles to actuate a PUMA-560 will identify whether or not current robots can be retrofitted, where and when desired for performance or other reasons. Also, a study concerning methods of component and system manufacture will be performed to identify processes and materials required to manufacture individual components and fabricate them into complete assemblies. Error analyses of each of the tests and evaluations will be performed to determine inherent or systematic errors with their distributions and confidence factors.

Finally, an evaluation of the potential applications of the technology, components, and systems will be performed with emphasis on space and hazardous environment applications. Potential applications include almost any applications where actuators are in use today. Other applications are wherever spatial and temporal parameter estimation, guidance and control are needed. These applications include sensor data fusion, tracking, and adaptive and compliant control problems. Specifically, telemedicine, telesurgery, prostheses, flexible structure control, theater and national missile defense, naval propulsion, exoskeletal motion, solar system and planetary exploration vehicles are a few of the potential uses for the technology.

5. **Summary. An** very brief overview of the work proposed as part of the design and development, and test and evaluation of a synthetic anthropomorphic arm has been presented. A top level work breakdown structure shows the fundamental components of the project. Each

component is its own separate project which when integrated with the others will replicate an anthropomorphic arm, using existing technology along with that developed as part of the project. Testing and evaluation will provide characterizations of the individual components as well as the entire system such that they may be easily evaluated for a given application using existing engineering terminology.

Skeletal studies will extend confirmation or adjustment of current models. Development of the polymer based muscle as an actuator technology is valuable on its own. Use and application of artificial intelligence methods and tools for communications, sensing, and control of system components and the integrated system will give insight into their application to adaptive control problems. Finally, development and integration activities will provide a look at component and system design, scalability, manufacturability, and usability for various applications.

Application usability will be focused on space applications; however, while in some circumstances unique, they are readily transferred to other arenas of application.

References.

1. M. Shahninpoor, Lectures on Artificial Muscles as part of University of New Mexico course on Biotechnology, ChNE 437*, Fall 1997.

2. D. Thompson, Lectures on Biomedical Applications as part of University of New Mexico course on Biotechnology, ChNE 437*, Fall 1997.

3. M. A. Johnson, Payload Invariant Control via Neural Networks: Development and Experimental Evaluation. Master's Thesis, Air Force Institute of Technology, Air University, October 1989.'

4. M. A. Johnson and M. B. Leahy, Jr., Adaptive Model-Based Neural Network Control. Proc, 1990 Intl Conf on Robotics and Automation, Vol 3, pp 1704-1711, IEEE computer society press, Los Alamitos, CA, 1990.

5. E. Tunstel, Embedded Control of a Miniature Science Rover for planetary Exploration. Doctoral Dissertation, University of New Mexico, Albuquerque, 1996.

Vision Algorithms to Determine Shape and Distance for Manipulation of Unmodeled Objects

Leticia Montes, David Bowers, and Dr. Ron Lumia
NASA Center for Autonomous Control Engineering (ACE)
The University of New Mexico

Abstract

This paper discusses the development of a robotic system for general use in an unstructured environment. This is illustrated through pick and place of randomly positioned, un-modeled objects. There are many applications for this project, including rock collection for the Mars Surveyor Program. This system is demonstrated with a Puma560 robot, Barrett hand, Cognex vision system, and Cimetrix simulation and control, all running on a PC. The demonstration consists of two processes: vision system and robotics. The vision system determines the size and location of the unknown objects. The robotics part consists of moving the robot to the object, configuring the hand based on the information from the vision system, then performing the pick/place operation. This work enhances and is a part of the Low Cost Virtual Collaborative Environment which provides remote simulation and control of equipment.

Keywords: COGNEX Vision System, Robotic Transformations, Low – Cost Virtual Collaborative Environment, Mars Surveyor Program

1.0 Introduction

This work builds from the Low Cost Virtual Collaborative Environment (VCE), which is a method of simulation based control of equipment over an Internet link [1], [2]. One potential application for the VCE project is the NASA Mars Surveyor Program, where VCE could provide a means of testing robotic explorers to ensure capability prior to launch. This project demonstrates bow this is possible through an example of rock collection on Mars. The first task for this example was to develop vision algorithms that would compute the size and shape of an object, as well as the location of the object with respect to the camera. The second task was to determine the robotic transformations necessary to pick/place the object.

For robotic applications, it is important to determine the surroundings of the robot, especially if the environment is unfamiliar. Visual servoing is commonly used to observe surroundings, then furnishing the robot with this information so it may interact with those surroundings. Three common tasks for vision are determining the size, shape, and location of an object(s) with respect to the camera.

Without vision, a manipulator can only pickup an object if the size, shape, attitude, and position are known. An effort [3] has been made to overcome some of these obstacles, however, the size and shape of the object were predetermined, and the general location of the object was known. In [4], a method for shape recovery called range aspect hierarchy segments and estimates the shape of 3D objects, where the objects general shape, orientation, and position is known. A computer aided design model was used to facilitate shape recognition [5] in order to locate and identify objects. A method of constructing 3 D maps was developed [6] based on blurring and magnification, where the images are taken at two camera positions at a small displacement, and depth is calculated. In this paper, wc present an approach where no object information is required. The size, shape, and location of the object are all determined by images taken at two camera positions. This information is then used to guide the robot's end effecter to grasp and pick/place the object.

1.1 Project Overview

The goal of this project is to develop and demonstrate the feasibility of robotic systems operating in unstructured and unmodeled environments. Through the use of a vision sensor an unknown object's size, shape, and the distance the object is from the camera can be determined. This information, along with robotic transformations [7], is used to calculate the location of the object with respect to the robot's base via robotic transformations.

This particular application performs a pick/place operation. The system is programmed as follows:

1. The PUMA560 moves to a known region where objects might be, and the camera takes a picture. If there are no objects in that region, the PUMA560 moves to the next region. There are a total of four regions for our experiment.
2. If there are objects within that region, the PUMA moves to center the object in the field of view of the camera. A picture is taken and the data are processed through a vision algorithm, which determines the shape of the object. Robotic transformations are used to calculate the first position of the focal point of the camera (with respect to the PUMA560 base).

3. The PUMA560 moves toward (or away) from the object and a second picture is taken. Robotic transformations calculate the second position of the camera, and the distance between the first and second position of the camera is calculated. This distance and the data from the first and second images are used to complete the vision algorithm. The distance from the camera to the object, size, and shape of the object are now known.
4. A position vector from the object to the camera is computed, and robotic transformations determine the position vector of the object from the base of the PUMA560. The PUMA560 then moves near the object, configures the gripper based on the shape of the object, then performs pick/place. Steps 1-4 are repeated until the table has been cleared.

There are two main topics that will be discussed: vision algorithms and robotics. The vision algorithm portion will discuss the techniques used to determine the size, location, and shape of unknown objects. The robotics section will discuss how the vision information and the robotic transformations are programmed to pick and place an object. The robotics section will also discuss the how the Barrett Hand is configured for a sufficient grasp. Finally, some closing comments in the conclusion section.

2.0 **Vision Algorithms**

The vision software used in this project is Checkpoint Machine Vision System by COGNEX. This software contains various vision tools that are used to analyze a part or application. In this application, a *Blob Tool is* used. The *Blob Tool is* used to detect the number, location, shape, and orientation of 2-D blobs with in an image. The *Blob Tool can* locate one or more objects in the area of an image specified by a region. For each object that it locates, it can determine the angle, area, center of the object, elongation, enclosing rectangle, Euler number, perimeter, major axis, and principal moments of inertia.

2.1 Shape Determination

In order to estimate the shapes of 3-D objects from the camera data, an area fitting process is used. It consists of matching the enclosing rectangle and its dimensions in pixels (rectangular width and rectangular height), area, and perimeter provided by the *Blob Tool* with area of certain shapes. The shapes used are a rectangle, a triangle, a sphere, and an ellipse. The areas of these shapes are computed using the rectangular width and height and the perimeter. An error maybe found by subtracting the area given by COGNEX and the calculated area, then dividing by the area from COGNEX.

For example, let's say the object is a triangle. COGNEX would take a picture that would look close to the picture given in Figure 1. (Dashed lines represent the enclosing rectangle.) The areas for the shapes would be calculated from the data returned from COGNEX.

Area_Rectangle = rw*rh
Area_Triangle = ½*rw*rh
Area_Sphere = PI * (perimeter/2* PI)
Area_Ellipse = PI * (½*rw) * (½*rh)

The errors are calculated based on the calculated areas and the area returned from COGNEX.

Rectangle_Error = (Area_COGNEX - Area_Rectangle)/Area_COGNEX
Triangle_Error = (Area_COGNEX - Area_Triangle)/Area_COGNEX
Sphere_Error = (Area_COGNEX - Area_Sphere)/Area_COGNEX
Ellipse_Error = (Area_COGNEX - Area_Ellipse)/Area_COGNEX

Once these errors have been calculated, a process of elimination is performed using the areas that "match" the known shapes to the unknown shapes. The methodology works as follows:

1. IF Rectangle_Error < Triangle_Error AND Rectangle_Error < Sphere_Error AND Rectangle_Error < Ellipse_Error THEN it is a RECTANGLE. If it is not a rectangle, then it checks the other three shapes.
2. IF Triangle_Error < Rectangle_Error AND Triangle_Error < Sphere_Error AND Triangle_Error < Ellipse_Error THEN it is a TRIANGLE. If it is not a triangle, then it must be either a sphere or an ellipse.
3. IF the difference of the rectangle width and rectangle height (rw –rh) is less than 5, THEN it is a SPHERE. This is true because the enclosing rectangle's sides will be approximately the same. In other words, the enclosing rectangle will actually be a square.
4. IF these tests fail, THEN it is an ELLIPSE.

In the example, the algorithm would stop after Step 2 because it would have determined that the unknown object is a triangle. If the object can not be characterized as a rectangle, triangle, or sphere, the algorithm will conclude that it is an ellipse.

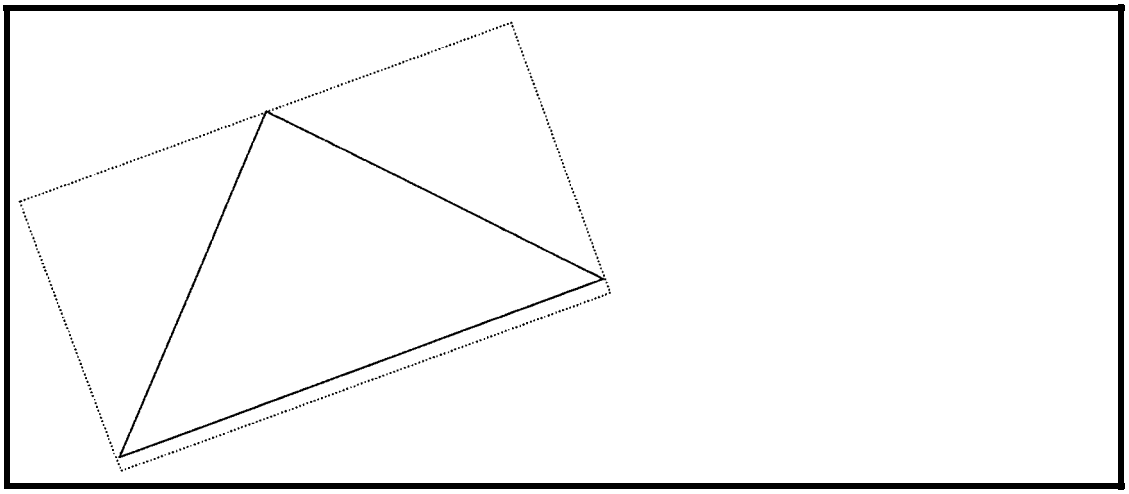


Figure 1: Picture of a triangle by Cognex

2.2 Determination of Size and Distance from Camera

The next algorithm, called "Triangulation Method", determines the distance the unknown object is from the camera's focal point, as well as the size. This method is based on how a camera lens views images. It works on the premise that if a camera with focal point f takes one picture, moves in closer (or further away) a known amount and takes another picture of the same object, the size and distance from the object to the focal point of the camera may be calculated. Figure 2 is a schematic that illustrates this idea.

The triangulation method is determined from the following equations:

Triangle ACP is similar to Triangle ODP, therefore

$$a \div f = h \div (f + X + DT)$$

Triangle BCP is similar to Triangle O 'D 'P, therefore

$$b \div f = h \div (X + f)$$

This creates a system of two equations and two unknowns, h and X. The next two equations solve for these two unknowns.

$$X = [(a*f) + (a*DT) - (b*f)] \div (b-a)$$

$$h = [b*(X + t)] \div f$$

These values are converted from pixels to millimeters using the following conversion:

1 pixel= 0.0098 millimeters

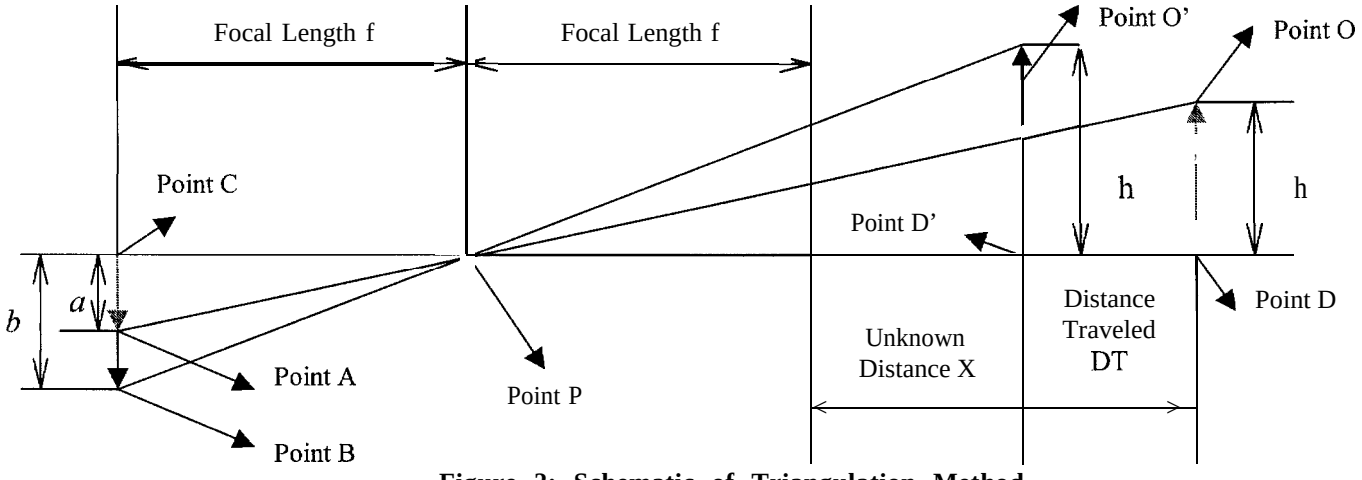


Figure 2: Schematic of Triangulation Method

The variables *a* and *b* used in this method are both determined by the rectangular height of the object. This rectangular height is used because the palm of the Barrett Hand is placed along the direction of this dimension. In other words, the Barrett Hand grasps the object along the shorter side.

3.0 Robotic Manipulation

In order to pick and place the objects seen by the vision system, it is necessary to represent the positions, orientations, and coordinate systems of these objects and for the PUMA560 robot. In this project the world coordinate system, to which everything discussed will be referenced to, is the center of the base of the PUMA560 robot (hereafter will be denoted as PB, for Puma Base). A constant homogeneous transformation is defined from the camera with respect to Joint 3 from a previous calibration scheme. Figure 3 illustrates the world coordinate system and the transformation from the camera with respect to Joint 3. (The notation used in this paper is defined as ${}^{B}T_{A}$, or 'P_{A} and reads "Transformation of B relative to A" and the "Position of B relative to A". For example, ${}^{Camera}T_{J3}$ is the transformation of the camera relative to Joint 3.)

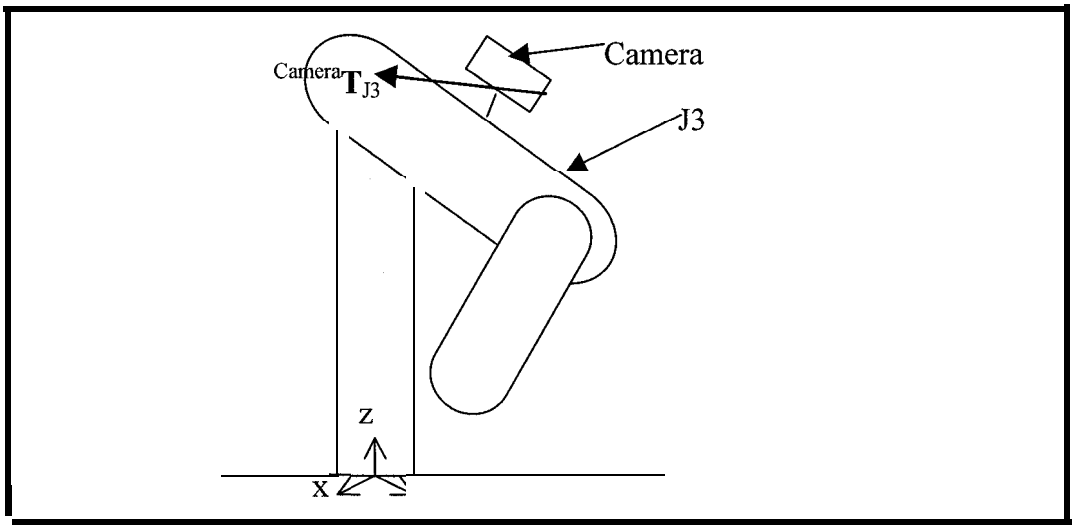


Figure 3: PUMA560 Robot

From the triangulation method, the distance between the object and the camera can be written as a position vector from the object with respect to the camera frame. It is desired to find the object's position vector with respect to some other frame, specifically the PUMA base. Once this is determined, a "teach point" is added within the coordinate frame of the PUMA base, and the end effecter is programmed to move to that point. Figure 4 shows the transformations used to represent the object with respect to the PUMA base.

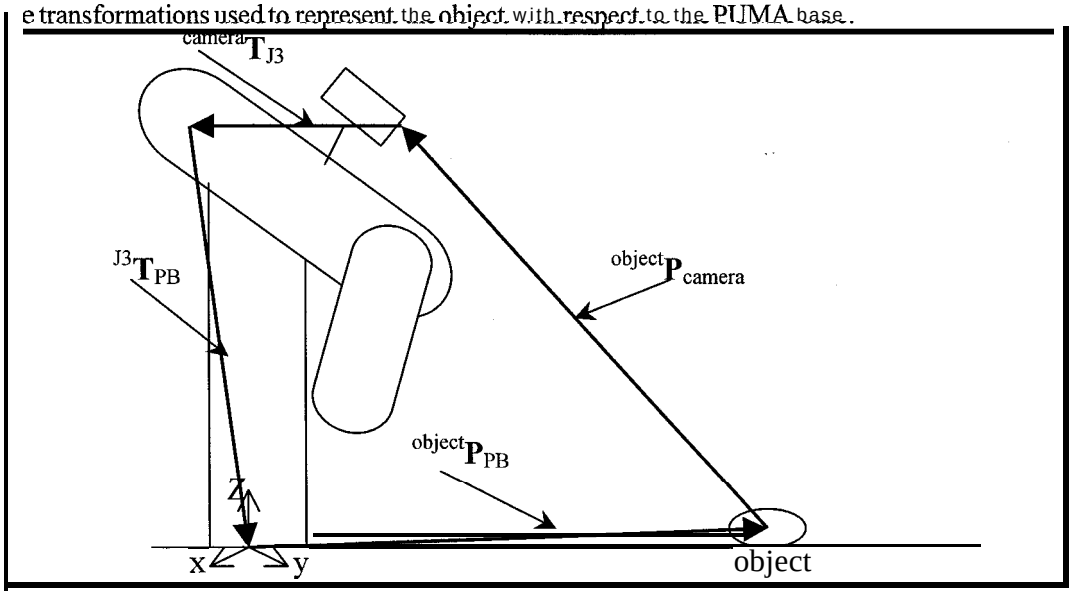


Figure 4: Transformations used to calculate the position vector of the object

The matrix mathematics is shown below.

$$[{}^{object}P_{camera}][{}^{camera}T_{J3}][{}^{J3}T_{PB}] = [{}^{object}P_{PB}]$$

3.1 Hand Configurations

The object's size and shape are used to configure the Barrett Hand for a successful grasp. The Barrett Hand is a three-fingered mechanical hand, equipped with clutches on each of the fingers that do not allow the hand [8] to continue closing once an object is encountered. By adjusting the degree of spread of the fingers, the hand can accommodate a variety of shapes and sizes. Based on the shape classification (rectangle, triangle, sphere, and ellipse) and a measure of the object's size, the hand is configured for a grasp. For example, if the vision algorithm identifies a rectangle 2 inches in height, the hand is configured with two fingers adjacent to each other, and the third approximately 180° from the two fingers. If it identifies a 5-inch diameter sphere, the three fingers spread approximately 120° from each other. A concurrent project is investigating alternative hand grasping schemes and determination of a successful grasp.

4.0 Experimental Results

Conducting a series of experiments in which a sphere was placed in the robot's workcell tested the algorithms. For each experiment a sphere was placed in some position in the workcell then the program executed, then a smaller sphere was placed in the same position, and the program executed. A total of three experiments (30 trials) were performed in order to understand how well the algorithms would calculate the position coordinates and size of the sphere. Table One lists the results (in mm) for the three experiments.

Experiment One

Sphere Size (mm)	96.52			
Trial Number	**Measured Sphere Size (mm)**	**X-dir Coordinate (mm)**	**Y-dir Coordinate (mm)**	**Z-dir Coordinate (mm)**
1	**96.52**	**-654.14**	**16.99**	**34.44**
2	**96.52**	**-653.98**	**16.97**	**31.75**
3	**96.52**	**-653.92**	**16.97**	**30.76**
4	**96.52**	**-654.03**	**16.98**	**32.72**
5	**96.52**	**-654.19**	**16.99**	**35.29**
Sphere Size(mm)	**60.96**			
Trial Number	**Measured Sphere Size (mm)**	**X-dir Coordinate (mm)**	**Y-dir Coordinate (mm)**	**Z-dir Coordinate (mm)**
1	**59.64**	**-653.06**	**5.46**	**13**
2	**59.64**	**-652.98**	**5.45**	**11.76**
3	**59.64**	**-653.07**	**5.46**	**13.23**
4	**59.64**	**-654.13**	**5.47**	**14.14**
5	**59.69**	**-653.15**	**5.47**	**14.59**
		Experiment Two		
Sphere Size (mm)	**96.52**			
Trial Number	**Measured Sphere Size (mm)**	**X-dir Coordinate (mm)**	**Y-dir Coordinate (mm)**	**Z-dir Coordinate (mm)**
1	**95.5**	**-629.8**	**10.5**	**38.99**
4	**95.5**	**-629.8**	**10.5**	**38.4**
5	**95.5**	**-629.7**	**10.5**	**37.9**
Sphere Size (mm)	**60.96**			
Trial Number	**Measured Sphere Size (mm)**	**X-dir Coordinate (mm)**	**Y-dir Coordinate (mm)**	**Z-dir Coordinate (mm)**
1	**60.45**	**-627.4**	**10.6**	**13.8**
2	**60.2**	**-627.6**	**10.6**	**15.4**
3	**59.94**	**-627.9**	**10.6**	**18.3**
4	**60.2**	**-627.6**	**10.6**	**15.1**
5	**60.45**	**-627.4**	**10.6**	**13.1**
		Experiment Three		
Sphere Size (mm)	**96.52**			
Trial Number	**Measured Sphere Size (mm)**	**X-dir Coordinate (mm)**	**Y-dir Coordinate (mm)**	**Z-dir Coordinate (mm)**
1	**92.96**	**-678.94**	**36.25**	**47.21**
2	**95.76**	**-678.56**	**36.05**	**31.6**
3	**96.01**	**-678.53**	**36.03**	**38.08**
4	**95.25**	**-678.6**	**36.07**	**33.18**
5	**95.00**	**-678.2**	**36.08**	**33.88**
Sphere Size (mm)	**60.96**			
Trial Number	**Measured Sphere Size (mm)**	**X-dir Coordinate (mm)**	**Y-dir Coordinate (mm)**	**Z-dir Coordinate (mm)**
1	**59.94**	**-665.41**	**26.05**	**6.99**
2	**59.18**	**-665.66**	**26.13**	**15.3**
3	**59.69**	**-665.48**	**26.09**	**10.93**
4	**59.18**	**-665.57**	**26.11**	**13.24**
5	**58.67**	**-665.79**	**26.16**	**18.56**

Table One: Experimental Data – Measured Size and Position

The results indicate that the position coordinates are repeatable for a particular sphere size. For different spheres at the same location the coordinates for the smaller sphere are consistently smaller or equal to those

recorded for a larger sphere, which may indicate a bias, The most variation is in the Z-dir coordinate, which maybe due to the varying sizes, or because the robot cannot move exactly along the focal line of the camera. Also, the measured sphere size was very accurate when compared to the actual sphere size.

5.0 Conclusions

In order to expand this project further, new features will be added and some of the algorithms will be modified. This includes the addition of a force sensor that calculates if the object is too heavy for the PUMA to lift. The vision algorithms may also be modified and enhanced, as well as an improved method to determine better grasp configurations. Furthermore, more testing and modifications are necessary to improve the accuracy and repeatability of the algorithms.

6.0 References

[1] Todacheeney, M. Quintana, S., Lumia, R., "Low-Cost Virtual Collaborative Environment Through Open System Technology," Proceedings of the World Automation Conference, Montpelier, France, May 1996

[2] Bowers, D., Montes L., Ramos, A., Joyce, B., Lumia, R., "Demonstration of the Low-Cost Virtual Collaborative Environment (VCE)," Proceedings of the NASA URC Technical Conference, Albuquerque, NM, February 1997

[3] Ikeuchi, K., Horn, B., Nagata, S., Callahan, T., Feingold, O., "Picking Up an Object from a Pile of Objects," Robotics Research, The MIT Press Series in Artificial Intelligence, 1984

[4] Dickinson, S., Metaxas, D., Pentland, A., "The Role of Model-Based Segmentation in the Recovery of Volumetric Parts From Range Data," IEEE Transactions on Pattern Analysis, Vol. 19, No. 3, March 1997

[5] Belles, R., Horaud, P., Hannah, M., "3DPO: A Three-Dimensional Part Orientation System," Robotics Research, The MIT Press Series in Artificial Intelligence, 1984

[6] Lee, S., Ahn, S., Meyyappan, A., "Depth from Magnification and Blurring," Proceedings of the 1997 IEEE International Conference on Robotics and Automation, Albuquerque, NM, April 1997

[7] Craig, J., Introduction to Robotics Mechanics and Control, Addison-Wesley Publishing Company, 1996

[8] BarrettHand Bh8-250 User's Manual, Barrett Technology Inc., 139 Main Street, Kendall Square Cambridge, Massachusetts 02142-1528

Acknowledgement

This work was supported in part by NASA under contract #NCCW-0087.

Image Acquisition and the Eye of the Fly

John A. Moya

NASA ACE Center
Dept. of Electrical and Computer Engineering
University of New Mexico, Albuquerque, NM, 87131-1356.
jamoya@unm.edu

Abstract Detailed study has been carried out on various animal eyes, such as those of the fly. The fly's eye possesses certain significant advantages which have made it among one of the most investigated. One of these is its structure, which allows one to study a very small portion and still gain an understanding of the larger eye. In this paper, a model of the sensory portion of the fly's eye is presented. The effectiveness of the model is briefly addressed by a comparison of its performance with experimental data,

1 Introduction

A most important part of a vision system is its photoreceptive layer. This layer provides the sensory input and therefore sets limits on the performance of the system. Thus, a good starting point for the study of a vision system is in the most distal parts of the eye.

If one wishes to study an insect eye, the eye of the fly is an excellent choice. It is experimentally convenient and much literature on its eye has been produced. In this paper, after an introduction to the fly's eye, a new model of its processing in its photoreceptor layer is presented, The effectiveness of the model is also addressed.

This paper extends and improves upon results presented earlier at last year's conference [1]. A more detailed coverage of this work is available in [2].

2 The distal fly's eye

The most distal layers in the eye of the fly are the optical layer, the retina and the lamina ganglionaris, usually referred to as simply the lamina [3]. The optical layer, the familiar compound structure seen at the eye's surface, is the most visible part of the eye. The term retina as applied to the fly differs from its use in vertebrates, referring only to the portions of photoreceptors distal to their own axons. The lamina contains the remaining portions of the photoreceptors and the second-order cells upon which the receptors synapse.

The retinal portions of the photoreceptors contained within the eye are grouped into eight-member units, each unit present under a single facet of the eye [3], As depicted in fig. 1A, these receptors are arranged in an asymmetric trapezoid. Six peripheral receptors, referred to as R1 to R6, surround a pair of tandemly arranged central ones, R7 and R8. The central receptors and peripheral receptors are in general of different types and appear to serve different functions [4]. The receptor units are separated from one another by pigment cells which prevent light scattering [3].

The fly's photoreceptor can be broken up into four parts: the rhabdomere, the soma, the axon and the (axon) terminal [3]. The rhabdomere acts as an absorbing waveguide [5] and transduces light into an ionic current [6]. The soma is the remaining retinal part of the cell and receives the current generated in the rhabdomere. The terminal is the portion of the cell that transfers information onto second-order cells in the lamina. The axon is a conduit that connects the soma and terminal.

Unlike many animals with compound eyes, the fly has evolved an "open rhabdom," where every rhabdomere in the receptor unit is optically isolated from its neighbors [3]. In the fly, these rhabdomeres also do not collect light from the same portion of space [7]. Although, as depicted in fig. 1B, receptors from six, adjacent receptor units do join together to sample similar regions. These receptors, however, do not sample concentrically [8]. The axons of these same photoreceptors, after penetrating the basement membrane, a high resistance barrier that separates the extracellular space (ECS) of the retina from the ECS of the lamina, also join together in the same element in the lamina [3]. In these lamina structures, new sets of R1-R6 photoreceptors arc grouped. These grouped receptors are referred to as the neuro-ommatidium.

The axon terminals in a single neuro-ommatidium are electrically coupled via gap junctions, allowing light induced currents in one photoreceptor to flow to the other neuro-ommatidial receptors [9]. This coupling is restricted to adjacent neighbors (i.e. RI is coupled to R2 and R6, R2 to R1 and R3, etc.). The central receptors bypass the neuro-ommatidia, join the efferent lamina axons and continue on to the next neuropil, the medulla.

Glial cells perform a similar role to that of the pigment cells in the retinal layer above, and physically separate the neuro-ommatidial axon terminal rings from one another in the lamina, creating distinct units called cartridges [10]. This glial partitioning also results in electrical isolation of the terminal rings [11]. Current does, however,

pass through the glial cells, effectively coupling these compartments as well. The neuro-ommatidia are repeated thousands of times across the eye [3].

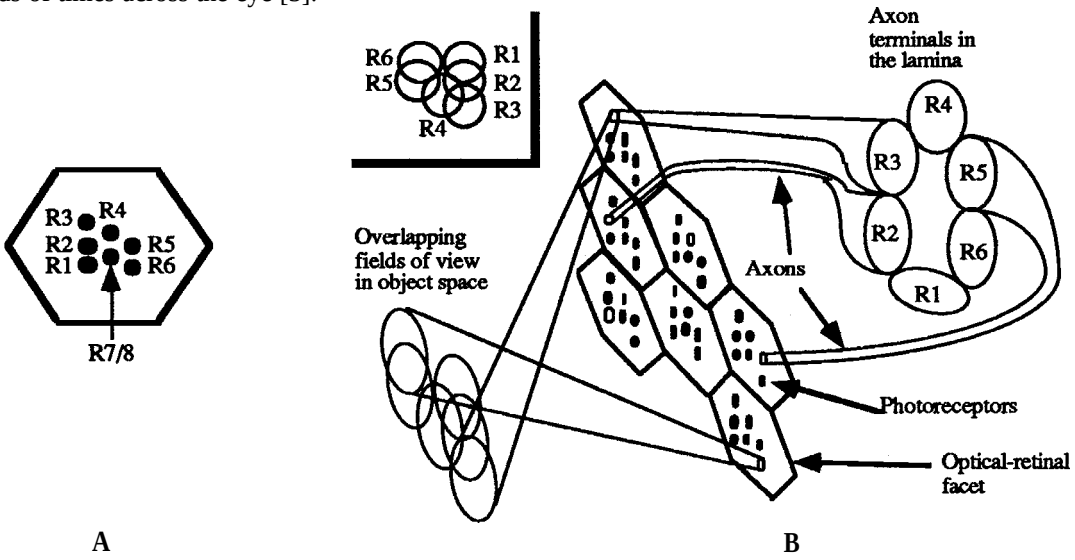


Figure 1 The neuro-ommatidium and the sampling pattern of its photoreceptors. **A.** Eight photoreceptors, **R1-R8,** forma receptor unit housed under each optical facet in the fly's eye. **B. The neuro-ommatidium** joins together photoreceptors from several receptor units. In the illustration the optical, retinal and **lamina** layers have been flattened, the optical and retinal layers combined into a single layer. The joined receptors are depicted as **white** circles within the hexagonal, optical-retinal facets. The **R1-R6** axon terminals of these latter receptors join together to form ring-like structures in the **lamina. Inset:** Every point in space is sampled by six peripheral receptors, each from a different receptor unit. A single **R1, R2,** etc. is used in each of these overlapped samplings.

3 Modeling

The development of a model for the **neuro-ommatidial** layer in the fly's eye starts with the creation a model for a single **neuro-ommatidium.** Once this is specified, the modeling of the fly's entire distal eye can be accomplished by taking into account the effects of the various structural gradients within a fly's eye (see [2]). **The neuro-ommatidial** model can be formulated in two parts, one to account for the collection of light power and the other to account for charge changes in its compartmental structure. The description of the power collection portion of the model is presented in [2] and is not elaborated on **here.**

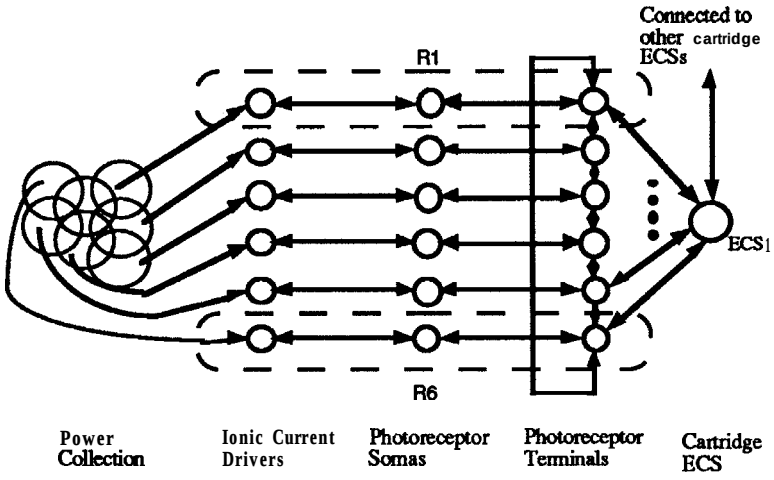


Figure 2 The nodal structure of a model for the flow of charge within the photoreceptor and cartridge **ECS** of a **neuro-ommatidium.** The arrows are used to indicate the type of coupling that exists between the nodes. **R1** and **R6** designate two of the peripheral photoreceptors and **ECS** 1 the node for the cartridge **ECS.**

Figure 2 illustrates the nodes and layers required to model the flow of charge within the photoreceptors and cartridge **ECS** of the **neuro-ommatidium.** The (ionic current) driver nodes **compartmentalizes rhabdomere and** certain **other** current generation, like the **nonstimulation-based** background current I_{bkgnd} (see [2]). In the equations presented here, the layers in the model will be **referred** to by **number,** starting with zero for references to the power collection one. Where a node can be related to one of the photoreceptors (i.e. **R1** through **R6),** it is assumed to be designated by the same number as that photoreceptor. When only one node is present in a layer, it is designated as node one. The notation used to refer to model quantities will **be** of two forms. A quantity of the form Am is a constant. A quantity of the form A_{mnp} refers to a concept associated with processing in the

m^{th} node of the n* layer in the p* neuro-ommatidium.

The ionic current driving the photoreceptor soma should be predominantly composed of sodium ions (Na^+) and potassium ions (K^+) [6]. Na^+ are typically concentrated on the outside of a cell and K^+ on the inside [12]. The increased flow of each of these into a cell has a different effect, increased Na ion flow reducing the lumped, inside-to-outside concentration difference and increased K ion flow raising it. Due to the complimentary effects of inward Na and K ion flow, one can think of the outward flow of K^+ as effectively equivalent to the inward flow of Na^+. One can then combine these and treat them as if they were effectively composed of a single ion.

With the lumping of the ions, equations to describe intracellular charge changes in R1-R6 photoreceptors and the cartridge ECS can be derived using a model like that in fig. 3 [2]. The resulting equations are

$$V_1\dot{X}_{m1p} = G_1(P^{eff}_{m0p})(1 - X_{m1p}) + G_2(X_{m2p} - X_{m1p}) + I_{bkgnd} \quad (1)$$

$$V_2\dot{X}_{m2p} = G_2(X_{m1p} - X_{m2p}) + G_3(X_{m3p} - X_{m2p}) \quad (2)$$

$$V_3\dot{X}_{m3p} = G_3(X_{m2p} - X_{m3p}) + G_4(X_{14p} - X_{m3p}) + G_5\sum_{i\in C}(X_{i3p} - X_{m3p}) \quad (3)$$

$$V_4\dot{X}_{14p} = G_4\sum_{i=1}^{6}(X_{i3p} - X_{14p}) + G_6\sum_{j\in S}(X_{14j} - X_{14p}). \quad (4)$$

The solutions of eqns. 1 through 4 give normalized, lumped, ionic concentrations (NLICs), relative to a lumped, ionic concentration in a remote part of the retina (i.e. in the other eye or in a distant unstimulated part of the same eye) for the four compartments of fig. 3. X_{mip} and Vi represent the NLIC and the volume contained within a compartment in the i^{th} layer. G is a coefficient of flow conductance, $G_1\left(P^{eff}_{m0p}\right)$ being a variable conductance associated with light stimulation. The set C in eqn. 3 defines the terminals within the cartridge ring that are adjacent to themth terminal. The set S in eqn. 4 defines the cartridges that arc apart of the pth neuro-ommatidium's surround, i.e. the cartridge neighborhood that due to proximity and the glial coupling in the fly's lamina significantly affect the NLIC within the cartridge ECS of the p* neuro-ommatidium [11,13,14].

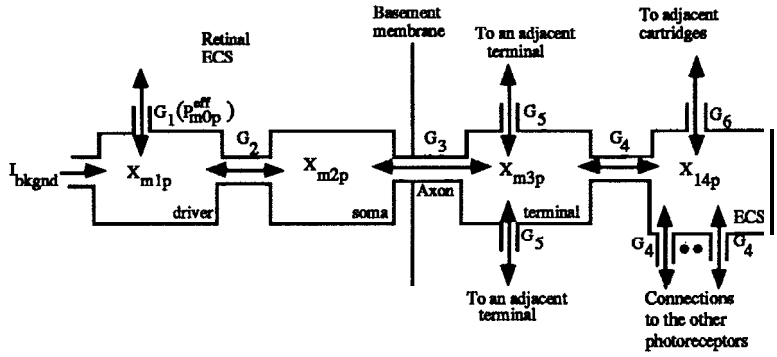


Figure 3 Intercompartmental charge flow within R1-R6 photoreceptors and the cartridge ECS. The arrows are used to indicate the possible directions for current flow. The G_i's are coefficients of flow conductance, $G_1\left(P^{eff}_{m0p}\right)$ being variable and capturing light-dependent aspects of the photoreceptor. X_{mip} is a normalized, lumped, ionic concentration (NLIC), relative to a lumped, ionic concentration in a remote part of the retina. The input current I_{bkgnd} is due to a pumping process which exchanges internal Na^+ for external K+.

$G_1\left(P^{eff}_{m0p}\right)$ is composed of three processes, one that expresses the "ON' conductance and two, one significantly faster than the other, that express cellular adaptation. The ON conductance will be defined as the conductance change, excluding adaptation, that is required to generate cellular responses and will be referred to as $H(P^{eff}_{m0p})$. Cellular adaptation reduces the cell's membrane conductance to a value less than the ON conductance. The faster cellular adaptation process, A^{fast}_{m1p}, responds with time constants on the order of 10 to 20 ms and the

slower one, A^{slow}_{m1p}, with time constants on the order of ten seconds [15]. $H(P^{eff}_{m0p})$, A^{fast}_{m1p} and A^{slow}_{m1p} are combined to give $G_1(P^{eff}_{m0p})$ using the equation

$$G_1\left(P^{eff}_{m0p}\right) = H\left(P^{eff}_{m0p}\right) - A^{fast}_{m1p} - A^{slow}_{m1p}. \quad (5)$$

$H(P^{eff}_{m0p})$ is directly dependent on the phototransduction process [4] and is considered much faster than either A^{fast}_{m1p} or A^{slow}_{m1p}. At least two types of activity comprise $H(P^{eff}_{m0p})$ [2] and these were model using the equation

$$H(P^{eff}_{m0p}) = H_1 - t\, J(P^{eff}_{m0p}) \quad (6)$$

Where

$$J(P^{eff}_{m0p}) = H_2 \tanh(H_3(P^{eff}_{m0p} - 1)) + H_4(P^{eff}_{m0p} - 1). \quad (7)$$

A^{fast}_{m1p} is proportional to $J(P^{eff}_{m0p})$ via a variable β_{m1p} [2], i.e.

$$A^{fast}_{m1p} = \beta_{m1p} J\left(P^{eff}_{m0p}\right), \quad (8)$$

where

$$\dot{\beta}_{m1p} = -D_1\beta_{m1p} + \left(D_2 - D_3\beta_{m1p}\right)\log_{10}\left(P^{eff}_{m0p}\right). \quad (9)$$

Recall the several order difference between the time constants applicable to A^{slow}_{m1p} and the other two conductance processes. For the short duration phenomena to be modeled, a polynomial relationship dependent only on the initial power presented to the model was assumed to be capable of capturing this process [2]. The relation settled upon was

$$A^{slow}_{m1p} = D_4 \log_{10}\left(P^{eff}_{m0p}(0)\right) + D_5 \left(\log_{10}(P^{eff}_{m0p}(0))\right)^4. \quad (10)$$

The model has been implemented in Matlab 4.2c. The coefficients required in the model were determined using anatomical and electrophysiological data (see [2]).

4 Model Effectiveness

Figure 4 depicts model and fly responses to wide-field light pulses. As shown in fig. 4B, the transient peak and fully- adapted steady state agree well with experimental data (cf. [15]). In addition, the trace of fig. 4A reveals that the model has also captured other general features seen in similar fly responses, including approximately the same time-to-peak and fast adaptation settling time (cf. [15]).

Figure 5 depicts sinusoidal response data for the fly and the model. The model can reproduce the experimental response characteristics seen in [16] (see fig. 5A). Although it may appear that the model phase significantly differs from the experimental results, the deviation can in fact be shown to be nearly linear, suggesting that a time delay of approximately 6.5 ms has been omitted from the model [2]. This delay compares well with the expected 2 to 10 ms latency of the animal (see [2], also cf. [14] and [16]}. When the background intensity was varied, the model revealed an adapting magnitude response, the response changing from a lowpass to a bandpass characteristic with increasing background intesity (see fig. 5B). This type of adaptation has also been seen in the fly (cf. [17]).

Figure 6 shows data for small-field, light stimulation of the model. Among the characteristics that the model can reproduce in this case is the time-to-peak delay seen for single photoreceptor stimulation (cf. [1 l]). When two neuro-ommatidially-adjacent receptors are stimulated the voltage responses generated suggest that photoreceptor individuality is preserved in the cartridge. A local contast encoding capability may thus be present in the neuro-ommatidium (see [2]). This result has not been previously predicted in the literature and points to a potential source that could feed into edge detectors in the fly.

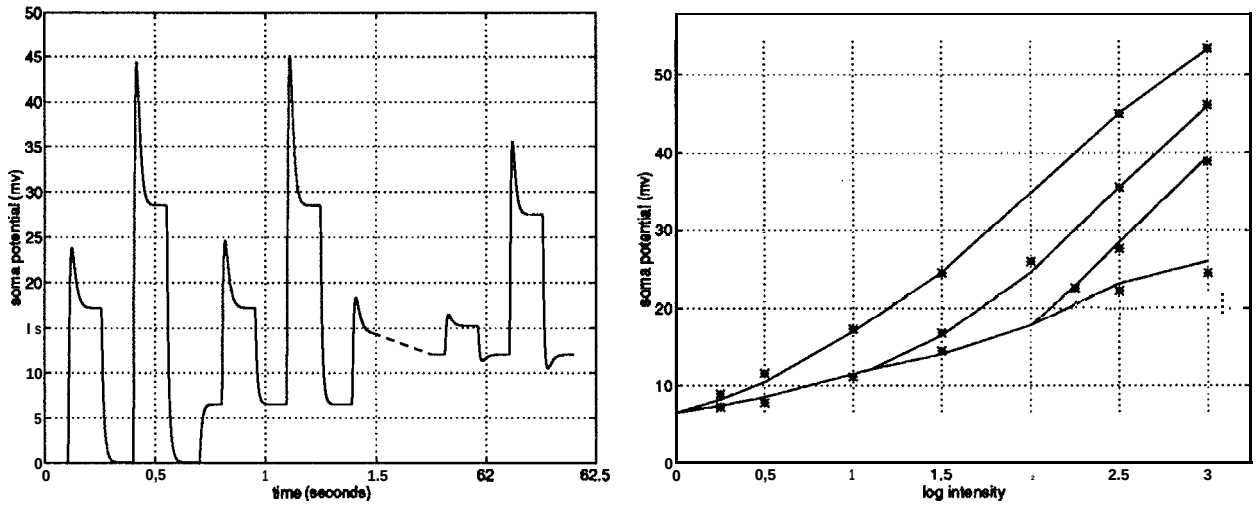
5 Conclusions

A comprehensive model of an eye can improve our understanding of the performance limits in an animal vision system. Working towards such a model for the eye of the fly, the equations presented here model the principles used in the photoreceptor layer of the fly's eye. This model is capable of reproducing the responses recorded in this animal under various types of stimulation. Further, the model predicts certain local contrast encoding capability that may be present in the animal.

Due to the method of derivation and the resulting structure of the model, various adaptational features can quite readily be incorporated as can the optical properties involved in the collecting of light and the process of neuro-transmission onto the subsequent neural layers. Thus, the model offers significant flexibility.

References

[1] Moya, J. 1997. NASA University Research Center Technical Conference.
[2] Moya, J. 1998. Dissertation, University of New Mexico.
[3] Braitenberg, V. 1%7. *Exp. Brain Res.* 3,271-298.
[4] Hardie, R. 1986. *TINS* 9,419-423.
[51 Snyder, A. 1975. In *The compound eye and vision of insects, G.* A. Horridge, ed. Clarendon Press, Oxford
[6] Muijser, H. 1979. *J. Comp. Physiol. A 132, 87-95.*
[7] Kirschfeld, K. 1%7. *Exp. Brain Res. 3, 248-270.*
[8] Pick, B. 1977. *Biol. Cyber. 26, 215-224.*
[9] Ribi, W. 1978. Cell *Tiss. Res. 195, 299-308.*
[10] Saint Marie, R. and Carlson, S. 1983. *J. Neurocytol. 12,243-275.*
[11] Shaw, S. 1984. *J. Exp. Biol. 112, 225-251.*
[12] Nicholls, J., Martin, A. and Wallace, B. 1992. *From Neuron* to *Bruin* 3. Sinauer Assoc., Sunderland, MA.
[13] Dubs, A. 1982. *J. Comp. Physiol. A* 146, 321-343.
[14] Mote, M. 1970. *J. Exp. Zool.* 175, 159-168.
[15] Laughlin, S, and Hardie, R. 1978. *J. Comp. Physiol. A 128,319-340.*
[16] Järvilehto, M. and Zettler, F. 1971. *Z. vergl. Physiol. 75,422-440.*
[17] Juusola, M. and Weckström, M. 1993. *Neurosci. Letters* 154,84-88.



A B

Figure 4 Model voltage responses to wide-field, uniformly-illuminating, light pulses of varying intensity and duration. The stimulus was assumed to be an unpolarized, broad spectrum source. **A.** Typical model responses to pulse log intensities of 1.5 and 2.5 with background log intensities of -3,0 and 1.1 are shown (see abscissa of B). Note the break in time. During this sixty see. period the photoreceptor would sustain an adaptation loss due to the slow adaptation process. **B. The** model photoreceptor peak response (–; upper three curves) for adapting background log intensities of 0,1.1, and 2 and steady state response after complete adaptation (–; lower curve) are shown. The data given with the "*" is from fig. 3 in [15].

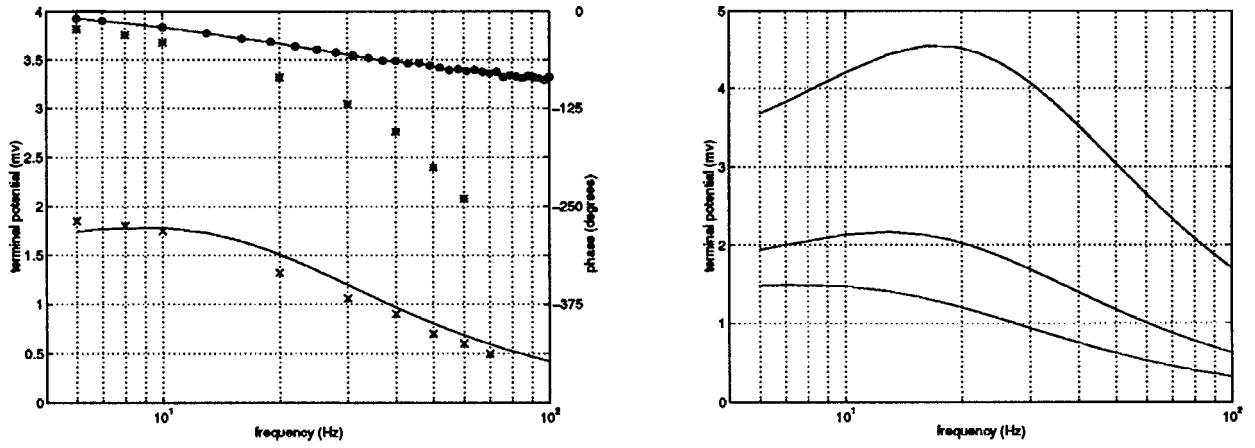


Figure 5 The magnitude and phase of the sinusoidal responses of the model and the fly. The details of the stimulation used can be found in [2]. **A. The** upper curve(~) is the model phase response and the lower curve (—) is the model magnitude. The data symbolized using "*" and "x" is corresponding experimental results **from** [16] for fly photoreceptor phase and magnitude responses, respectively. The phase error between the model and the fly is nearly linear [2]. This suggests that the model lacks a time delay. **B. The** model magnitude response is shown for effective background log illumination intensities of (from bottom to top) .5, 1.5 and 2.5, respectively. Note the change in the response with increasing background light intensity from one of **lowpass** filtering to one **bandpass** filtering. The phase response in this case is similar **to** that in **A. The** filter adaptation is seen in the animal (cf. **[17]).**

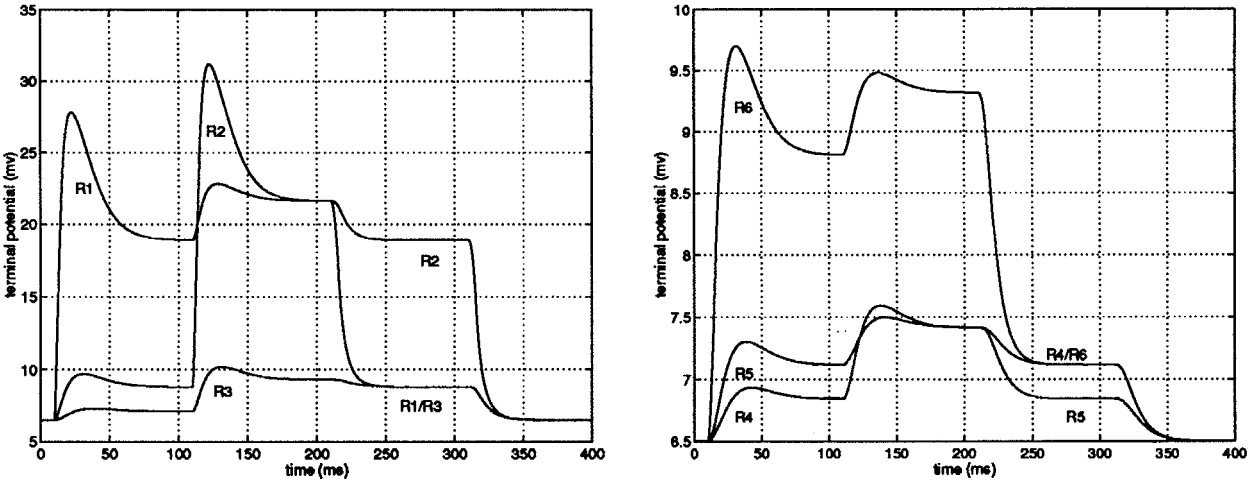


Figure 6 Electro-coupling effects in the photoreceptors associated with a single model **neuro-ommatidium.** These voltage responses were induced in the model **neuro-ommatidial** terminals when 200 **ms,** light pulses of an effective log intensity of 2.5 were sequentially shown on **R1,** then **R1** and **R2,** and finally on the **R2** alone. The two separate light stimuli had a time overlap of 100 **ms.** Note the increasing delay in the **transient** peak (at the left of each plot) as the measured voltage is taken **from** photoreceptors that are increasingly distant around the terminal ring from the singly illuminated **R1.** This characteristic has also been observed **experimentally** (cf. [1 l]). When two receptors are illuminated, these traces suggest that the individuality of the six photoreceptors in a **neuro-ommatidium** is preserved. In such a case, this element may be capable of local **contrast** encoding. **Note: The left and right figures have different ranges.**

98 URC149

Motion Planning in A Society of Mobile Agents

James Campbell and Albert Esterline
Department of Computer Science/NASA ACE
North Carolina A&T State University
Greensboro, NC 27400
Email: {jrc, esterlin}@ncat.edu

1. Introduction

Groups or societies of agents may collaborate on a common goal and thereby achieve more than by working individually. This paper discusses the use of societal theories to model the carrying out of a plan by a group of agents. We focus on plans that relate to agent motion since the low-level aspects of such plans are well understood and motion provides clear criteria for successful planning. We use statecharts to represent these plans, which are hierarchically structured and require concurrent behavior. We represent norms as statements in deontic logic, and we annotate statechart plans with such statements to indicate that certain norms apply to certain groups. In fact, the transitions in a statechart are often naturally interpreted in a normative way. Using deontic logic allows us to distinguish ideal behavior from actual behavior that violates norms and hence to represent fault tolerance. The follow section introduces societal norms and relates them to motion planning. Section 3 presents statecharts and introduces deontic statecharts. Section 4 is an extended example, and Section 5 discusses planning, specifically, how a successful multi-agent motion plan of the form we discuss may be constructed. Section 6 is the conclusion.

2. Societal Norms and Motion Planning

Central to social group theories is the idea of social norms [Tu95]. Social groups with structure have some sort of task or role structure, such as organizations, institutions and some teams. The members of these structured groups act on behalf of the group and are governed by means of formal or informal rules. The rules can be characterized in terms of social norms, which maybe rules called r-norms or proper social norms called s-norms. Norms in a society or group are reasons for members of that group or society either to act or not to act, with sanctions applied accordingly based on permissions. The social position of a societal or group member has attached to it certain tasks, roles and rights. An r-norm task is what is required of a social position by virtue of an r-norm, such as the duties of a job. An example of an s-norm task is what is required in the role of a father-figure as specified by proper social norms, such as doing yard work.

It is critical to note that norms can be violated without being invalidated. For example, an employee may fail to fulfill the duties of his job and a father may fail in his role. The agent violating the norm, not the norm itself, is blamed. In contrast, if certain behaviors run counter to an empirical generalization, we reject that generalization as false and do not blame the agent.

Collision-free movement of an agent through an environment populated with obstacles is a major objective of motion-planning research [Fu9 1]. With multiple agents, the agents in a group or in multiple groups must work together in coordinating moves, and the size of the search space can increase exponentially. Societal norms governing how the agents collaborate can reduce the number of parameters, thereby pruning the search space. For example, with the appropriate norm, an agent would not need to decide whether to stay on the right or the left side of the road – the norm would be reason to perform the appropriate behavior. Also, communication between agents is reduced since the norms of the group dictate many of the aspects of the actions of each agent. Note that norms may be violated, often with serious consequences for the group endeavor, Generally, group members depend on the other members to follow the norms and apply some sanction to members violating the norms.

3. Statecharts

Statecharts [Ha87] are useful in modeling complex reactive systems. Statecharts are visual formalisms used to describe or model states and transitions. Statecharts extend conventional state-transition diagrams to include the ideas of concurrence y, hierarchy and communication. Hierarchy gives the ability to encapsulate or to group substates into superstates. For example, Figure 1 shows a statechart with two states, *D* and *B*. State *D* has two substates, *A* and C. This configuration is an Exclusive Or (XOR) decomposition: if the system is in state *D*, then it is in either substate *A* or substate *D* (but not both). Suppose that the system is in state *B*. On event e, it goes directly to substate *A* of state *D* – *see* the arrow labeled "e". On event *h*, it goes to state *D* as a whole – see the arrow labeled *"h"*, which stops at the boundary of *D*. But the unlabeled arrow to substate *C* indicates that C is the default start state of state *D; so*, on event *h* in state *B*, it ends up going to C. Now suppose that the system is in state *D* (either substate). On event

f, it goes to state *B* - see the arrow labeled "*f*", which starts at the boundary of *D*. Finally, suppose that the system is in substate *A*. On event g, it goes to substate C as long as condition *S* holds – see the "*S*" in parentheses after the label "g".

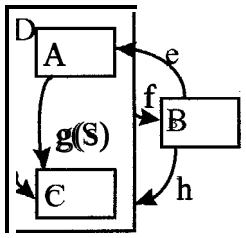


Figure 1.

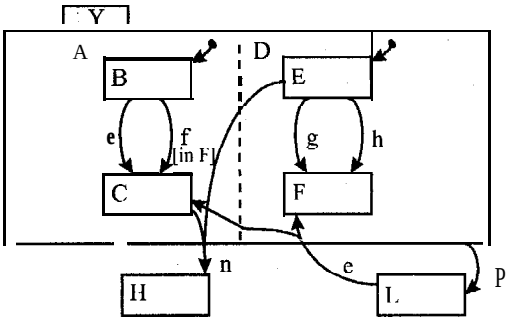


Figure 2.

Statecharts show concurrency through orthogonality. Orthogonality is an AND decomposition and is captured by the partition of the statechart as illustrated in Figure 2. For example, Figure 2 shows a statechart with three states, *H*, *L*, and Y. Y has two orthogonal components, *A* and *D – see the* dashed boundary. State *A* has two substates, *B* and *C*; *D* has two as well, *E* and *F*. If the system is in Y, it must be in one of four possible pairs of substates (one from *A*, one from *D);* it could be in *B* and *E*, *B* and *F*, *C* and *E*, or *C* and *F*. *Now* suppose that the system is in state *L*. On event *e*, it enters substates *C* and *F*. If the system is in substate *E* (and either *B* or C,), then, on event g, it goes to *F* (and stays in whatever substate of *A* it was in). If the system is in substate *B* (and either *E* or *F*), then, on event e, it goes to C. If the system is in *B*, it can go to Con cvent *f* if and only if it is also in substate *F* – see the *"in F' in* square brackets added to the label "f". If the system is in substates C and *E*, then, on event *n*, it goes to state *H*. If the system is in state Yin any of the four ways mentioned above, then, on event *p*, it goes to state *L*.

There are a few enhancements to statecharts that we shall find useful. To begin with, we can add a time interval $[l,u]$ to the event labeling a transition [La95]. This indicates that the event must occur no earlier than l time units after the condition on the transition becomes true (or the source state is entered if there is no condition) and no later than u time units after this comes about. Similarly, we can append an interval $[l,u]$ to the action labeling a transition to indicate that the action can take no less than l and no more than u time units. Finally, we use *T*, possibly subscripted, as a lable on a transition to indicate an internal event that triggers that transition. This is useful when an agent is free to determine when it will perform an action – as long as it is in an appropriate state.

The most significant enhancement is to introduce into statecharts deontic logic operators, enabling one to handle deviations from ideal behavior [Lu97]. Deontic Logic [MW93] is the logic of reasoning about ideal and actual behavior. The modal (specifically, deontic) operators in deontic logic express permission, prohibition, and obligation. Modal operators can modifiy a given proposition. For example, the proposition *Betty washes the clothes* could be modified using the obligation operator, thus forming the sentence *It is obligated that Betty washes the clothes.* In relation to norms, Betty's role in the group could be maid. The duties of a maid include to wash clothes. In formal notation, O is the obligation operator; if wc let *p* stand for the proposition *Betty washes the clothes*, then the above statement of obligation could be expressed as *Op*. There arc also operators for permission (*P*) and prohibition (*F*). We can define any two of these operators in terms of the third; as is usual, we shall take obligation to be the primitive noiton.

We use a version of deontic logic in which a deontic operator can be subscripted with the name of an agent to indicate that the obligation, permission, or prohibition applies to that agent. For example, if *A is the name* of some agent and *p* stands for "The book is returned," then $O_A p$ means that *A* is obligated to see to it that the book is returned. The name could actually denote a group of agents. For example, let G denote a group of dockworkers and let *p* bc the proposition "The pallet is lifted on the truck." Then $O_G p$ asserts that the group of dockworkers has the obligation to see to it that the pallet is lifted on the truck. The case where a group as a whole has a certain obligation must be distinguished from the case where each member of the group, individually, has a certain obligation. By abuse of notation, wc apply deontic operators, not only to propositions, but also to phrases denoting actions, giving propositions with the obvious meaning.

The example in the next section shows how we annotate statecharts with deontic statements. Once we allow deontic statements, we can generally interpret transitions deonticall y as well, except when the system represented is a natural, physical system with no intended purpose. Suppose, for example, that thc behavior of a person is being described. Then there is a general obligation that the person perform the actions and change state as indicated by the

part of the statechart in question. Similarly, there is a general prohibition on performing an action when that action is not sanctioned by the statechart. Finally, there are obligations that relate to the time intervals added to events and actions labeling transitions. We allow that those obligations and prohibitions might be violated. A violation will lead to a violation state where wrongs are righted and the agent is penalized and possibly reformed before he returns to the normal states. If the agent is an artifact that fulfills a purpose, norm violation is faulty behavior, and violation states involve such things as rectifying adverse consequences of the fault, testing the agent, and repairing or adjusting the agent. Thus, deontic statecharts allow us to represent fault tolerance.

4. *Example*

We give an example in which a society of mobile agents follows a plan. Here a society of dockworkers has the collective goal to ship all requested pallets by the end of each day. This society of dockworkers is composed of two groups. Group *B* is responsible for inspecting and moving pallets from a storage location to a shipping and receiving location. Group *A* is responsible for receiving pallets and preparing them for shipment. Two moving obstacles, *Ca* and *Cb*, are in the environment of the dockworkers and may impede their paths. Referring to Figure 3, obstacle *Ca* is first at location 4, moves to location 5, then back to 4. *Cb starts* at location 6, moves to location 5, then back to 6. We assume that *Ca* and *Cb* are never at 5 at the same time. The dockworkers in group *B* include *Ba*, *the* boss, and loaders *Bb* and *Bc*. Their task is to move a pallet from location 8 to location 7, have it accepted by the receiver and his boss, and return to location 8 for another pallet. *Ba*, *Bb*, and *Bc* start at locations 1,2, and 3, respectively. They fist assemble at location 4. They then proceed to location 8, where the boss inspects the pallet and the loaders pick it up. The loaders, carrying the pallet, proceed to location 7 via location 5; they must wait, if necessary, for *Ca* or *Cb* to vacate location 5. At 7, *Ab* is waiting for the pallet, and Aa (his boss) comes from location 9 to 7 to inspect the pallet. If it passes the inspection, then *Ab* receives it and Bb and Bc leave the pallet and go to 10. There they wait (if necessary) for location 5 to become unoccupied so they may go via it to location 4, where the cycle begins again, If the pallet does not pass the inspection, then the loaders go back via the same locations but bring the pallet with them as far as location 8. When *Aa* is finished inspecting the pallet, he returns to location 9; *Ab* remains at location 7.

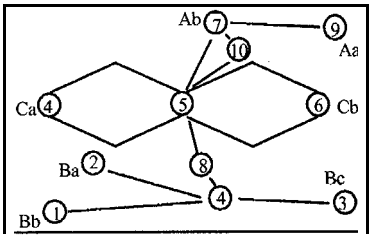


Figure 3.

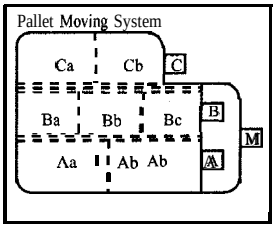


Figure 4.

Figure 4 shows the higher-level composition of the statechart of the pallet moving system. Dashed lines are repeated to show larger units of orthogonal composition. The system contains two orthogonal components, C and M, representing, respectively, the society of workers and the group of obstacles. Component C contains two orthogonal subcomponents, the moving obstacles *Ca* and *Cb*. Component *M* contains two subcomponents, *B* and *A*. *B*, the group of workers that moves the pallets, contains three orthogonal subcomponents, *Ba*, whose role is loader-boss, and *Bb* and *Bc*, both with the role of loader. *A*, the group of workers that receives the pallets, contains two orthogonal subcomponents, *Aa*, whose role is receiver-boss, and *Ab*, whose role is receiver. The duty of a loader agent in this society is to lift and carry pallets to a specified location. The duty of a loader-boss is to inspect a pallet before moving it and to coordinate the loader agents. The duty of a receiver agent is to receive pallets and to prepare necessary documentation for shipping. Finally, the role of a receiver-boss is to inspect and approve a pallet for transport and receipt.

Figure 5 shows the states and transitions for component *Bb* (a loader agent) from Figure 4. The states are grouped into two superstates: *Violation,* which contains as substates all the states that *Bb* enters as the result of a norm violation, and Non- *Violation,* which contains the remaining, "normal" states. The names of the normal states all begin with the name of the agent, "*Bb*", *so* that similar agents may have similar states, distinguished from these only by the prefixes of their names. The next part of the name of a normal state indicates the location of *Bb* when it is in that state. If there is more than one state corresponding to a given location, then lowercase letters are used as suffixes to supply enough distinct names. For example, *Bb* is at location 8 when it is in either *Bb8a* or *Bb8b*.

The general flow is as follows. From *Bb1*, *Bb goes* to *Bb4*. Once *Ba* and *Be* are also at location 4, *Bb goes* to state *Bb8* (picking up the pallet) then (carrying the pallet) to *Bb7*, where the pallet is inspected. If the pallet is approved, it is received, *Bb goes* to location 10 (*Bb10a*) to wait for the obstacles to clear, and then goes back to the

staging state, *Bb4*, where the cycle begins again. If the pallet is not approved, then *Bb* again goes to the staging location, now indicated by state *Bb10b* (since we must distinguish the case where a pallet is being carried from the case where one is not), and stops off at location 8 – to leave the pallet – on the way back to the staging state, *Bb4*. Every action labeling a transition between normal states is a *move.* The various maximum and minimum times for the moves arc indicated. In the "delivery" sequence, from *Bb1* to *Bb7*, transitions are assumed to be triggered by a *signal* event from the loader-boss. A *received* event (from the receiver agent) starts the return to *Bb4* without carrying a pallet, while a *not received* event starts the return while carrying the pallet to location 8. The events triggering the return transitions once state *Bb7* is left are all internal events, indicated by a subscripted "T'; the maximum time (as well as the minimum time) for the *T* event to happen is O.

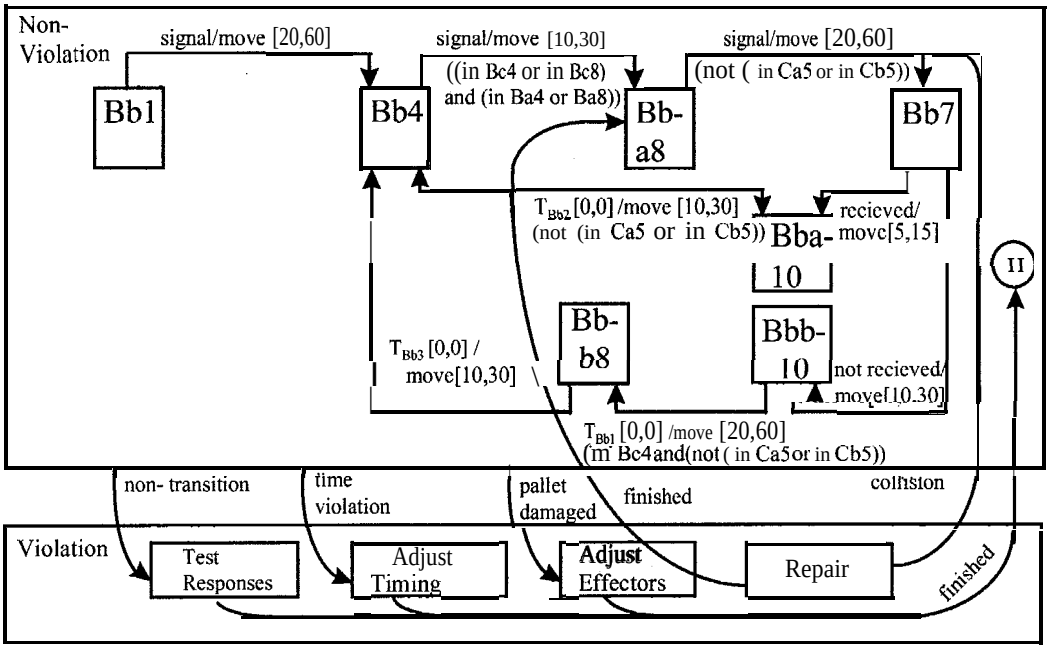


Figure 5

Each transition in Figure 5 involves several obligations. For example, consider the transition from state *Bb4* to state *Bb8*. The statechart indicates that, when the event *signal* happens, the component goes from state *Bb4* to state *Bb8*. We can paraphrase this in terms of Figure 3 as: when the signal is given, *Bb* ought to move from location 4 to location 8. This is an obligation. All of *Bb*'s violations of obligations to make transitions are handled uniformly *Bb*'s responses arc tested. This in indicated in Figure 5 by the arrow labeled *"non-transition"* from the boundary of the non-violation state to the *Test Responses* substate of the *Violation* state. *Test Responses* itself will have substates sufficient to express the steps of the testing procedure. The arrow labeled "*finished*" from *Test Responses* to the *H in* the non-violation state indicates that, when *Bb* is in the *Test Responses* state and the *finished* event happens (the tests have been completed and any adjustments have been made), it goes back to the state where the violation occurred. This will be the source state of the transition that failed. With agent *Bb* now properly adjusted, when event *signal* next occurs, it should indeed make the required transition.

Referring again to the transition from *Bb4* to *Bb8 in* Figure 5, we see that the action *move* ought to be completed in no less than 20 sec. (any faster would be dangerous) and no more than 60 sec. (any slower would be undue delay). Violation of this obligation is taken to be a *time violation* event. Such an event can occur whenever any *move* action is undertaken. Figure 5 shows that this leads to the *Adjust Timing* violation state. Exit from this state (as from the *Test Responses* state) is an event *finished* and again goes back to where the violation occurred (see the arrow to H). Here the state we return to is the destination of the transition. The violation was in the time it took to get there, so the adjusted agent resumes from that state.

Several of the transitions in Figure 5 are labeled with conditions requiring that other components be in certain states. For example, the transition from *Bb8a* to *Bb7* has the condition *not(in Ca5 or in Cb5)*. This is violated if either of the mobile obstacles (*Ca* or *Cb*) *is* at location 5, through which *Bb* must pass in going from location 8 (state *Bb8a*) to location 7 (state *Bb7*). *So* this violation is a *collision* event. It results in the transition from *Bb8a* to *Bb7* on event *signal* becoming a transition to the violation state *Repair* on the event pair *signal-collision* (in that order). The transition from *Repair,* labeled *finished*, is back to state *Bb8a, so the* agent can try again to go from location 8 to location 7. Note that here the violation does not occur at either the source state or the destination state, but in the

course of the transition between the two. Any other condition could be violated, and this component of the statechart could become quite cluttered with violation transitions. An alternative to representing each such violation with an arrow in the diagram is to tabulate the violations outside that statechart. In such a table, there would be no need to invent names (such as *collision)* for condition violations.

The states and transitions for component *Bc* will be just like Figure 5 but with "*Bb*" and "*Bc*" exchanged throughout. The states and transitions for the loader-boss, *Bc*, will have the same structure as in Figure 5 for the delivery sequence from location 1 to location 7 since, even though *Ba* does not go beyond location 8, it must signal to the loaders to go to 7. The return for *Ba is* simply moving from 8 back to 4. The triggering events for *Ba can* all bc internal events since it controls the moves of group *B. Ba*'s actions in the delivery sequence must include *signal* events that trigger transitions by *Bb* and *Bc.*

As an example of a norm that is associated with a state or component, not a transition, consider the obligation of the entire society *M* of agents that all pallets arc shipped by 5:00 PM. Violation leads to state *Run Late.* This applies to the entire society of agents. The individual agents do not have *Run Late* violation states. Rather, there is a *Run Late* violation state for the entire society *M.* This state could be exploited as follows. Suppose wc conjoin the following to the condition on the transition from *Bb4* to *Bb8a*, where *time* is the local time and $<_t$ is the total order on time points expressed by "earlier than":

(*) *time* $<_t$ 4:55 PM *and not in Run Late*

Suppose similar conditions are conjoined to the corresponding transitions of *Ba* and *Bc.* Next, suppose that we add a transition from *Bb4* to *Bb1* on event *signal with the* negation of (*) as condition, and similar transitions arc added for *Ba* and *Bc.* Then, when the loaders are back at location 4, it is 4:55 PM or later, and *M* is not in the Run Late state, then everyone will go home (*Ba* to *2, Bb* to 1, and *Bc* to 3). If, however, the loaders are back at 4, it is 4:55 PM or later, but M is in the *Run Late* state, then *Ba, Bb, Bc* will proceed, as before, to carry a pallet. If *M is* in the *Run Late* state, it will exit this state when the pallets are exhausted – the details do not concern us here (nor does how the pallets are delivered).

5. *Applying Statecharts in Shortest-Path Motion Planning*

So far, we have been concerned with carrying out plans, not with how to form plans, that is, not with planning. When using a finite-state representation (such as statecharts), it is critical to identify discrete state values for continuous parameters, in our case, two spatial dimensions. We make several standard simplifying assumptions to allow us to identify discrete states. We assume that all obstacles are convex polygons, that an agent occupies a point, and that sources and destinations of paths are points.

When obstacles are stationary, the visibility graph [Fu91] has been an important combinatorial structure in planning shortest paths. Vertices of such a graph are vertices of the obstacles as well as the given start and goal points. A line segment that connects two vertices without passing through the interior of any obstacle is an edge of the graph. A shortest path from a start point to a goal point is given as a finite sequence of edges of the visibility graph, The locations, then, to which states in our statecharts relate are the vertices on these shortest paths.

If wc allow obstacles and destinations to move, we can generalize the visibility graph to an accessibility graph [Fu91]. As long as the agent can move faster than any of the obstacles and the destination, a time-minimal motion for the agent is represented as a scquencc of edges of the accessibility graph. Given a velocity (speed and direction), an agent can travel a certain distance before meeting some object; this defines a *collision point* with respect to the velocity. The set of collision points around the start point with respect to a certain speed is partitioned into *collision fronts,* consisting of curves or line segments. The agent first moves to an endpoint, e, of one of the collision fronts computed for the start point. At e, the agent is at a vertex of some obstacle (which has moved while the agent moved). We now compute a set of collision fronts with e as start point, and the agent moves to the endpoint of one of these. This procedure is repeated until the destination point is reached. Here the locations to which states in our statecharts relate are the endpoints of collision fronts that give a shortest path from a start point to a goal point. Dependence on speed is captured by including generally narrow time intervals on the move actions.

In the context of planning, we view statecharts as scripts or schemas [Tu94], or possibly even previously recorded cases [Ha89], that arc selected, modified, and applied as required by the case at hand. A key aspect of applying a statechart is to identify vcrticcs in the accessibility graph with locations used in defining the states. Planning becomes a hierarchical process, and ideally as distributed as possible. Such planning could be done in either a top-down or a bottom-up maner. In the top-down approach, the overall structure of the statechart would be selected, and one would zoom in for the detail, which ideally would be left to the subcomponcnts, constrained by norms inherited from supercomponents. In the bottom-up approach, we would start with statecharts for individual

agents, which would make contracts to form groups and to establish norms as we zoom out to the overall structure. In either approach, the agents in a group must work out certain norms and procedures or centralize this aspect by delegating an agent to represent the collective. Either approach is tractable since we assume that wc have an adequate number of statechart schemas that can be selected, modified, and applied, and we assume that there are a manageable number of norm patterns that can be applied to prune the search space. Finall y, we could allow the planning to be reactive, that is, wc could modify or further elaborate the plan as more information becomes available.

6. Conclusion

In this paper we have discusses how social norms can be used in plans for a society of agents. Such norms can simplify planing as well as the execution of plans. We find statecharts to be an excellent notation for representing plans for societies of agents. In particular, to express norms, we allow deontic statements (about obligations, etc.) to occur as annotations on states and components, indicating that, for example, the stated obligation applies throughout that state or to every subcomponent of that component. Also, transitions arc now interpreted as normative. that is, in deontic terms. To handle norm violations, we specify certain sub-ideal, violation states, where faults arc corrected. We thus also represent fault tolerance. Multi-agent motion planning in this setting involves selecting and adapting statechart schemes and identifying critical locations that can be identified with states in the statecharts. Planning can be top-down or bottom-up, but in any case it is critical to associate the agents into groups or societies and for them to accept norms.

The two most pressing topics arc selection and adaptation of statechart schemes for new cases and communication among agents leading to acceptance of norms and roles. Selecting appropriate schemas requires identification of key features and ways to summarize both a statechart schema and the situation at hand so that meaningful similarities can be found with reasonable effort. Adapting statechart schemas to the situation at hand requires operators that can modify statechart structures and norms in principled ways. Agent communication actions, especially in the form of speech acts, have been studied [CL97], but generally in a context where two agents establish commitments. We must look at communication where norms encompassing entire societies of agents arc accepted.

References

[CL97] Cohen, P.R. and Levesque, H. J., "Communicative Actions for Artificial Agents", in J.M. Bradshaw (ed.), *Software* Agents. Menlo Park, CA: AAAI Press/The MIT Press, 1997, pp. 418-436.

[Fu91] Fujimura, Kiuo, *Motion Planning in Dynamic Environments.* Tokyo: Springer-Verlag, 1991.

[Ha89] Hammond, K. J. ,*Case-Based Planning: Viewing Planning as a Memory Task. New* York: Academic Press, 1989.

[Ha87] Harel, David., "Statecharts: A Visual Formalism for Complex Systems," *Science of Computer Programming 8, 1987.* pp. *231-274.*

[La95] Lane, K., *Formal Object Oriented Development.* London: Springer-Verlag, 1995.

[Lu97] Luu, T.V., *Enhancing Statecharts Using Deontic Logic Operators,* Technical Report, NASA Sharp Plus Program, North Carolina A&T State University, Aug., 1997.

[MW93] Meyer, J.-J.Ch. and Wieringa, R. J., "Deontic Logic: A concise Overview, in J.-J. Ch. Meyer and R.J. Wieringa (eds.), *Deontic Logic in Computer Science: Normative System Specification*. John Wiley & Sons, 1993.

[Tu95] Tuomela, Raimo., *The Importance of Us: A Philosophical Study of Basic Social Notions.* Stanford, CA: Stanford University Press, 1995.

[Tu94] Turner, Roy M., *Adaptive Reasoning for Real- World Problems: A Schema-Based Approach.* Hillsdale, New Jersey: Lawrence Elrbaum Associates, 1994.

The Deontic Transaction Model in Multi-Agent Normative Systems[1]

Yonghua Huang and Albert Esterline
Department of Computer Science/NASA ACE
North Carolina A&T State University
Greensboro, NC 27400
Email: {yonghua, esterlin} @neat.edu

1. Introduction

In the area of multi-agent systems, much research is devoted to the coordination of the agents. There exist several issues, two of which are summarized by [Dig96]. The first is that, although agents are said to be autonomous, they always react in a predictable way to each message, and they cannot decide to violate the conventions that are hard-wired into the protocol. In fact, there might be circumstances in which the agent violates a convention in order to realize a private goal that it considers to be more important. Another issue is that, if the protocols that agents use to react to the environment are fixed, they have no way to respond to changes. However, an important characteristic of agents is that they can react to a changing environment.

Although transaction models ([BOH92], [GR93]) evolved from the database domain, they establish a general execution paradigm that ideally covers all the subsystems invoked in a sequence of transactions. So transaction models apply to multi-agent systems. Recently, some research has been devoted to overcoming the limitations of the traditional transaction models which are suitable for conventional systems and focus on system integrity, e.g., [SJ97].

Here we solve above issues by turning to a deontic concept: obligation. In multi-agent systems, agents interact with each other according to norms, We use deontic logic ([And58], [Aqv84], [BC96], [JS94], [MW93]) to model norms. Here the norms prescribe how the agents ought to behave, but — and this is essential — they do not exclude the possibility of "bad" behavior (i.e., the actual behavior may deviate from the ideal), and so they also prescribe what should be done in circumstances of norm violation.

Thus, we propose a new approach -— a deontic transaction model for multi-agent normative systems. Our approach improves the protocol of "abort/commit" of traditional transaction models to a protocol of "abort/exception/commit". In multi-agent normative systems, wc can see the violation of a norm as an exception and how the other agents react to the violation as exception handling.

This paper concentrates on our formal model; for more details, see [Hu97]. In section 2, we sketch how we extend traditional transaction models to include deontic notions, especially obligation. In section 3 (the main section of the paper), we present our formal model and our specification language, a dynamic deontic propositional language. Section 4 briefly discusses our application of our model using the Java-based Aglets technology. Section 5 is the conclusion.

2. The Structure of Deontic Transaction Model

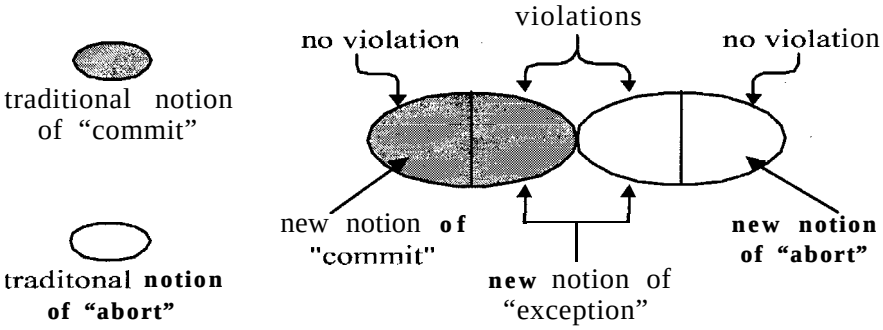


Figure1. Definition of New Notions

Briefly, our approach supports exception occurrence and exception handling by supporting the obligations that different autonomous agents have toward one another. As shown in Figure 1, we redefine the "commit" of a

[1] The first author has been partially support by grant NAG 5-4102, "Formal Foundations of Agents," form NASA/GSFC. The second author has been supported by the same grant and by grant NAG 2-1150, "Motion Planning in a Society of Intelligent Mobile Agents," from NASA/ARC.

transaction as the traditional meaning with a special requirement that no obligation is violated, and the "abort" of a transaction as the traditional meaning with a special requirement that there is no obligation violated. Thus, we can define the "exception" of a transaction as the union of the traditional "commit" in a situation where at least one obligation is violated, and the traditional "abort" in a situation where at least one obligation is violated. We see that the notion of "exception" includes two general cases: we call them "commit with violation" and "abort with violation", respectively.

We use the state transition paradigm portrayed in Figure 3. to present our idea of a deontic transaction model.

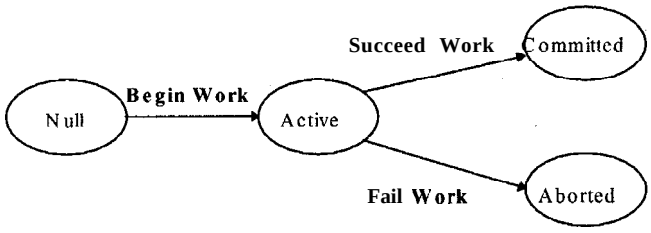


Figure 2. State Transition of the Flat Transaction

Compared with traditional transaction model shown in Figure 2, we have two more states: the *Prepare* state and the *Excepted* state. In the *Prepare* state, the negotiation between agents occurs; different results of the negotiation lead to different states. When an obligation is violated, Vi 01 a t e Work is issued, and the system goes into the *Excepted* state; when the system undertakes some measures to correct the violation, Exception Hand 1 ing Work is issued. The *Committed* state is an ideal state. If the actual state is just the ideal state, then the transaction commits; if the actual state is incorrigibly inconsistent with the ideal state, then the transaction aborts; if the actual state is not optimal but is acceptable — we call it a sub-ideal state — then the transaction goes into the *Excepted* state,

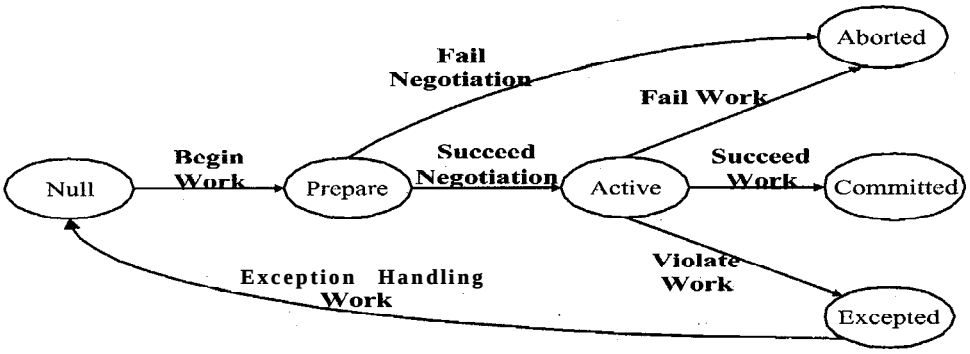


Figure 3. The Deontic Transaction Model

Because we focus on the obligations that different agents have toward one another, we distinguish two roles in an obligation. One is that of the obligater, who either directly invokes the obligation on another by a speech act, or accepts an obligation, which is imposed by another on himself, as with a promise; the other is that of the obligatee, whose behaviors are governed by the obligation. So we can see that not only can the obligater invoke an obligation but also the obligatee could impose an obligation on himself (as with a promise and similar speech acts). A role represents a behavior an agent is authorized to play in the system. Each role is associated with a set of norms. Notice that a given real agent maybe authorized to play several roles.

AU obligation is initiated or violated, respectively, by satisfying the initiation condition or the violation condition. For an Obligation, if both obligater and obligatee agree to it, then the initiation condition is satisfied, and thc obligation is initiated (comes into effect). If the obligatee breaks the obligation, then the violation condition is satisfied, and thc obligation is violated.

In our work, we distinguish general and special obligation, primary and secondary obligation. General obligations hold for all agents, while special obligations hold for some (perhaps only one) particular agents. Primary obligations describe what is ideal to the whole system, and secondary obligations arc concerned with what is to be done when violation of a primary obligation has occurred; they are usually called contrary-to-duty obligations and

describe a sub-ideal state of the system. We also discussed the relation between obligation and other notions, such as violation, negotiation, action. Our work focus on the deontic transaction formal model.

3. The Deontic Transaction Formal model

Our specification language LDT is a simple dynamic deontic propositional language (in the spirit of [DMW96]) whose alphabet consists of a set *Prop* of atomic propositions $(p, p_1, p_2, \ldots)$. a set *Act* of action expressions $(\alpha, \alpha_1, \alpha_2, \ldots)$. a finite set *G* of agents $(i, i_1, i_2, \ldots, j, j_1, j_2, \ldots)$, parentheses and square brackets, the usual truth-functional connective, O (obligation), Do (what happen next), PREV (what just happened), and PAST (what has happened before). The sentenses of LDT constitute the set *Form* defined as the smallest set of formulas closed under

(a) $Prop \subseteq Form$

(b) $\phi_1, \phi_2 \in Form \Rightarrow (\phi_1 \wedge \phi_2), (\phi_1 \vee \phi_2), (\phi_1 \rightarrow \phi_2), (\phi_1 \leftrightarrow \phi_2) \in Form$

(c) $\alpha \in Act, \phi \in Form \Rightarrow [\alpha]\phi \in Form$

(d) $\alpha \in Act, i \in G \Rightarrow Do_i(\alpha) \in Form$

(e) $\alpha \in Act \Rightarrow PREV(\alpha), PAST(\alpha) \in Form$

(f) $\phi \in Form, i, j \in G \Rightarrow O_{i,j}(\phi) \in Form$

Here $[\alpha]\phi$ means that doing α necessarily leads to a state where ϕ holds; $O_{i,j}(\phi)$ means that agent *i* is obligated to agent *j* to bring about ϕ; PREV(α) means that the present state is actually reached by performing α; PAST(U) means that *a* has actually been performed on the past; and $Do_i(\alpha)$ means that agent *i* intends to perform a possible action α next.

The semantics of LDT requires a semantics for action (and transaction) expression [DMW96]. The set *Act* of action expression $(\alpha, \alpha_1, \alpha_2, \ldots)$ includes the set AA of atomic action $(\underline{a}, \underline{b} \ldots)$ and contains as an element **any** ("don't care what happen") and **fail** ("failure"). *Act* also contains expression of the form $\alpha_1 \cup \alpha_2$ (the non-deterministic choice of between *al* and α_2), $\alpha_1 \& \alpha_2$ (the parallel execution of α_1 and α_2), and $\neg\alpha$ (the non-occurrence of *a).* The set *Trans* of transaction expressions (T, $T_1, T_2, \ldots$) includes the set *Act* of action expression and contains as an clement **time** (denoting a transaction lasting one time unit during which **any** is executed enough times to match the length of the unit). *Trans* also includes expression of the form $T_1; T_2$ (the sequential composition of the transactions T_1 and T_2).

The semantics of action expression is given in two stages [DMW96]. In the first stage, algebraic action semantics, we interpret action expression in terms of a class A of events $(a, b, c, \ldots)$. Every atomic action expression $\underline{a}$ is associated with an event $a \in A$. a special failure or deadlock event $\varepsilon \notin A$ is assumed. The relation between an action expression $\underline{a}$ and the associated event *a* involves not only interpreting $\underline{a}$ as a, but also keeping open the possibility of other compatible events. We thus define a *step* to be either a finite no—empty subset of A or {e}, and a *choice set* to be a set of steps. We then define a semantic function $\|\ \|: Act \rightarrow D$, where D is the set of choice sets over A. Now, $\|\underline{a}\|$ is the choice set consisting of all steps with a as a member: performing any of these steps consists as executing $\underline{a}$. Secondary, $\|\mathbf{any}\|$ is the choice set with every step over A as a member; $\|\mathbf{fail}\|$ is the choice set whose only member is the failure step {c}. Again, $\|\alpha_1 \cup \alpha_2\|$ (respectively, $\|\alpha_1 \& \alpha_2\|$) *is* the union (respectively, intersection) of $\|\alpha_1\|$ and $\|\alpha_2\|$, where, however, special provision must be made for *(E).*

In the second stage in the semantics of action expression, state-transition action semantics, wc define a function *change:* $A \rightarrow (\Sigma \rightarrow \Sigma)$, where Σ is the set of states. Thus, for $a \in A, \sigma, \sigma' \in \Sigma$, $change(a)(\sigma) = \sigma'$ means that in state σ the occurrence of event *a* leads to the state σ'. And, for a step S, *change*(S)(σ) is the state representing the joint effect of the events in S on σ (where these events are assumed compatible: their joint effect is independent of their order of occurrence). Finally, where *K* is a choice set, change is a set of states, generally one for each step in *K.*

In our language, wc focus on the obligation one agent has to another and the violation of obligations. We give the semantics of our model by means of the notion of a Kripke structure $M = (Z, G, A, \pi, Norm, Subject, Role, R_A, R_O, \leq_A, Z)$, where

- Σ is a set of states.
- G is a set of agents.

. *A* is a set of events.

- π is a truth assignment function to the atomic propositions relative to a state, that is, π is a function $\Sigma \rightarrow (Prop \rightarrow \{1, O\})$, where 1 and O denote truth and falsity. For example, for $p \in Prop, \sigma \in \Sigma$, nil means that the atomic proposition *p* is true in state cr.
- ***Norm*** is a set of norms, each of which supports obligations (either general or special) and indicates what is required for certain deontically ideal state. A norm is a token, not a type, and is established by a speech act or

other conventional action. Here, Norm is finite, that is, only a finite number of obligation initiations have been performed at any state of the system.

- *Subject* is a function: $Norm \to Form$. It relates each norm with the corresponding formula that was involved in the speech act or other conventional action that established the norm.
- *Role* is a function $Norm \to (G \to \{obligatee, obligater\})$, where *obligatee* and *obligater* denote the two roles of an obligation. For instance, for $n \in Norm, i \in G$, $role(n)(i) = obligatee$ means that agent i is an *obligatee* under norm n. Once an agent is authorized a role, this role will only be canceled by the system, not by himself.
- R_A is an accessibility relation which specifies how actions can change states. R_A can be defined as follow:
  $R_A(\sigma, \sigma') =_{def} \exists S \subseteq A\, [change(S)(\sigma) = \sigma']$
- R_O is the deontic relation which specifies how a state transforms to an ideal state. By putting various restrictions on R_O, we can get different families of Kripke structure. In our model, we consider this relation on the state/history pairs, not only on states, in order to distinguish the state that is deontically ideal from the state that is actually achieved. The history of a state is expressed by a trace of actions that happened before. A trace is a sequence $S_1 S_2 \ldots S_n \ldots$ of steps. The set Γ of *trace expressions* can be given by the following BNF:
  $\gamma ::= \theta \mid \gamma' \bullet S$ for any $S \subseteq A$
  Here "*" is a concatenation operator between the trace and step; θ means the empty trace. R_O is now defined as follow:
  $R_O((\sigma, y), (\sigma', \gamma')) =_{def} R_A(\sigma, \sigma') \wedge (\exists S \subseteq A : \gamma' = \gamma \bullet S)$
  $\wedge (\forall n \in Norm : (M, \sigma') \models Subject(n))$
  Thus, the state-trace pair (σ', y') is related by R_O to the pair (σ, y) iff there is a step S that changes σ to the ideal (violation free) state σ' and y' is the trace γ with a step S added at its end.
- $\leq_A(\sigma)$ is the deontic ordering relation on the set of states reachable from a given state σ by the occurrence of a step over the events in A. For any $\sigma \in \Sigma$, $\leq_A(\sigma)$ is a partial order over the states in $\{\sigma' \in \Sigma \mid R_A(\sigma, \sigma')\}$. For example, for σ, IS', $\sigma'' \in \Sigma$, if $R_A(\sigma, \sigma')$ and $R_A(\sigma, \sigma'')$, then wc can have $\leq_A(\sigma)(\sigma', \sigma'')$ or $\leq_A(\sigma)(\sigma'', \sigma')$ or neither (i.e., σ' and σ'' are not comparable by the relation $\leq_A(\sigma)$). Here $\leq_A(\sigma)(\sigma', \sigma'')$ means that, for a given state σ, state σ'' is deontically better than or equal to state σ'. In terms of the deontic relation R_O, we have that, if $R_O((\sigma, \gamma), (\sigma', \gamma'))$, then $\forall \sigma'' \in \Sigma : R_A(\sigma, \sigma'') \to \leq_A(\sigma)(\sigma'', \sigma')$.
- I is a family of accessibility relations. We define it in terms of the relations I_i; for a given agent i, I_i specifies which states arc possible and intended to be reached by i. I, then, is given as follow
  $I =_{def} \{I_i \mid i \in G\}$
  The relation I_i is a subset of the accessibility relation R_A, that is, $\forall i \in G, \sigma, \sigma' \in \Sigma : I_i(\sigma, \sigma') \Rightarrow R_A(\sigma, \sigma')$. But it cannot be simply related to the deontic ordering relation on states, which we just defined. For example, for $\sigma, \sigma', \sigma'' \in \Sigma$, we may have $I_i(\sigma, \sigma') \wedge \leq_A(\sigma)(\sigma', d')$, which means that even though state IS" is deontically better than state σ', agent i still intends to do actions which lead to state σ'. So this relation reflects the autonomy of agents, that is, they can violate obligations.

Given $M = (\Sigma, G, A, \pi, Norm, Subject, Role, R_A, R_O, \leq_A, I)$ as above and a state/trace pair (σ, y), we can give the interpretation of the formulas in specification language in the Kripke structure M as following:

1. $(M, (\sigma, \gamma)) \models p \Leftrightarrow \pi(\sigma)(p) = 1$ (for $p \in Prop$)
2. $(M, (\sigma, \gamma)) \models \phi_1 \wedge \phi_2 \Leftrightarrow (M, (\sigma, y)) \models \phi_1$ and $(M, (\sigma, \gamma)) \models \phi_2$
3. $(M, (\sigma, \gamma)) \models [\alpha]\phi \Leftrightarrow \forall S \subseteq A : S \in \| \alpha \|, \forall \sigma' \in \Sigma\, [R_A(\sigma, \sigma') \Rightarrow (M, (\sigma, \gamma \bullet S)) \models \phi]$
4. $(M, (\sigma, \gamma)) \models Do_i(\alpha) \Leftrightarrow \exists S \in \| a \|, \exists \sigma' \in \Sigma\, [(change(S)(\sigma) = \sigma') \wedge I_i(\sigma, \sigma')]$ (for $i \in G$)
5. $(M, (\sigma, \gamma)) \models PREV(\alpha) \Leftrightarrow \exists S \in \| a \|, \exists \gamma' \in \Gamma\, [\gamma = \gamma' \bullet S]$
6. $(M, (\sigma, y)) \models PAST(\alpha) \Leftrightarrow \exists \sigma' \in \Sigma, \exists \gamma', \gamma'' \in \Gamma\, [(1' = \gamma' \bullet \gamma'') \wedge (M, (\sigma', y')) \models PREV(\alpha)]$
7. $(M, (\sigma, y)) \models O_{i,j}(\phi) \Leftrightarrow \forall (\sigma', \gamma') \in \Sigma \times \Gamma, \exists n \in Norm\, [(Role(n)(i) = obligatee) \wedge (Role(n)(j) = obligater) \wedge ((k?, (\sigma', \gamma')) \models Subject(n) \to \phi) \wedge (RO((\sigma, \gamma), (\sigma', \gamma')) \Rightarrow (M, (\sigma', \gamma')) \models \phi)]$ (for i, $j \in G$)

Then general and special obligation operators over actions are given as follows.

$O^+_{i,j}(\alpha) \equiv \forall i \in (G \setminus \{j\}) : O_{i,j}(\alpha)$

$O^-_{i,j}(\alpha) \equiv O_{i,j}(\alpha) \wedge \forall k \in (G \setminus \{i\}) : \neg O_{k,j}(\alpha)$

Here O' expresses general obligation and 0^- expresses special obligation.

In our model, some of the important properties that follow immediately from the definitions areas follows:

$\models \neg (O_{i,j}(\alpha) \wedge O_{i,j}(\neg\alpha))$

$\models \neg (Do_i(\alpha) \wedge Do_i(\neg\alpha))$

The former means that the obligations one agent has to another cannot be logically conflicting. The latter means that an agent cannot intend to perform logically conflicting actions. But we allow that the following propositions may hold in some worlds:

$O_{i,j}(\alpha) \wedge O_{i,k}(\neg\alpha)$ and $O_{i,j}(\alpha) \wedge Do_i(\neg\alpha))$

This means that conflicting obligations can be generated by different agents and an agent can intend to do an action that violates an obligation.

4. Application

Our model has also been implement using IBM Aglets technology [LO97]. Aglets are Java-based software agents. There are three things that are important when using our model to do an application. The first is to identify the different obligations in the system; the second is to identify and model the entities (agents) inherent in the system, the last is to identify the relationships between the obligations and the entities. As we stated, wc see a violation as an exception. So any kind of violation can be defined as a subclass of the Exception class in Java. When the violation happens, we throw a corresponding exception; afterward, we usc the *catch* statement to do exception handling. Space limitations preclude further discussion; the reader is referred to [Hu97].

5. Conclusion

We described the deontic transaction model for normative multi-agent systems, which we have implemented in Java-based Aglet technology. The main contribution of this model is in extending traditional transaction concepts and their scope to capture the autonomy of agents, which requires more support with exception occurrences and exception handling than is offered by traditional systems. We have focused on the obligations that one agent has toward another, and wc view the violations of obligations as exceptions and corrective action as exception handling. We have introduced several new concepts, such as role, obligater, obligatee, and constraints. In contrast to existing transaction models, which use a commit-abort protocol, our model uses a commit-exception-abort protocol. Other main points of the deontic transaction model are the following. It is a distributed, object-oriented model that supports all kinds of complex applications in normative systems. Also, it can provide as much flexibility as is required by the applications. Finally, it provides concurrency control by identifying transaction serializability for different obligations.

There are many issues that relate to the deontic transaction model that require further investigation. We have focused on the obligations one agent has to another; we did not discuss obligations that do not have a particular obligater. Also, we did not discuss other kinds of normative concepts, such as permissions and prohibitions (although these are addressed to an extent in [Hu97]). Again, the formalization of the semantics of transactions, beyond the semantics of actions, must be completed. And temporal dependencies amen transactions are not currently captured. Finally, much more of our model will be implemented using Aglets so that practical insights may fill out the theoretical landscape.

Reference

[And58] Anderson "A Reduction of Deontic Logic to Alethic Modal Logic", *Mind 671958,* pp. 100-103

[Åqv84] L. Åqvist, "Deontic Logic", in: D.M. Gabbay & F. Guenther (eds.), *Handbook of Philosophical Logic Vol. II,* Reidel, Dordrecht/Boston, 1984, pp. 605-714.

M.

[BC96] Brown, M. A. and Carmo, Jose (eds). *Deontic Logic, Agency and Normative Systems. DEON'96*: Third international Workshop on Deontic Logic in Computer Science, Sesimbra, Portugal, 11-13 January, 1996.

[BOH92] Buchmann, Alejandro Ozsu, M. Tamer Hornick, Mark Georgakopoulos, Dimitrios, and Manola, Frank A. "A Transaction Model for Active Distributed Object Systems," in [Elm92], pp. 123-158.

[Dig96] Dignum, Frank. "Autonomous Agents and Social Norms", *ICMAS '96 Workshop in Norms, Obligations, and Conventions, 1996.* Available at http://wwwis.win.tue.nl/~dignum/papers.html.

[DMW96] Dignum, F. Meyer, J.-J.Ch Wieringa, R. J., and Kuiper, R. A. "Modal Approach to Intentions, Commitments and Obligations: Intention Plus Commitment yields Obligation". In [BC96], 1996, pp.80-97.

[Elm92] Elmagarmed, A. K., (cd.), *Database Transaction Models for Advanced Applications,* San Mateo, CA: Morgan Kaufmann, 1992.

[MW93] J.-J. Ch. Meyer and R. J. Woeronga (eds). *Deontic Logic in Computer Science: Normative System Specification,* John Wiley& Sons Ltd, 1993.

[GR93] Gray, Jim and Reuter, Andres. *Transaction Processing: Concepts and Techniques.* San Francisco: Morgan Kaufmann 1993.

[Hil93] Hilpinen, R.; "Actions in Deontic Logic," in Meyer, J. J. Ch. and Wieringa, R. J. (eds.) *Deontic Logic in Computer Science: Normative System Specification.* Wiley, *1993.*

[Hu97] Huang, Yonghua, The Deontic Transaction Model in Multi-Agent Normative Systems, masters thesis, Dept. of Computer Science, North Carolina A&T State University, Greensboro, NC, Dec. 1997.

[LO97] Danny B. Lange and Mitsuru Oshima, *"Programming Mobile Agents in Java™ —With the Java Aglet API", 1997.* Available at http://www.trl.ibm.co.jp/aglets/aglet-book/index.html

[JS94] Jones, Andrew J. I. and Sergot, Marek (eds.). *Workshop Proceedings. DEON'94: Second International Workshop on Deontic Logic in Computer Science.* Oslo, Norway, *6-8* January 1994.

[SJ97] Singh, Munindar P. and Jain, Anuj K. "Using Spheres of Commitment to Support Virtual Enterprises". *4th ISPE Int. Conf. on Concurrent Engineering: Research and Applications.* Oakland Univ., Rochester MI, Aug. 20-11, 1997.

Anisotropic optical-response of Eu-doped yttrium orthosilicate

Huimin Liu, Miguel Santiago and Weiyi Jia
Department of Physics, University of Puerto Rico, Mayaguez, PR 00681

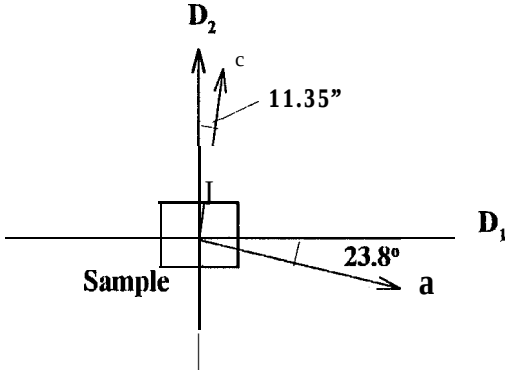
Shoudu Zhang
Shanghai Institute of Optics and Fine Mechanics, Academia Sinica, Shanghai, China

1 INTRODUCTION

Eu-doped yttrium orthosilicate (Eu^{3+}: Y_2SiO_5) had been a subject being investigated for coherent time-domain optical memory and information processing applications since its ultraslow optical dephasing was discovered several years ago [1-4]. In this crystal the weakly allowed $^7F_0$ - $^5D_0$ transition of europium ions exhibits a sufficient long dephasing time and no spectral diffusion on a time scale of several hours at low temperature, thus an information pattern or data can be stored as a population grating in the ground state hyperfine levels [5]. On the other hand, the study on photon-echo relaxation shows that the dephasing time T_2 of Eu^{3+} and other rare-earth ions doped YAG, $YAlO_3$ strongly depends on the intensity of the excitation pulses [6,7]. In Eu^{3+}:$YAlO_3$, Kroll et al [6] observed an exponential decay of photon-echo with T_2 = 53 μs if the excitation pulses are weak (5 vJ/pulse). However, when the excitation pulses are strong (80 μJ/pulse) they observed a much shortened T_2 with a highly nonexponential decay pattern. The conclusion they derived is that the intensity-dependent dephasing rate effects are quite general, and it depends on how much the excitation intensity varies. In this paper we use transient grating formation technique showing that a temporal lattice distortion may only occur along crystal c axis, caused by Eu^{3+} excitation. At high excitation level the produced exciton in conduction band may also couple to the dynamical lattice relaxation process, giving rise to an apparently much shortened dephasing time.

2 EXPERIMENTAL

Y_2SiO_5 belongs to monoclinic biaxial crystal and a C^6_{2h} space group with eight molecules per unit cell, Due to the biaxial nature, three principal axes of polarization (n_x, n_y, n_z) can be defined, and for our samples these axes were determined and is shown here.



n_x = 1.772

n_y = 1.773

n_z = 1.793

Trivalent rare-earth ions can substitute for Y^{3+} ions and occupy two inequivalent crystallographic sites with no rotational point symmetry (Cl). The crystal used in this study contains nominally 0.12 at .% Eu^{3+}. It was determined that crystal a and c axes have oblique angle of 11.350 and 23.8° with respect to the polished faces. In a degenerate- four-wave-mixing (DFWM) experiments the laser-induced transient grating was produced by using a frequency doubled $Y_3Al_5O_{12}:Nd^{3+}$ (Nd:YAG) laser as the excitation source which was operated at 532 nm with a pulse width of 30-35 psec. The two pump pulses derived from the laser are cross inside the sample to "write" an interference pattern in the form of a sinewave grating (see Fig. 1). The third pulse named as probe was aligned in a direction that was phase conjugate to the one of the pump beam and was partially diffracted by the grating, named as signal to be measured. Previous DFWM studies showed that there are in general two distinctly different contributions to the FWM signal in the temporal range of picosecond to nanoseconds [8-1 O]. The first is an instantaneous response signal with a time width of -30 ps which is close to the cross-correlation width of the two write-beam pulses. It was suggested that this signal is associated with contributions from fast electron processes, such as bound charges, free carriers. The second contribution to the FWM signal is a slow response that has a decay time corresponding either to the population grating lifetime if it is the origin of slow response, or to the lattice relaxation time if it is the origin. The transient grating formed in pure Y_2SiO_5 crystal appears to be caused by two-photon excitation of the host to create the free carriers trapped in conduction band in a grating peak. Owing to the difference in polarizability of electrons with different configurations when change from valence to conduction band, a modulation of the refractive index is created with the sinusoidal interference pattern of the two write-beams. In a Eu^3 '-doped Y_2SiO_5 crystal, however, the 532nm laser pulse can nearly resonantly excite europium ions if the population at $^7F_1$ level is taken into account. The population grating of Eu^{3+} ions will also contribute to the

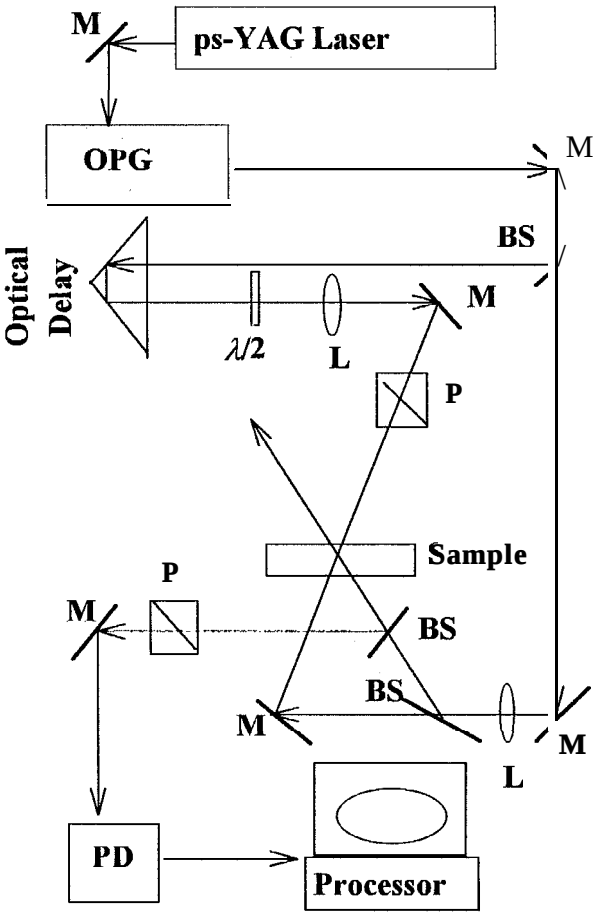


Figure 1. Schematic of the DFWM system.
M - mirror, P - polarizer, BS - beam splitter,
L - lens, PD - photodetector,

transient grating signal. The decay pattern of the diffracted signal, in the former case depends the lifetime of the trap while in the latter case, depends on the excited state lifetime. From grating peak-to-valley, the absolute diffraction efficiency can be measured and the induced change in susceptibility $\Delta\chi$ can be obtained through the expression [7]

$$\eta - \left(\frac{\pi d}{2n\lambda\cos\theta}\ \Delta\chi\right)^2 exp(-2kd) \qquad (1)$$

where 2$ is the crossing angle of the two write-beams, d is the grating thickness, and η is diffraction efficiency. The exponential factor in Eq. (1) represents the absorption loss and can generally be neglected for YSO crystal.

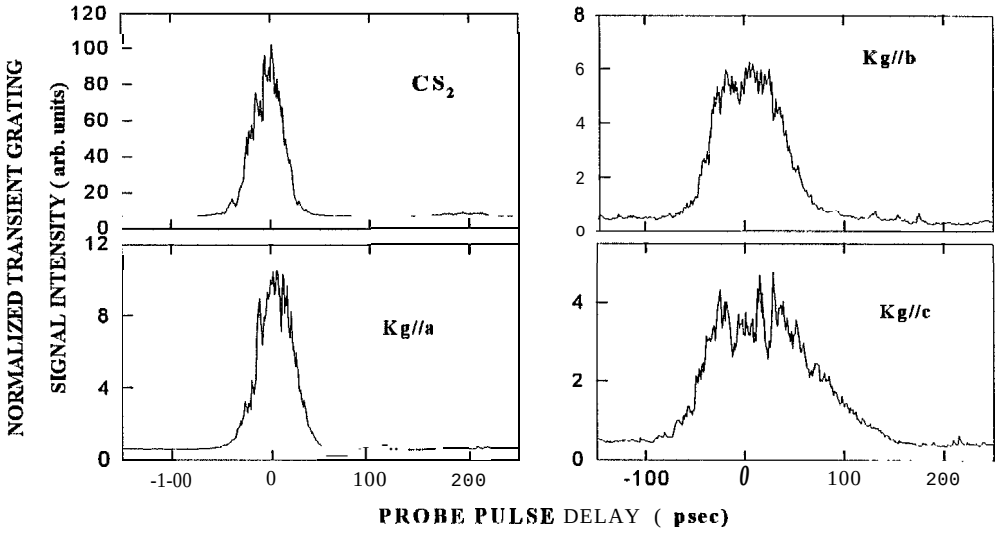


Figure 2. Normalized transient grating signal of YSO crystal shows a strong orientation dependence. CS_2 signal represents a standard optical response.

3. ANISOTROPY OF OPTICAL RESPONSE

Figure 2 shows the optical response signal of Eu-doped yttrium orthosilicate crystal as a function of delay time of the probe pulse in a different orientations configurations. The polarization of all laser pulses was vertical, perpendicular to the incident plane of two pump pulses with the polarization configuration of (ssss). From the figure it is seen that the instantaneous-response-signal intensity as well as the signal width varies when crystal orientation changes. As shown in Fig. 2 for **Kg// a**, where the optical electric field **E** of the pump pulses is parallel to crystal c axis (**E // c**)and the formed grating vector **Kg// a**, the obtained coherent nonlinear-optical-response profile is similar to that of CS_2 standard (up-left). Its full width at half maximum (FWHM) is ~50 ps, slightly broader than that of CS_2 standard, which is close to the autocorrelation width of the three pump pulses. The

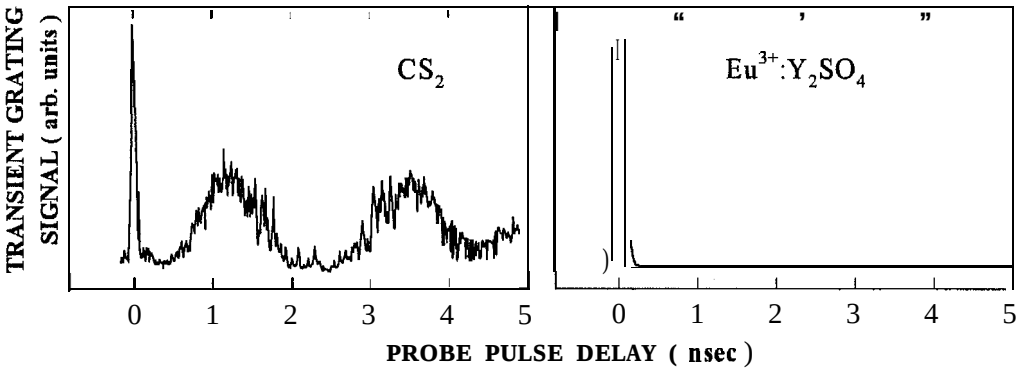


Figure 3. DFWM profile of Eu-YSO crystal compared to CS_2 reference,

signal intensity reflects the strength of the third-order susceptibility $\chi^{(3)}$ of the crystal. Note that in this configuration of the polarization, the linear optical absorbance at 532nm is greater than any other configuration. It varies as 0.7 :2.4:2,6 when the polarization changes as **E // a : E// c : E// b.** Furthermore, in the case where **E // c** and **Kg // b,** the obtained diffraction pattern is similar to the above case except a slight reduction in the diffraction efficiency. Rotating the crystal by a small angle where both **E // c** and **Kg // b** are slightly offsetting, the observed diffraction pattern become broadened. To carefully compare each with other diffraction signal pattern we find that the diffraction signal was broadened most obviously only in the case when **Kg // c.** The extent of broadening, on the other hand, was found to depend on the laser intensity. In the case of **Kg // c,** at the total incident laser intensity less than 20 mJ/cm^2 the obtained signal profile is a slightly broadened coherent peak followed by a weak but long lived tail. Resolve the long-lived component from the coherent peak, the decay rate of the tail was estimated to be in the order of 10^{-6} to 10-8 /s, which is associated with the excited state relaxation process of Eu^{3+} ions. From ground 'F to excited ^5D states the electronic configuration changes, resulting in deformation of the local structure. In the distorted octahedral cells where Eu^{3+} ions replace Y^{3+} in the center, while the central ions changes its electronic configuration realignment of the adjacent ligand must happen in order for the energy match. Further distortion along the long axis (c-axis) seems likely to occur based on our measurement. However we do not know if the further distortion is towards elongation, or in opposite, towards shortening. This kind of so-called lattice relaxation related optical response was detailed in the literature [9, 11], where the change in polarizability $\Delta\alpha$ due to the distortion Δr in one direction can be expressed as

$$\Delta\alpha = (3r^2/\xi + 1)\Delta r \qquad (2)$$

where r is the lattice constant of the crystal and ξ is a crystal dependent parameter. In our experiment Δr is actually the effective lattice distortion Δ_{eff} along **Kg** direction, and its projection along crystal c-axis is

$$\Delta r = \Delta_{eff} \cos\phi. \quad (3)$$

The diffracted signal intensity is proportional to the square of the polarizability change and therefore the diffracted signal can be given by

$$\eta = \eta_0 \cos^2\phi \quad (4)$$

where η_0 represents the maximum diffraction efficiency due to distortion at ϕ = O. It involves factors such as write-beam crossing angle, wavelength, laser fluence and pulse duration etc. In this case, the third-order susceptibility of $\chi_{xxxx}^{(3)}$ was calculated to be 6.9×10^{-15} esu.

When the laser intensity increases, except the contribution due to Eu^{3+}-induced distortion another contribution from exciton related lattice distortion becomes possible,. At the excitation intensity greater than 50 mJ/cm^2 the two-photon absorption has to be taken into account even it is known to be vet-y weak. The electron excited from valence band into conduction band, behaving as free carrier, will be trapped at some deformation center leaving hole in the valence band. This kind of exciton is mobile but heavy with radius of the volume within wavefunction comparable with the lattice constant. A strong coupling between electron and hole results in localization, forming a small polaron. Once again, the configuration change-induced lattice distortion results in change of polarizability, which gives an additional diffraction pattern in a pump-probe experiment. The stronger the excitation intensity, the higher the carrier density is, the greater the contribution due to this mechanism will be observed, The rate of signal decay depends on the lifetime of this type of exciton, or on the lattice relaxation time. In our experiment the decay rate was estimated to be faster than 10^{-10}/s at relatively lower excitation level. The apparent decay rate seemed to increase with excitation intensity. However, the actual lattice relaxation time does not show significant change because at relative lower excitation level the weaker broadened signal is added on the tail of the coherent response signal while at higher excitation level the broadening enhanced is comparable to the coherent signal component resulting in a much broad signal with sharp wing on the right side. In Eu^{3+}-doped Y_2SiO_5, the two mechanisms of lattice distortion are actually combined. The net decay rate is therefore to be excitation-intensity dependent, and it seems to decay faster when the laser pump power increases.

This project is supported by NASA-MURC NCCW-0088, US ARO-DAAH04-96-10416, US DoE(DE-FRO2-94ER757640) and NSF-EHR-9108775. S. Zhang is supported by NSF of China.

REFERENCES:

1) H. Sasaki and K. Karaki, Appl. Optics, 36 (1997) 1742.
2) R.W. Equall, Y. Sun and R.L. Cone, R.M. Macfarlane, Phys. Rev. Lett., 72 (1994) 2179.

3) R. Yano, M. Mitsunaga and N. Uesugi, Jour. Opt. Soc.Am.B, 9 (1992) 992.
4) X.A. Shen and R. Kachru, Opt. Lett., 21 (1996) 2020.
5) M. Zhu, W.R. Babbitt and C.M. Jefferson, Opt. Lett., 20 (1995) 2514.
6) S. Kroll, E.Y. Xu and M.K. Kim, M. Mitsunaga and R. Kachru, Phys.Rev.B, 41 (1990)1 1568.
7) J. Huang, J.M. Zhang, A. Lezama and T.W. Mossberg, Phys.Rev.Lett., 63 (1989)78.
8) H. Liu, R. C. Powell, and L. A. Boatner, J. Appl.Phys., 7020 (1991).
9) H. Liu, R. C. Powell, L. A. Boatner, Phys.Rev.B. 442461 (1991).
10) H. Liu, B. Taheri and W. Jia, Phys.Rev.B. 49 10166(1994).
11) H. Liu, R. C. Powell, L. A. Boatner, Opt.Mat. 2175 (1993).

Evaluation of Manganese doped $Ca_5(PO_4)_3F$ as a near infrared (1-2 μm) solid-state laser material

Matthew Turner and Uwe Hömmerich*
Hampton University
Department of Physics
Research Center for Optical Physics
Hampton, VA 23668

George B. Loutts
Norfolk State University
Center for Materials Research
2401 Corprew Avenue
Norfolk, VA 23504

Abstract

Tunable solid-state lasers are of enormous interest for applications including fundamental spectroscopy, remote sensing of the earth atmosphere, medical surgery, and optical communications. Efficient and widely tunable lasers have been developed for the 800-1100 nm region based on transition metal doped insulators like e.g. Ti: Sapphire. The development of transition metal lasers operating at longer wavelength, however, has been limited by the luminescence efficiency of existing materials. We are currently evaluating Mn doped $Ca_5(PO_4)_3F$ as a new solid-state laser material for the 1-2μm region. Preliminary spectroscopic studies revealed that Mn: $Ca_5(PO_4)_3F$ exhibits an intense near infrared luminescence which extends from ~1100-1300 nm. Based on lifetime measurements we estimated the luminescence quantum efficiency to be as high as 90°/0 at room temperature. The near infrared luminescence properties of Mn doped $Ca_5(PO_4)_3F$ and its potential for solid-state laser applications will be discussed in detail,

Keywords for Indexing:

solid-state laser, optical communications, remote sensing, transition metal, tunable lasers, infrared luminescence

Introduction

The development of tunable solid-state lasers for the near infrared region (1-2 pm) is of significant current interest for applications in laser remote sensing and optical communications. Efficient and broadly tunable lasers have been developed for the 800-1100 nm region based on transition metal doped insulators like e.g. Ti:Sapphire, Cr:$BeAl_2O_4$, and Cr:LiSAF/LiCAF [1]. The development of efficient transition metal lasers operating at longer wavelength, however, has been hampered by the low luminescence quantum efficiency of existing materials.

Promising candidates for near infrared laser operation are insulating crystals doped with $3d^2$ transition metal (TM) ions (e.g. Cr^{4+}, Mn^{5+}, and Fe^{6+}). $3d^{2+}$ TM ions have interesting optical properties which make them potential active centers in solid-state laser materials. In a tetrahedral oxygen coordination, $3d^2$ ions exhibit strong visible absorption bands arising from the electric dipole-allowed transition $^3A_2 \rightarrow {}^3T_1$ [2]. The strong absorption is attractive for both flashlamp and diode laser pumping. Crucial for laser applications is the crystal field strength the $3d^2$ TM ion experiences. Depending on whether the crystal field strength is weak or strong the metastable laser level is the $^3T_2$ or ^1E state, respectively. In the weak crystal field case the luminescence of the $3\,d^2$ ion is a broad band and laser activity is described by a quasi four level scheme (e.g. Cr^{4+} lasers). Contrary, in the strong crystal field case the luminescence is typically a narrow band and the laser operates in a three level mode similar to the well known ruby laser [1]. Interesting for laser applications is also an intermediate crystal field strength in which case the laser level is an admixture of ^1E and $^3T_2$ levels leading to a quasi four level scheme [2].

Among $3d^2$ TM ions, Cr^{4+} has received most attention and Cr^{4+}:forsterite [3,4] as well as Cr^{4+}:YAG [5] lasers have become commercially available. The efficiency and operation of Cr^{4+} laser materials, however, is limited by nonradiative decay which reduces the room temperature quantum efficiency to less than 20%. These drawbacks of Cr^{4+} lasers have stimulated research activities focusing on finding host materials for other $3d^2$ ions such as Mn^{5+} [6-10]. Laser action has been reported in Mn^{5+}:$Sr_5(VO_4)_3F$ [8,9] and Mn^{5+}: $Ba_3(VO_4)_2$ [10] but the efficiency was low because of poor crystal quality and excited state absorption. In this paper we will discuss preliminary results of the near infrared luminescence properties of a new Mn^{5+} host material, namely $Ca_5(PO_4)_3F$, and we evaluate its potential as a solid-state laser material.

Experimental Details

The investigated Mn doped $Ca_5(PO_4)_3F$ crystal was grown at the Center for Materials Research at Norfolk State University using the Czochralski growth method. The nominal Mn concentration in the melt was 1%. The tetrahedral PO_4^{3-} complex in $Ca_5(PO_4)_3F$ is well suited for the incorporation of Mn^{5+} ions. The crystal exhibited the characteristic blue color as observed in other Mn^{5+} doped insulators [6-10]. The degree of coloration, however, varied significantly along the crystal axis indicating an inhomogeneous Mn^{5+} doping profile.

The room-temperature absorption spectrum was recorded using a Cary5E Spectrophotometer. The 514nm output of a Nd:YAG pumped Optical Parametric Oscillator (OPO) was used as the excitation source for luminescence spectra and lifetime measurements. The luminescence was dispersed by 1 meter monochromator using a grating blazed at 1.6 μm. Detection was achieved with a high sensitivity, liquid nitrogen-cooled Ge-detector. The signal was digitized by a Boxcar Averager and stored in a personal computer. For luminescence lifetime

measurements the high sensitivity Ge-detector was replaced with a faster responding InGaAs detector. The decay transient was averaged by a digitizing oscilloscope and processed using a personal computer. Temperature dependent measurements were carried out using a closed-cycle helium refrigeration system which allowed the temperature to be varied between 12K and 300K. The experimental setup is shown in Figure 1.

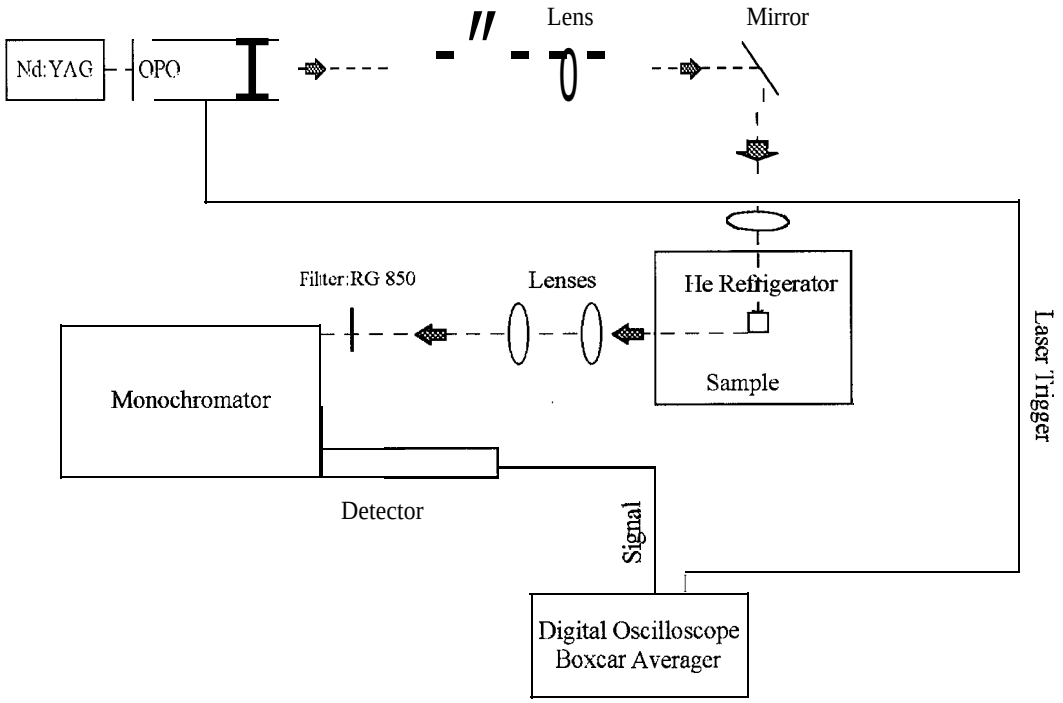


Figure 1: Experimental setup for luminescence spectra and lifetime measurements.

Results and Discussion

Energy levels of tetrahedrally coordinated Mn^{5+} ions in insulators

Mn^{5+} has two d-electrons in the unfilled (3d) shell. The ground state of the free Mn^{5+} ion is a ^3F state (Russel Saunders notation). The next higher energy levels are ^1D, ^3P, and ^1G. Optical transitions of Mn^{5+} occur within the unfilled 3d shell. Mn was doped into $Ca_5(PO_4)_3F$ with the intent to achieve a tetrahedral oxygen coordination of Mn^{5+}. The oxygen ions form a so called "ligand" or "crystal" field which modifies the free ion Mn^{5+} energy levels [2]. The influence of the ligand field on the Mn^{5+} energy levels is well described by ligand field theory [2]. The left side of Figure 2 shows a schematic of the local Mn^{5+} environment in a tetrahedral ligand field. The effect of the ligand field on the free ion energy level for a $3d^2$ system has been calculated by Tanabe-Sugano as a function of the crystal field strength parameter (Dq), and the Racah parameters (B and C) (see Figure 2), Dq is directly related to the interatomic distance R between the Mn ion (central ion) and the four surrounding oxygen ions (ligands), i.e., $Dq \sim 1/R^5$ [2]. Therefore, the shorter R is, the stronger is the crystal field and its effect on the free ion level. The Racah parameters B and C consider variations of the Mn^{5+} energy levels due to covalency effects, and are typically determined from the experiment [2].

The Tanabe-Sugano diagram in Figure 2 shows that the free ion energy levels are obtained for Dq/B = O. The effect of the crystal field (Dq/B > O) leads to the splitting of the ground state into three energy levels: $^3A_2$, $^3T_2$, and $^3T_1$. The order of the energy levels is $^3A_2 < {}^3T_2 < {}^3T_1$. The ligand field notation "A", "E", and "T" of the energy levels is obtained from group theory and indicates the orbital degeneracy of 1 (non-degenerate), 2, and 3, respectively. Distortion of the tetrahedral site can remove the orbital degeneracy of the E and T states. Ligand field theory considerations will be used in the following discussions to identify the 5+ valence state of Mn in $Ca_5(PO_4)_3F$ and to discuss its energy levels.

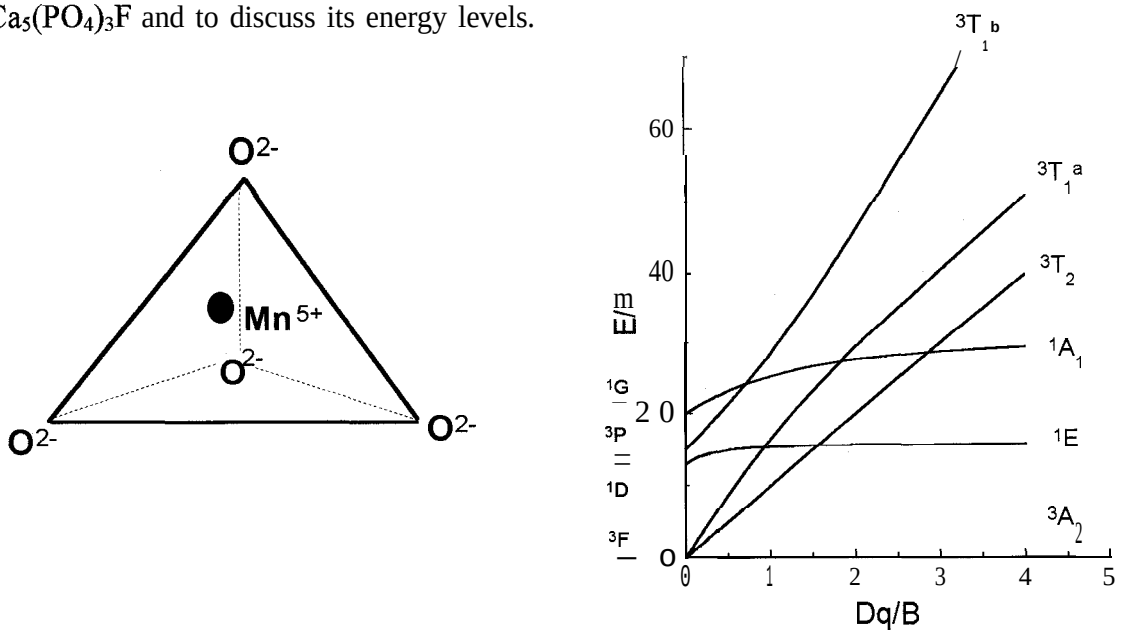


Figure 2: The left side shows the incorporation of Mn^{5+} in $Ca_5(PO_4)_3F$. The right side depicts the Tanabe-Sugano diagram of the lowest energy levels of a $3d^2$ ion in tetrahedral symmetry [2].

Absorption and Emission Spectrum

Group theory predicts [2] that the only electric-dipole allowed transition of a $3d^2$ TM ion is $^3A_2 \rightarrow {}^3T_1$. Therefore, a characteristic absorption feature for $3d^2$ ions is a strong and broad visible absorption [3- *10*]. Another "fingerprint" for $3d^2$ ions is a weaker infrared absorption which arises from the electric dipole forbidden transition $^3A_2 \rightarrow {}^3T_2$. Figure 3 shows a survey of the unpolarized absorption and emission spectrum of Mn doped $Ca_5(PO_4)_3F$. A striking feature of the absorption spectrum is the broad band spanning the entire visible region and extending into the near infrared. This strong absorption is very attractive for either flashlamp or diode laser pumping at ~800nm. A significantly weaker absorption is observed in the longer wavelength region (-900- 1200nm). Since the absorption spectrum exhibits the two characteristic features expected for $3d^2$ ions, we conclude that Mn was successfully incorporated into the crystal lattice as Mn^{5+}. Further support for this conclusion is provided by the luminescence properties of $Mn:Ca_5(PO_4)_3F$. Figure 3 shows the near infrared luminescence spectrum of the crystal at room temperature. The spectrum spans the region from 1100-1300 nm, and exhibits intensive emission at -1155 nm (zero-phonon line) and some weaker vibronic structure around 1200 nm and 1300 nm. The emission spectrum is rather narrow and suggest that Mn occupies a strong crystal field (Dq/B>1. 5) with the ^1E being the metastable level.

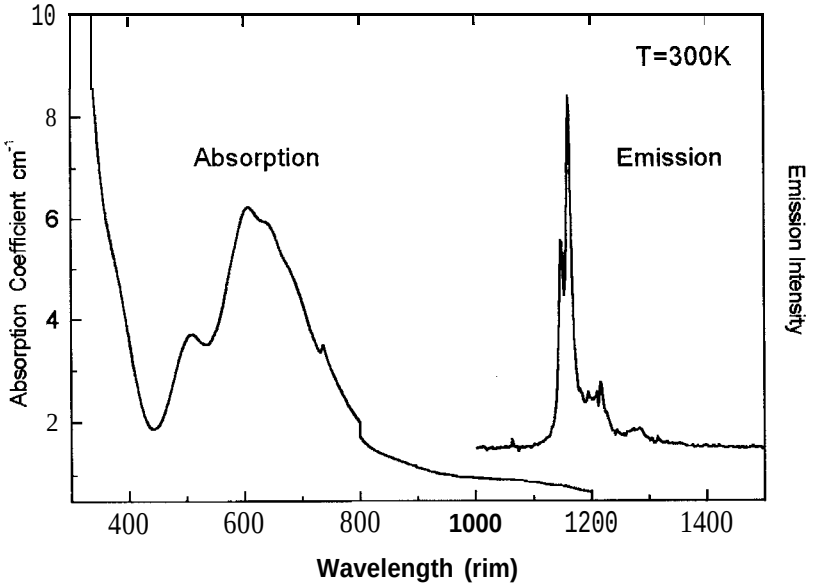


Figure 3: Room temperature absorption and emission spectra of Mn^{5+}:$Ca_5(PO_4)_3F$.

Luminescence Lifetime and Quantum Efficiency

Examples of luminescence transients and the temperature dependence of the decay time are shown in Figure 4. The luminescence lifetime is approximately single exponential which indicates only one emitting center. The l/e lifetime at T = 12 K and T = 300 K was 745μs and 643 μs, respectively. The small change in lifetime with respect to temperature suggests a high quantum efficiency of the near infrared luminescence (~90%).

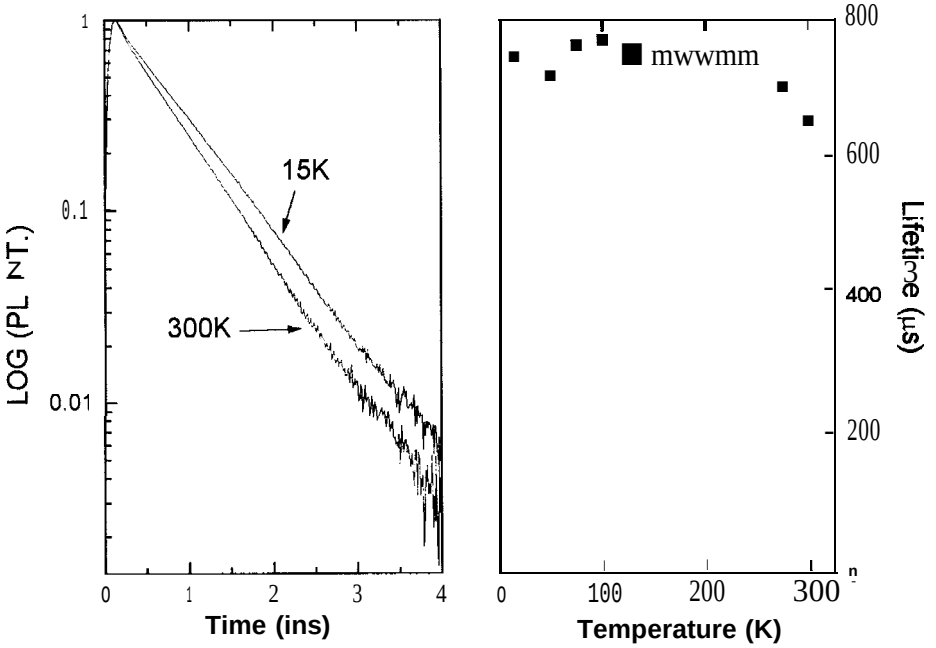


Figure 4: The left side shows examples of the luminescence transient at 12 and 300 K. The right side depicts the decay time as a function of temperature for the range 12-300K.

The infrared luminescence of Mn doped $Ca_5(PO_4)_3F$ is similar to observations made for other Mn^{5+} doped crystals [3-10]. The energy position, narrow band shape, and long lifetime of the luminescence is consistent with the spin-forbidden $^1E \rightarrow {}^3A_2$ transition of a $3d^2$ TM ion in a strong crystal field environment. Therefore, possible laser operation at ~1155nm would be ascribed to a 3 level system. We are currently investigating if heating the crystal allows a thermal population of the $^3T_2$ level which would result in a broadening of the luminescence spectrum and allow a quasi 4 level laser operation.

Conclusions

The spectroscopic investigation of Mn doped $Ca_5(PO_4)_3F$ showed a broad absorption spectrum and a strong near infrared luminescence characterized by an intense peak at -1155 nm. The absorption and luminescence features indicate that Mn was incorporated into the crystal lattice as tetrahedrally coordinated Mn^{5+}. A luminescence quantum efficiency of= 90% was estimated from temperature dependent lifetime measurements. The high luminescence quantum efficiency in conjunction with the fact that Mn^{5+}: $Ca_5(PO_4)_3F$ can be grown with a high optical quality, makes this system attractive as a potential solid-state laser material. Tunability of this laser system, however, will be limited by the narrow luminescence band. Laser experiments under pulsed pumping conditions are currently in progress. In addition, luminescence experiments at temperatures greater than 300 K will be performed in an attempt to induce broad band luminescence from the $^3T_2$ level. This would provide the prospect of obtaining tunable laser activity from Mn^{5+} : $Ca_5(PO_4)_3F$.

Acknowledgments

Matthew Turner and Uwe Hömmerich would like to acknowledge financial support of NASA through Grant NCC-1-251.

References

[1] Peter F. Moulton, Proceedings of the IEEE, **Vol. 80, No. 3, (1992) 348.**
[2] B. Henderson and G. F. Imbusch, Optical Spectroscopy of Inorganic Solids, (Clarendon Press, Oxford, 1989).
[3] V. Petricevic, S. K. Gayen, and R. R. Alfano, Appl. Phys. Lett. 52 (1988) 1040 and Appl. Phys. Lett. 53 (1988) 2590.
[4] H. R. Verdun, L. M. Thomas, D. M. Andrauskas, T. McCollum and A. Pinto, Appl. Phys. Lett 52 (1988) 2593.
[5] A. P. Shkadarevich, in OSAProceedings on Advanced Solid-State Lasers, Optical Society of America, Washington DC (1989) **Vol 5, 60.**
[6] U. Oetliker, M. Herren, H. U. Gudel, U. Kesper, C. Albrecht, and D. Reinen, J. Chem. Phys. **100** (1994) 8656.
[7] U. Hömmerich, H. Eilers, and W. M. Yen, Chem. Phys. Lett. 213 (1993) 163.
[8] L. D. Merkle, A. Pinto, H. R. Verdun, and B. McIntosh, Appl. Phys. Lett. 61 (1992)
[9] L. D. Merkle, Y. Guyot, and Bruce H. T. Chai, J. Appl. Phys. 77 (1995) 474.
[10] L. D. Merkle, B. Zandi, and Bruce H. T. Chai, Avanced Solid-State Lasers, Technical Digest, Optical Society of America, Washington DC (1996) 320.

98 URC153

FABRICATION AND CHARACTERIZATION OF THIN FILM ION IMPLANTED COMPOSITE MATERIALS FOR INTEGRATED NONLINEAR OPTICAL DEVICES

S. Sarkisov[a], M. Curley[a], E.K. W illiams[b], A. Wilkosz[a], D. Ila[b], D.B. Poker[c]
D. K. Hensley[c], C.Smith[d], C. Banks[d], B. Penn[d], and R. Clark[e]

'Department of Natural and Physical Sciences (Physics), 4900 Meridian Street, P. O. Box 1268, Alabama A&M University, Normal, Alabama 35762
[b] Center for Irradiation of Materials, Alabama A&M University, 4900 Meridian Street, Normal, AL 35762
[c] Solid State Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[d] Space Science Laboratory, NASA Marshall Space Flight Center, Huntsville, Alabama 35812
[e] New Mexico Highlands University, Las Vegas, NM 87701

Introduction

Ion implantation has been shown to produce a high density of metal colloids within thyn layer regions of glasses and crystalline materials. The high-precipitate volume fraction and small size of metal nanoclusters formed leads to values for the third-order susceptibility much greater than those for metal doped solids [1]. This has stimulated interest in use of ion implantation to make nonlinear optical materials. On the other side, $LiNbO_3$ has proved to be a good material for optical waveguides produced by MeV ion implantation [2]. Light confinement in these waveguides is produced by refractive index step difference between the implanted region and the bulk material. Implantation of $LiNbO_3$ with MeV metal ions can therefore result into nonlinear optical waveguide structures with great potential in a variety of device applications. We describe linear and nonlinear optical properties of a waveguide structure in $LiNbO_3$-based composite material produced by silver ion implantation in connection with mechanisms of its formation.

Experiment

The sample was made of 1 -mm thick $LiNbO_3$ (z-cut) implanted with 1.5-MeV Ag ions to a dose of 2.0×10^{16} cm^{-2}. Implantation was done at room temperature. Fig. 1 presents the distribution of the implanted silver along the depth of the sample calculated by the Monte-Carlo simulation program TRIM96 [3]. The position of the peak of the distribution defines the depth of the implanted layer at 0.41 μm and FWHM of the distribution gives an estimate of 0.24 μm for the thickness of the layer.

Light transmission along the implanted layer (treated at 500^0C) was studied using the prism coupling technique [4]. The technique is based on light tunneling from a prism with high refractive index (rutile, index is 2.8643) to the implanted light guiding layer through a small air gap (Fig. 2). The phase matching condition for light coupling is

$$n_p \sin\theta_m = N_m , \quad (1)$$

where n_p is the prism refractive index; θ_m is the angle of incidence; N_m is the propagation

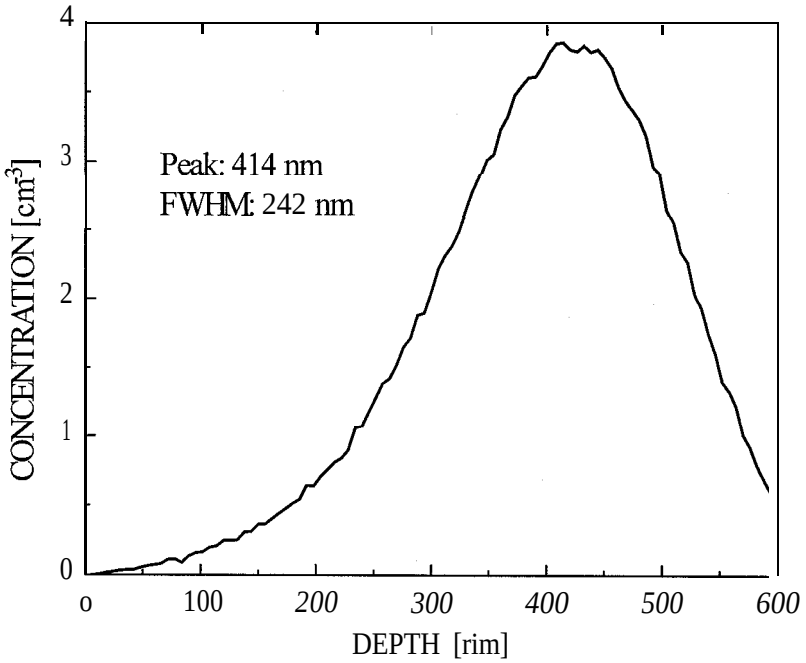


Fig. 1. The concentration of the implanted Ag ions versus the depth of the $LiNbO_3$ sample calculated by the Monte-Carlo method. The energy of the Ag ions is 1.5 MeV, the fluence is $2.0x10^{16}$ cm^{-2}.

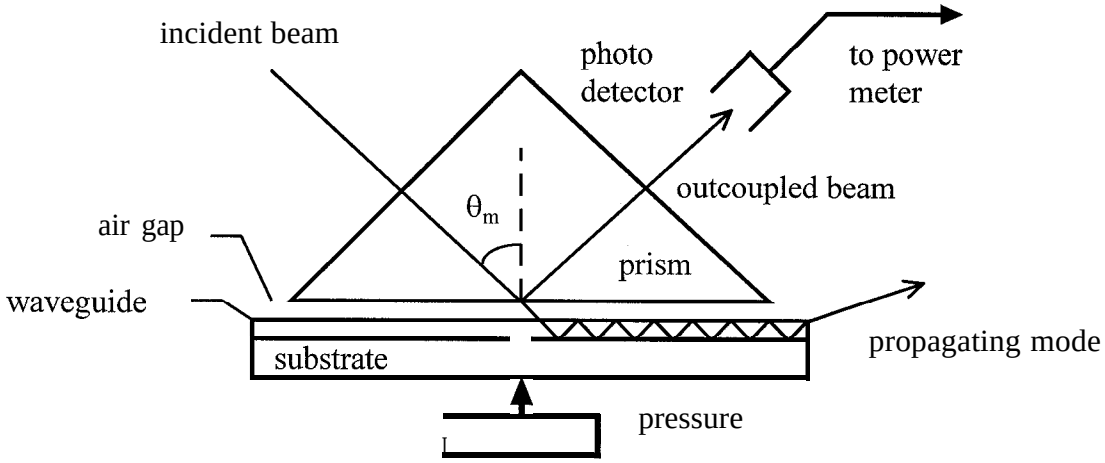


Fig. 2. Scheme of the prism coupling experiment.

number for the *m*-th order mode in the waveguide. Energy channeling to a guided mode leads to the drop of the intensity of the light outcoupled from the prism which is monitored with a photo detector. Measuring the intensity dependence on the incidence angle, it is possible to obtain θ_m and then to calculate N_m. The implanted layer refractive index n_f and thickness h_f can be found from the solution of the system of dispersion equations at least for two modes. For *s* - polarization (TE - modes) these equations can be written as:

$$\frac{2\pi h_f}{\lambda}(n_f^2 - N_m^2)^{1/2} - \tan^{-1}\left(\frac{N_m^2 - n_c}{n_f^2 - N_m^2}\right)^{1/2} - \tan^{-1}\left(\frac{N_m^2 - n_s^2}{n_f^2 - N_m^2}\right)^{1/2} - m\pi = 0 \qquad (2)$$

where λ is the light wavelength, n_s and n_c are the substrate and the cladding (air) refractive indices respectively . Similar equations can be formulated also for TM-modes. The prism coupler Metricon 2010 was used in our measurements. Fig, 3 shows experimental results which

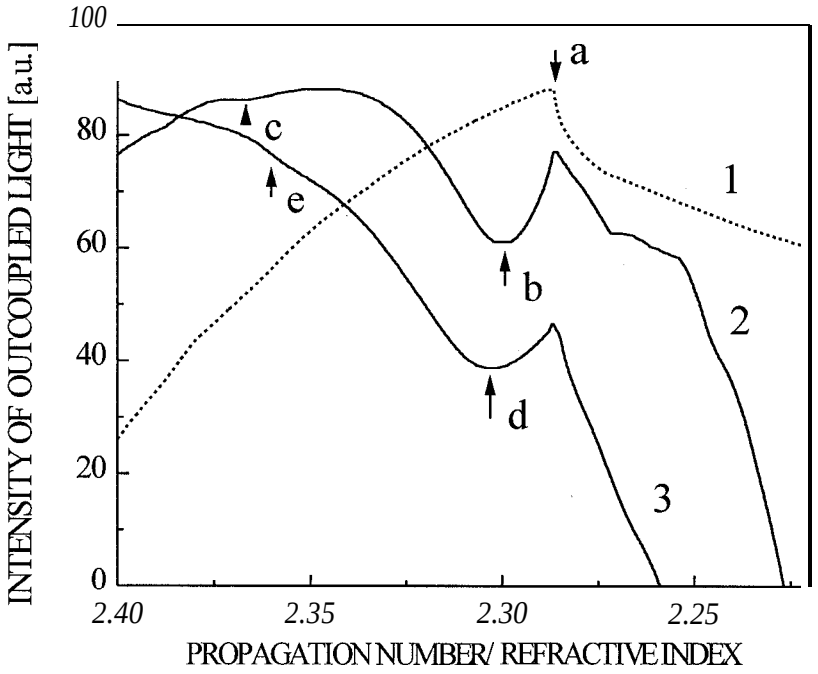


Fig. 3. Intensity of the light outcoupled from the coupling prism versus the propagation number of modes injected into the light guiding implanted layer. Curve 1 corresponds to virgin $LiNbO_3$ crystal, where point (a) marks the threshold of total internal reflection which determines the refractive index at 2.2867. Curves 2 and 3 correspond to two different locations in the implanted crystal. Points (b) and (c) mark the dips related to the zeroth order and the first order propagating modes in the first location with propagation numbers N_0= 2.3652 and N_1= 2.3003 respectively. Points (d) and (e) mark the modes N_0= 2.3938 and N_1= 2.3029 for the second location. The average refractive index and thickness of the implanted layer in the first location is calculated to be 2.3887 and 0.748 μm respectively while the same parameters for the second location are 2.4254 and 0.6228 pm. All the measurements are taken at 633 nm.

indicate that the light guiding layer with the index greater than that of the substrate (2.2867) is spread along the depth up to 0.75 μm and therefore goes beyond the limits of the 0.24-μm-wide nuclear stopping region where the implanted Ag ions were originally located (Fig. 1). This is indicative for silver ion diffusion out of the original location during heat treatment accompanied by the increase of the refractive index of the crystal.

The linear optical absorption spectra of the implanted sample are shown in Fig. 4. The spectrum right after implantation exhibits the prominent single surface plasmon peak near 430 nm which is the signature of nanometer size clusters of metallic silver formed in the host. The radius of the clusters can be estimated by using the relation $R = V_f / \Delta\omega_{1/2}$, where V_f and $\Delta\omega_{1/2}$ are Fermi's velocity of silver (1.39×10^8 cm/s [5]) and the half width of the absorption peak, respectively [6]. In our case $R \approx 1.1$ nm. Heat treatment after the implantation at 500^0C for 1 hour lead to the red shift of the absorption peak to 550 nm (Fig. 4) without significant change of $\Delta\omega_{1/2}$ (and, correspondingly, of the radius *R)*. Shang, et. al. proposed to explain this effect as a result of the volume fraction increase due to Ag precipitation near the surface [7]. Another

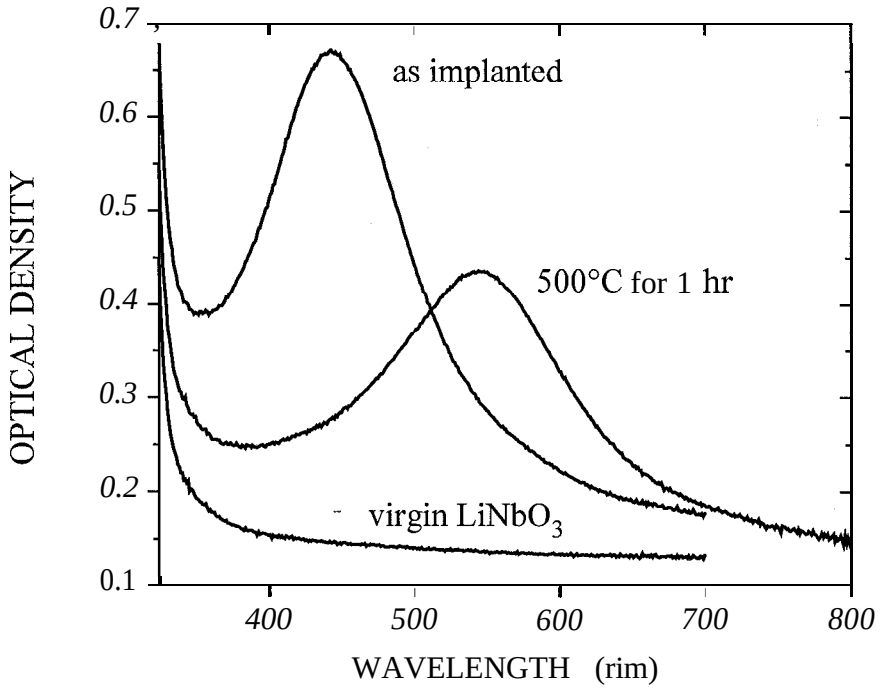


Fig. 4. Optical absorption spectra of the Ag implanted $LiNbO_3$ sample.

likely contributing factor is that the heat treatment is working to remove implantation damage [8]. $LiNbO_3$ implanted with 190 KeV Ag at $1x10^{17}$-cm^{-2} fluence has been shown to undergo full epitaxial regrowth after heat treatment at 400^0C but an anneal of 800^0C for 1 h is necessary for full removal of the implantation damage [9]. At the end of range ions implanted into $LiNbO_3$ have been shown [2, 8] to decrease the index of refraction by over 5 percent. Reducing the host index by 5 to 10 percent would result in a shift of the expected absorption peak for Ag from 520 nm to approximately 480 nm.

The nonlinear refractive index of the sample after heat treatment was characterized using the Z-scan technique [1 O]. The laser source was a tunable dye laser (with laser dye Rhodamine 6G) pumped by a frequency-doubled mode-locked Nd:YAG laser (76-MHz pulse repetition rate). The tuning range was 555 to 600 nm. The average power of the laser radiation applied to the sample varied from 100 to 350 mW. The laser pulse duration (FWHM) was 4.5±0.8 ps. The beam diameter was 3.0 mm. Estimated laser peak power density in the sample placed near the focus of the lens (125-mm focal distance) in the Z-scan experiment was 0.025 to 0.088 GW/cm^2. The closed aperture Z-scan demonstrated typical behavior of a nonlinear refractive medium with positive nonlinear refractive index [1 O]. The open aperture Z-scan showed saturation of nonlinear absorption of the sample (optical transmission peak) at the distance z = O from the focus. The measured nonlinear refractive index of the sample is plotted against the wavelength in Fig.5 (circles) . It repeats the linear absorption spectrum fraction of which is depicted by a solid line 1. This is a typical picture when the surface plasmon resonance contributes to the intrinsic nonlinear response of nanoclusters originated from interband, intraband, and hot electron photo excited transitions [1 1].

The nonlinear refractive index of the silver implanted $LiNbO_3$ sample was compared against Cu-implanted silica sample as a reference. The sample was prepared at the conditions

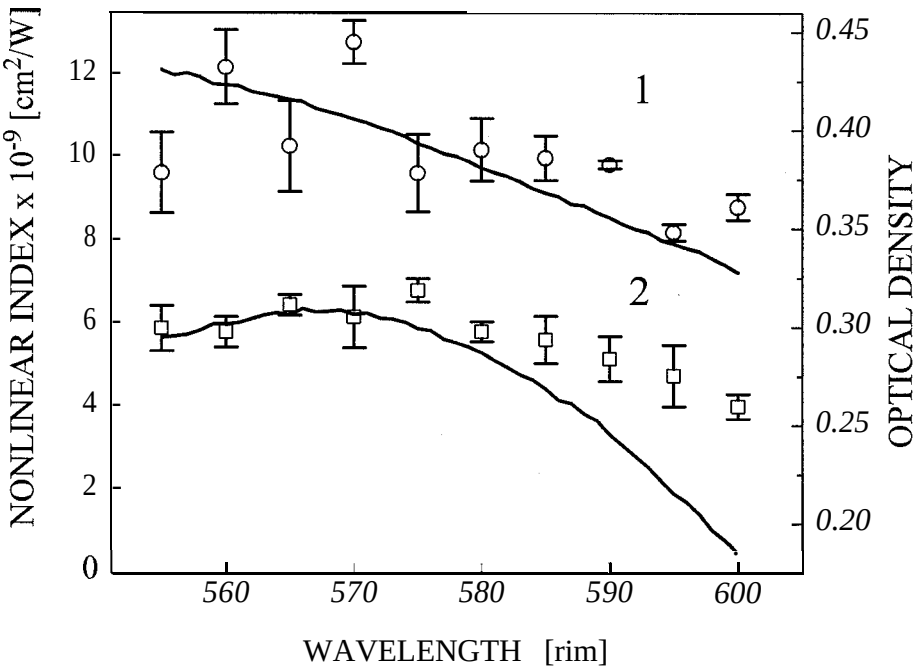


Fig. 5. Nonlinear refractive index of the Ag:$LiNbO_3$ sample (circles) and Cu:silica reference sample (squares) versus wavelength. Solid lines 1 and 2 represent optical absorption spectra of Ag:$LiNbO_3$ and Cu:silica respectively.

similar to those in Ref. 1. The energy of the Cu ions and the fluence were 2.0 MeV and 1.0×10^{17} cm^{-2} respectively. Implantation was performed at room temperature. Heat treatment was done at 1000^0C in the air for 1 hour. The nonlinear refractive index of the reference sample was characterized with the Z-scan technique at the same conditions as the Ag-implanted $LiNbO_3$ sample. It is plotted together with the absorption spectrum in Fig. 5 (squares and solid line 2 respectively). The nonlinear refractive index of the reference also repeats the shape of the Cu surface plasmon resonance peak in full agreement with Ref. 1. The nonlinear index of the Ag-implanted sample is approximately two times greater than the index of the reference. The optical density is greater as well. At the same time, the fluence used to implant the silver sample is 5 times less than the fluence used for the reference. We believe that this enhancement of the nonlinear refractive index and the optical density can be attributed to the high volumetric density of Ag nanoclusters in the ion-implanted layer remaining even after intensive heat treatment and possible recrystallization of the nuclear damage region in the $LiNbO_3$ matrix. The absolute values of the nonlinear index of both samples ($(8.21 \pm 0.20) \times 10^{-9}$ $cm^2/W \le n_2 \le (12.82 \pm 0.52) \times 10^{-9}$ cm^2/W and $(4.02 \pm 0.30) \times 10^{-9}$ $cm^2/W \le n_2 \le (6.82 \pm 0.28) \times 10^{-9}$ cm^2/W for silver and copper respectively) are at least one order of magnitude greater than those for similar materials reported in the literature [1] possibly due to the cumulative thermal self-focusing effect at relatively high pulse repetition rate (76 MHz versus 3.8 MHz in Ref. 1).

Conclusions

We have fabricated a planar nonlinear optical waveguide using implantation Ag ions in $LiNbO_3$. The implanted composite light guiding layer exhibits a Kerr-type nonlinear susceptibility. The nonlinear refractive index for the Ag:$LiNbO_3$ composite compares well to

other metal colloid composites prepared by ion implantation. It is particularly twice as high as the index for Cu:silica composite.

Acknowledgments

This research was supported by the Department of Natural and Physical Sciences, the Center for Irradiation of Materials, Alabama A&M University, Alliance for Nonlinear Optics (NASA Grant NAG5-5121), and the Division of Material Sciences, U.S. Department of Energy (Contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp.).

References

1. R.F. Haglund, L. Yang, R.H. Magruder, J.E. Wittig, K. Becker and R.A. Zuhr, Opt. Lett. 18, 373 (1993).

2. S.S. Sarkisov, E.K. Williams, D. Ila, P. Venkateswarlu, D.B. Poker, Appl. Phys. Lett. 68,2329 (1996).

3. J.F. Zeigler, J.P. Biersack and U. Littmark, The Stopping and Range of Ions in Solids (Pergamon Press, New York, 1985).

4. R. Ulrich and R. Torge, Appl. Opt. 12,2901 (1973).

5. A.E. Hugs and S.C. Jian, Adv. Phys. 28,717 (1979).

6. G.W. Arnold, J. Appl. Phys. 46,4466 (1975).

7. D.Y. Shang, Y. Saito, R. Kittaka, S. Toniguchi and A. Kitahara, J. Appl. Phys. 80, 6651 (1996).

8. E.K. Williams, PhD thesis, Alabama A&M University, 1996.

9. D.B. Poker and D.K. Thomas, Nucl. Inst. and Meth. B 39,716 (1989).

10. M. Sheik-Bahae, A.A. Said, T.H. Wei, D.J. Hagan, E.W. Van Stryland, IEEE J. Quantum Electronics, 26,760 (1990).

11. R.F. Haglund, Jr. Quantum-Dot Composites for Nonlinear Optical Applications, in Handbook of Optical Properties. Volume II. Optics of Small Particles, Interfaces. and Surfaces, edited by R.F. Hummel and P. Wismann (CRC Press, Boca Raton, 1997) p. 198.

IN SITU INFRARED SPECTROSCOPY OF THE GASEOUS SPECIES PRESENT IN A DIAMOND CHEMICAL VAPOR DEPOSITION SYSTEM

G. Morell and B.R. Weiner
University of Puerto Rico
PO Box 23323
San Juan, PR 00931 (USA)

ABSTRACT

We interfaced a Hot-Filament Chemical Vapor Deposition (HFCVD) system to the emission port of an FT-IR spectrometer, in order to study the gas phase species present during the deposition of diamond thin films. The implementation of the infrared (IR) emission technique ***in situ*** allowed the study of various carbon-containing species believed to be crucial in diamond film growth. The two IR-active vibrational fundamentals of methane, $\nu_3(f_2)$ and $\nu_4(f_2)$, were observed at three different filament temperatures: 1000, 1500 and 2000 'C. However, the net signal of ν_3 was emission, while that of ν_4 was absorption. These results indicate that the ν_4 fundamental is excited beyond equilibrium, while the ν_3 fundamental remains mostly in the ground state. This is due to the small concentration of methane, the low energy of ν_4 compared to ν_3 or to the H_2 vibrational mode, and symmetry considerations that forbid interaction among the four fundamentals of methane. Thus, the excitation of ν_3 is more likely than its decay under HFCVD conditions, producing a non-equilibrium population. At a filament temperature of 2000 "C, the ν_3 (σ_u^+) fundamental of acetylene and a band at 1328 cm-l also ascribed to acetylene ($\nu_5(\pi_u) + \nu_4$) appear in net absorption. This correlates well with the onset of molecular hydrogen breaking by the filament, which occurs at temperatures around 2000 'C and above. The hydrogen atoms produced in this heterogeneous reaction give rise to a chain of reactions that lead to acetylene, among other carbonaceous species.

I. INTRODUCTION

The growth of diamond thin films by enhanced chemical vapor deposition has been widely demonstrated.[1] However, the details of the chemical reaction mechanism that leads to diamond deposition are not well understood. The technique can be described, in general, as follows: a mixture of a carbon-containing precursor molecule (e.g, CH_4, C_2H_2) highly diluted in hydrogen passes through a high energy source (e.g., a tungsten filament at 2000 "C, or a klystron) placed close to a heated substrate (800-900 "C) at low pressures (20-30 Torr). The two leading deposition mechanisms proposed to account for the chemistry of this deposition process point at the methyl radical[2] or the acetylene[3] molecule as the main species responsible for diamond growth. To date, there is no conclusive evidence in favor of either of these deposition mechanisms.

We have implemented the ***in situ*** infrared (JR) spectroscopy technique in a Hot Filament Chemical Deposition (HFCVD) reactor used for the deposition of diamond thin films, in order to gather detailed information about the gas phase species present in the HFCVD reactor and their behavior as a function of temperature. The interpretation of these data enhances our

understanding of the basic gas-phase processes taking place during the diamond deposition reaction, and can help elucidate the actual deposition mechanism.

Methane (CH_4) is probably the most commonly employed carbon-precursor gas in diamond film deposition, and also the precursor gas of our choice. It is a spherical top molecule with four fundamental modes: ν_1 (a_1) at 2914 cm-l, ν_2 (e) at 1524 cm-l, ν_3 (f_2) at 3020 cm-l, and ν_4 (f_2) at 1306 cm^{-1}. Only ν_3 and ν_4 are IR-active.[4] They are both quite strong and of roughly the same intensity under standard IR absorption conditions.[5] From symmetry considerations, interaction among these modes is forbidden for any spherical top molecule.[6] In actual experiments, the high sphericity of the methane molecule significantly delays reaching equilibrium.[7]

Another commonly employed carbon-precursor gas is acetylene (C_2H_2), which is a centrosymmetric linear molecule. Acetylene has two IR-active fundamentals: $\nu_5(\pi_u)$ at 730 cm-l and ν_3 (σ^u_+) at 3287 cm^{-1}.[8] A third band appears in IR at 1328 cm-l, which is interpreted as arising from ν_4 (612 cm-l)+ ν_5.[8] At the conditions usually employed for ***the deposition*** of diamond ***films,*** the identity of the original carbon-precursor gas is scrambled due to the presence of a high energy source that breaks the hydrogen molecule into atoms.

We have performed a set of ***in situ*** IR experiments in a HFCVD reactor dedicated to the deposition of diamond films. We studied the behavior of a mixture of 20 Torr containing 10% methane in hydrogen at three different filament temperatures: 1000, **1500,** and 2000 'C. The spectra collected fall in the mid-infrared region from 1000 to 5000 cm-l. Analysis of these data shows fundamental differences between the processes taking place in the hot filament and the microwave chemical vapor deposition reactors.

II. EXPERIMENTAL

The Hot-Filament Chemical Vapor Deposition (HFCVD) system consists of a ultra high vacuum (UHV) stainless steel cylindrical chamber of 100 liter volume. The substrate holder sits at the center of the chamber with the filament directly above. The gases are fed from above the substrate while the chamber is pumped from below it. The distance from the filament to the substrate can be varied from zero to 10 cm. The substrate holder is heated from inside by a graphite element encapsulated in boron nitride, that can reach temperatures up to 1200 "C. It also has convection cooling for optimum temperature control. A turbomolecular pump is employed to reach base pressures down to 5 x 10^{-9} Torr before admitting the gases into the chamber. The filament consisted of a tungsten wire of 0.5 mm diameter and 30 cm length. This filament was connected to a HP stabilized DC power supply that can deliver up to 50 Volts and 40 Amps. A mixture of research grade 10% methane in hydrogen was used for these experiments at a flow rate of 100 seem. The gases were continuously pumped out by the diaphragm mechanical pump that is backing the turbomolecular pump, while the turbomolecular pump remained off.

The infrared spectroscopy measurements were carried out in a Bruker IFS 66/S FT-IR spectrometer. The infrared light from the HFCVD chamber was directed through a lateral ZnSe window into the emission port of the FT-IR spectrometer. The interface was accomplished with a set of two spherical and two planar mirrors that image the IR light from above the substrate holder and match the f-number of the spectrometer. The FT-IR spectrometer and the CVD optical interface were continuously purged with dry, CO_2-free air during the measurements.

In this ***in situ*** set-up for infrared measurements, the filament inside the HFCVD chamber acts as an IR source, thus yielding the absorption spectra of the gases around it. However, since the gases get hot by interaction with the filament, they also emit IR radiation characteristic of their

vibrational modes. Therefore, any measured spectral profile results from the combination of these two processes, so that net absorption or net emission profiles can be observed depending on the level of excitation of each vibrational mode. The IR radiation was collected at three different filament temperatures: 1000, 1500, and 2000 'C. These temperatures were measured with a 1.065 μm optical pyrometer through a quartz viewport and corrected for the non-unity emissivity of tungsten and the percent transmission of quartz. A resolution of 1.0 cm^{-1} was employed in all the experiments. The filament emission under a 2 x 10-7 Torr vacuum at each of the above temperatures was collected and used as background. The voltage and current required to reach the above temperatures were also recorded for the filament under vacuum and at 20 Torr. From these data it was possible to estimate the power delivered to the filament at each temperature, and to characterize the additional power consumption required when the gas is present.

III. RESULTS AND DISCUSSION

We have observed a significant increment in the power delivered to the filament under 20 Torr of 10% methane in H_2, compared to the power required at 2 x 10^{-7} Torr, in order to reach any given temperature (Fig. 1). This increment goes from around 200 J/s at 1000 "C to 400 J/s at 2000 'C. Therefore, this amount of energy is being transferred by direct contact continuously from the filament to the gas mixture surrounding it. In addition, a fraction of the radiated energy is absorbed by the gas before it reaches the chamber walls or escapes through the windows. This energy transferred to the gas can excite vibrational-rotational levels or can break up bonds. How this energy is distributed ultimately affects the growth rate and quality of the diamond films. This follows from the fact that filament temperatures around 2000 "C and pressures around 20 Torr are commonly reported as the optimum parameters for diamond film deposition. When these parameters are changed above or below, the chemistry that leads to diamond deposition is fundamentally perturbed, leading to a high percent of graphite deposition or no deposition at all.

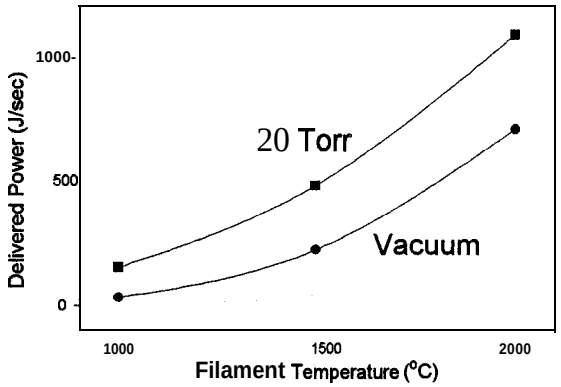


FIGURE 1. Power delivered to the filament under 20 Torr of 10% methane in H_2, compared to the power required at 2 x 10-7 Torr, in order to reach any given temperature.

As mentioned in the experimental section, any one spectral profile obtained with our *in situ* IR set-up is a combination of absorption from the filament background and the molecule's own emission. Whether it turns out to be net emission or net absorption will depend on the level of excitation of the vibrational mode under study. We have found that, at the three temperatures studied, the spectra of ν_4 is net emission while that of ν_3 is net absorption (Figs. 2 and 3). These results indicate that the ν_4 fundamental is excited beyond equilibrium, since at 1000 'C the equilibrium ground state population is around 77%, 65% at 1500 "C, and 56% at 2000 °C. The fact that emission is observed requires a ground state population below 50%, which corresponds to an effective temperature around 2400 "C or above. On the other hand, the ν_3 fundamental

should remain mostly in the ground state, from the observation that it is net absorption and very strong compared to ν_4, while they are of comparable intensity in standard absorption spectroscopy.[3] These non-equilibrium populations can be explained as due to the combination of the following factors:

1. The small concentration of methane (10% per volume) which minimizes the methane-methane interactions.
2. The fact that the energy of ν_4 is less than half that of ν_3 or the H-H stretch, thus requiring mostly four-body interactions to excite them.
3. The weak interaction among the four fundamentals of methane, due to its symmetry.

From these considerations, it follows that the excitation of ν_3 is more likely than its decay, thus producing a non-equilibrium population under HFCVD conditions.

At 2000 "C, a very weak spectral profile appeared around 3300 cm-l (Fig. 4), accompanied by another weak feature at 1304 cm-l (Fig. 2), both in absorption. We have preliminarily assigned these bands to the ν_3 (σ_u^+) fundamental of acetylene and to its ν_5 (π_u) + ν_4 combination band, respectively. This correlates well with the onset of molecular hydrogen breaking by the filament, which occurs at temperatures around 2000 "C and above.[9] The hydrogen atoms produced in this heterogeneous reaction give rise to a chain of reactions that leads to acetylene, among other carbonaceous species. That the 1304 cm-l band of acetylene is not appearing in emission but in absorption, opposite to the ν_4 fundamental of methane that falls around the same energy (Fig. 2), is easily understood from its origin as a combination mode, which makes it less likely to be readily excited.

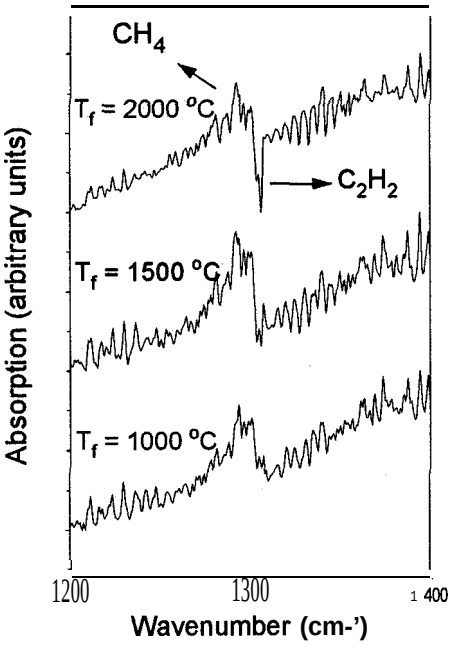


FIGURE 2. The ν_4 (f_2) fundamental mode of methane in net emission and the ν_5 (π_u) + ν_4 combination band of acetylene in net absorption (only at 2000 "C).

The above interpretation allows us to understand some fundamental differences between the Hot-Filament CVD and the Microwave-Enhanced CVD reactors. An experiment similar to the one described above was recently carried out in a Microwave-Enhanced CVD reactor. 10 In this reactor, the background IR radiation was provided by the heated substrate. Jin et al 10 found both the ν_3 and ν_4 fundamental modes of methane in net absorption and with comparable intensities. This indicates that the microwave power is not exciting any particular methane vibrational mode. They

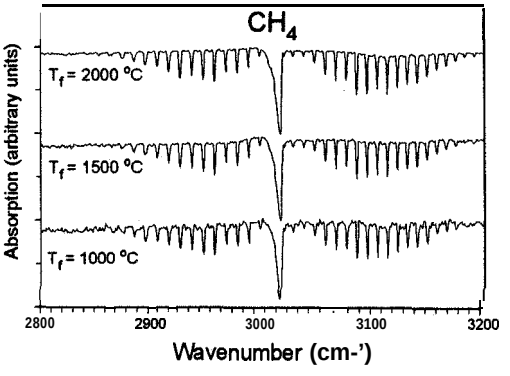


FIGURE 3. The ν_3 (f_2) fundamental mode of methane at 3020 cm-l, observed in absorption.

also found the C_2H_2 IR-active fundamentals[11] in net absorption: $\nu_5(\pi_u)$ at 729 cm-l and $\nu_3(\sigma^u{}_+)$ at 3287 cm-l. 10 Therefore, the microwave power (typically 500 Watt at 2.45 GHz) is directly targeted at breaking the H-H and C-H bonds, while scrambling the identity of the original carbon-precursor gas.

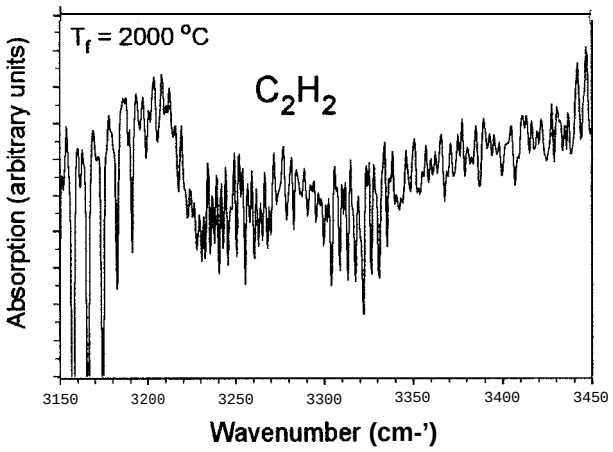


FIGURE 3. The $\nu_3(\sigma^u{}_+)$ fundamental mode of acetylene in net absorption, very weakly observed only at 2000 "C. The lines between 3150-3200 cm-l belong to methane.

IV. CONCLUSION

We have shown that the implementation of the infrared (IR) emission technique ***in situ*** allows the study of various carbon-containing species believed to be crucial in diamond film growth. The two IR-active vibrational fundamentals of methane, $\nu_3(f_2)$ and $\nu_4(f_2)$, were observed at three different filament temperatures (1000, 1500 and 2000 'C), the former in net absorption and the later in net emission. These experimental findings were interpreted as indicating that the ν_4 fundamental of methane is excited beyond equilibrium, while its ν_3 fundamental remains mostly in the ground state. This was explained in terms of energy and symmetry considerations, which indicate that the excitation of ν_3 is more likely than its decay under HFCVD conditions, thus producing a non-equilibrium population. At a filament temperature of 2000 'C, the $\nu_3(\sigma_u^+)$ fundamental of acetylene and a band at 1328 cm-l also ascribed to acetylene $(\nu_5(n")+\nu_4)$ appear in net absorption. This correlates well with the onset of molecular hydrogen breaking by the filament, which occurs at temperatures around 2000 "C and above, producing hydrogen atoms that give rise to a chain of reactions leading to acetylene.

ACKNOWLEDGEMENTS

This work received partial support from NASA through Grants NCCW-0056 and NAGW-4059. Partial support was also provided by the U.S Department of Energy Grant No. DE-FG02-94ER75764 and by FIPI (University of Puerto Rico) Grant 8-80-032.

REFERENCES

[1] For example: M.N. Yoder, in ***Synthetic Diamond: Emerging CVD Science and Technology,*** K.E. Spear and J.P. Dismukes editors, (John Wiley and Sons, New York, 1994), pp. 3-17; and P.R. Chalker and C. Johnston, Phys. Stat. Sol. (a), 154,455 (1996); and T.H. Borst and O. Weis, Phys. Stat. Sol. (a), 154,423 (1996).

[2] M. Tsuda, M. Nakajima, and S. Oikawa, J. Am. Chem. Sot., 108, 5780(1986); and S.J. Harris and A.M. Weiner, J. Appl. Phys., 67,6520 (1990); and S.J. Harris, Appl. Phys. Lett., 56,2298 (1990); and S.J. Harris and A.M. Weiner, J. Appl. Phys., 75,5026 (1994).

[3] M. Frenklach, and H. Huang, Phys. Rev. B, 43, 1520 (1991). and D. Huang, M. Frenklach, and M. Maroncelli, J. Phys. Chem., 92, 6379(1988).

[4] G. Herzberg, in Molecular Spectra and Molecular Structure, II. Infrared and Raman Spectra of Polyatomic Molecules, Krieger Publishing Company, 1991, p. 307.

[5] D.M. Dennison, Reviews of Modern Physics, 12, 175 (1940).

[6] G. Herzberg, in Molecular Spectra and Molecular Structure, II. Infrared and Raman Spectra of Polyatomic Molecules, Krieger Publishing Company, 1991, p. 39.

[7] G. Herzberg, in Molecular Spectra and Molecular Structure, II. Infrared and Raman Spectra of Polyatomic Molecules, Krieger Publishing Company, 1991, p. 41.

[8] G. Herzberg, in Molecular Spectra and Molecular Structure, II. Infrared and Raman Spectra of Polyatomic Molecules, Krieger Publishing Company, 1991, p. 288.

[9] L.L. Connel, J.W. Fleming, H.-N. Chu, D.J. Vestyck, Jr., E. Jensen, and J.E. Butler, J. Appl. Phys., 78, 3622 (1995).

10 S. Jin, L. Bourget, and E. Sevillano, Surface and Coatings Technology, 68-69,394 (1994).

11 G. Herzberg, in Molecular Spectra and Molecular Structure, II. Infrared and Raman Spectra of Polyatomic Molecules, Krieger Publishing Company, 1991, p. 288.

Formal Modeling of Multi-Agent Systems using the π-Calculus and Epistemic Logic'

Toinette Rorie and Albert Esterline
NASA ACE
Dept. of Computer Science
North Carolina A&T State University
Greensboro, NC 27411
{rorie,esterlin}@ncat.edu

1. Introduction

Multi-agent systems have become important recently in computer science, especially in artificial intelligence (AI). We allow a broad sense of agent, but require at least that an agent has some measure of autonomy and interacts with other agents via some kind of agent communication language. See [WJ95] for a good discussion of the notion of an agent. We are concerned in this paper with formal modeling of multi-agent systems, with emphasis on communication. We propose for this purpose to use the n-calculus, an extension of the process algebra CCS. Although the literature on the n-calculus refers to agents, the term is used there in the sense of a process in general. It is our contention, however, that viewing agents in the AI sense as agents in the n-calculus sense affords significant formal insight.

One formalism that has been applied to agents in the AI sense is epistemic logic, the logic of knowledge. The success of epistemic logic in computer science in general has come in large part from its ability to handle concepts of knowledge that apply to groups. We maintain that the n-calculus affords a natural yet rigorous means by which groups that are significant to epistemic logic may be identified, encapsulated, structured into hierarchies, and restructured in a principled way.

This paper is organized as follows. Section 2 introduces the n-calculus. Section 3 takes a scenario from the classical paper on agent-oriented programming [Sh93] and translates it into a very simple subset of the n-calculus. Section 4 then shows how more sophisticated features of the n-calculus may be brought into play. Section 5 discusses how the n-calculus may be used to define groups for epistemic logic. Section 6 is the conclusion.

2. The π-Calculus

The n-calculus can be used to model a system that consist of agents which interact with each other, and whose environment is constantly changing [MPW92]. The basic concept behind n-calculus is naming or reference. The names are the primary entities that refer to links or channels, and the processes, sometimes referred to as agents, are the only one other kind of entities. The calculus assumes an infinite set of names where the letters u, v, w, x, y range over the names, and a set of agent identifiers, where each one has an arity represented by an integer ≥ 0. We let A, B, c..... range over agent identifiers. Also, P, Q, R..... range over agents or process expressions, of which there are six kinds

1. A summation $\Sigma_{i \in I} Pi$ where the set I is a finite index set. The agent behaves like one or another of the Pi. The empty summation is referred to as an inaction, and represented by the symbol O. This is considered to be the agent that can do nothing. The binary summation is written as $P_1 + P_2$.

2. A prefix is presented in the form of $\bar{y}x.P$, $y(x).P$ or $\tau.P$. "$\bar{y}x.P$" is called a negative prefix. $\bar{y}$ can be thought of as an output port of an agent which contains it. "y(x)" is called a positive prefix, where y is the input port of an agent: it binds the variable x. At port y the arbitrary name z is input by $y(x).P$, which behaves like $P\{z/x\}$, where $P\{z/x\}$ indicates the result of substituting z for all free (unbound) occurrences of x in P. τ is the silent action, and $\tau.P$ first performs the silent action and then acts like P.

3. A composition $P_1 | P_2$. This is an agent that consists of P_1 and P_2 that act in parallel. P_1 and P_2 also have the ability to act independentl y.

[1] The first author has been support by grant NAG 5-4102, "Formal Foundations of Agents," form NASA/GSFC. The second author has been supported by the same grant and by grant NAG 2-1150, "Motion Planning in a Society of Intelligent Mobile Agents," from NASA/ARC.

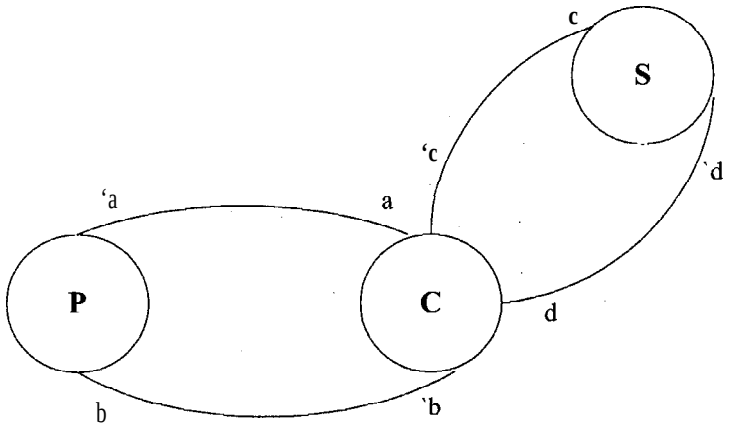
4. A restriction $(x)P$ is an agent that acts like P, and prohibits actions at ports x and $\bar{x}$, with the exception of communication that takes place between components of P along the link x. Restriction like positive prefix, binds variables.

5. $[x = y]P$ is referred to as a match, where an agent behaves like P if the names x and y arc identical.

6. A defined agent is represented by $A(y_1..., y_n)$. There must be a unique defining equation $A(x_1..., x_n) = "f P$, for any agent identifier A (with arity n) that is used, where xl..., x_n arc distinct names and the only names that may occur free in P. Then $A(y_1..., y_n)$ behaves like $P\{y_1/x_1 . . . . y_n/x_n\}$, for the simultaneous substitution of y_i for all free occurrences of x_i. (for $1< i \leq n$) in P.

3. *Modeling Agents with the π-Calculus - An Example*

We will give an example of an airline scenario taken from [Sh93] that will consist of three agents that interact with each other. The three agents are a passenger, clerk and supervisor. An example of a n-calculus translation will be given, based on the actions that take place between these agents. The example can be represented by the following diagram, where P is the passenger, C is the clerk and S is the supervisor. P sends messages to C along link a, C sends messages to P along link b, C sends messages to S along link c, and S sends messages to C along link d. The output port of a link x is indicated in the diagram as 'x not $\bar{x}$.



Airline Scenario

March

P to C: Please inform me what flights you have from San Francisco to New York on April 18.

C to P Flight #354 departs at 08:30, flight #293 departs at 10:00, flight # 441 departs at noon.

P to C: Please book me on #354.

C to P That is sold out.

P to C: Please book me on # 293.

c to P: That is confirmed, your reservation number is 112358.

P to C: Please book me also on # 441.

c to P: That conflicts with #293, I am not allowed to double book a passenger.

P to C: Please get permission to do so.

C to S: 1 request permission for the following booking.

S to C: Permission denied.

c to P: Sorry, I cannot get approval.

April 18, at the airport

P to C: My name is P, I have a reservation for flight # 293.

C to *P*: Here is your boarding pass.

π -Calculus Translation

Each step of the above scenario will be represented by two n-calculus identities, one showing the actions of the agent who sends messages and one showing the actions of the agent who receives the messages. Each identity is of the form

$$X\langle n\rangle = \sigma X\langle n+1\rangle$$

Here X is *P, C'*, or S. For *P*, $0 < n \leq 12$, for Q, $0 < n \leq 13$, and for S, $0 \leq n \leq 1$. $P\langle 3\rangle$, for example, is written as *P3*, and *PO, CO*, and SO appear as, respectively, *P, C*, and S. σ is a sequence of prefixes. At any step, one agent outputs a sequences of messages – constants in negative prefixes – on a single link and the agent at the other end of that link inputs those messages in the order sent by means of variables (x, *y*, . ..) in positive prefixes. The steps show how agent *P* evolves into agent *P1*, which evolves into *P2*, and so on. Similar evolution happens for agents *C* and S. We imagine, however, that there are really the same three agents throughout. We could reduce the number of steps by interleaving sequences of positive and negative prefixes. In this translation, the agents have minimum flexibility one agent expects an input sequence matching in structure the sequence that another outputs. Also, the contents of messages have no effect. We shall see later how to relax these restrictions.

$P = \overline{a}\ (4/18).\ \overline{a}(SF).\ \overline{a}(NY).P1$

$C = a(x).a(y).a(z).C1$

$C1 = \overline{b}\ (354).\ \overline{b}(8{:}30).\ \overline{b}(293).\ \overline{b}(10{:}00).C2$

$PI = b(x).b(y).b(z).b(w).P2$

$P2 \cdot \overline{a}(354).P3$

$C2 = a(x).C3$

$C3 = \overline{b}(\text{flight sold out}).C4$

$P3 \cdot b(x).P4$

$P4 \cdot a(293).P5$

$C4 = a(x).C5$

$C5 \cdot \overline{b}(\text{confirmed}).\ \overline{b}(112358).\ C6$

$P5 \cdot b(x).b(y).P6$

$P6 = \overline{a}(441).P7$

$C6 = a(x).C7$

$C7 = \overline{b}(\text{conflicts with } 293).\ \overline{b}(\text{cannot double book}).\ C8$

$P7 = b(x).b(y).P8$

$P8 \cdot \overline{a}(\text{get permission to double book}).\ P9$

$C8 = a(x).C9$

$C9 \cdot \overline{c}(\text{request permission for double booking}).\ C10$

$S = c(x).S1$

$S1 = \overline{d}(\text{permission denied}).\ O$

$C10 = d(x).C11$

$Cl\ 1 = \overline{b}(\text{not approved}).\ C12$

$P9 = b(x).P10$

April 18 at the airport

$P11 = \overline{a}(\text{Passengers name}).\ \overline{a}(\text{reservation for } 293).P12$

$C12 = a(x).\ a(y).\ C13$

$C13 = \bar{b}(boarding\ pass).\mathbf{0}$

$P12 = b(x).\mathbf{0}$

The system we have described is represented by

$P \mid Q \mid S$

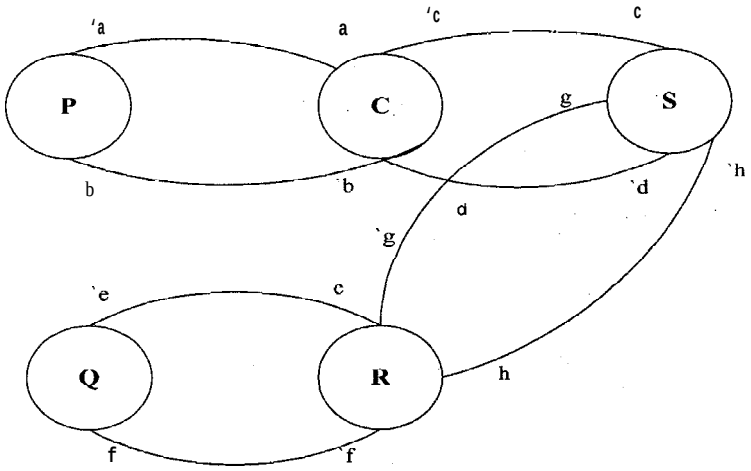
That is, our three agents execute in parallel. Since the *P-C* system is relatively self-contained, we might restrict the links. For example,

$SYS1 = (a)(b)(P \mid C \mid S)$

is a system that prohibits actions at ports *a*, $\bar{a}$, *b*, and $\bar{b}$ except for communication between *P* and C along links *a* and *b*. *When* such a communication occurs, *SYS1* performs a silent action, τ: the communication is not accessible outside the system.

4. *Enhancements to π-Calculus Modeling*

We can enhance and modify this specification in several ways. Suppose that there is another passenger Q and another clerk *R*. Suppose also that *Q* and *R* communicate in the same fashion as do *P* and C. Finally, suppose that *R* communicates with *S* just as C communicates with S. Let e be the link on which Q outputs to *R*, *f* be *the link* on which *R* outputs to Q, g be the link on which *R* outputs to S, and *h* be the link on which S outputs to *R*. We can portray this situation as follows



We have already defined the *P-C* subsystem, *SYS1*, hiding the *a* and *b links*. We now define a similar *Q-R* subsystem, restricting over links e and *f*:

$SYS2 = (e)(f)(Q \mid R)$

The only unrestricted ports in *SYS1* are $\bar{c}$ and *d*, through which it communicates with S. And the only unrestricted ports of *SYS2* are $\bar{g}$ and *h*, through which it communicates with S. It is natural, then, to view the overall system as the parallel composition of *SYS1*, *SYS2*, and S, where now we restrict over the remaining unrestricted links so that the system has no windows to the outside world:

$(c)(d)\ (''(h)\ (SYS1 \mid SYS2 \mid S)$

Parallel composition and restriction are natural operations when wc consider systems of intelligent agents. Parallel composition allows us to associate intelligent agents into group hierarchies. Restriction allows us to encapsulate these groups so that only certain ports are accessible from outside.

Several other features of the n-calculus could be used to define systems of agents in a way more flexible and responsive than in the above example. We can easily define an agent that repeats its behavior. For example, given

$P = \bar{a}(5).P,$

P continually outputs 5 on link *a*. Mutual recursion is natural in this context. For example, *P* defined as

$P = \bar{a}(5).P1$
$P1 = \bar{b}(6).P$

repeatedly outputs 5 on link *a* and 6 on link *b*. In the airline scenario above, wc could have the system repeat the same behavior again and again if we replace the O in the definition of *C13* with C, replace the **0** in the definition of *P12* with *P*, and replace the O in the definition of *S1* with S.

We can endow an agent with memory by including variables that occur in positive prefixes in the subsequent part of the agent expression. For example, agent *P* defined as

$P = a(x).Pi(x)$
$Pi(x) = \bar{b}(x).\mathbf{0}$

inputs a value, to which *x* is bound, on link a, then it outputs that value on link *b* and stops. More sophisticated uses of variables could allow one to model beliefs, commitments, and other attitudes an agent might acquire.

More flexibility can be introduced by allowing alternatives, using '+', and by using the match form, *[x=y]*, to select the appropriate alternative. For example, in the above scenario, we have the step

$P4 = \bar{a}(293).P5$
$C4 = a(x).C5$

representing the passenger's request to be booked on flight #293 and the clerk's confirmation. We could make the clerk flexible by using the variable x to select among several alternatives:

$C4 = a(x).([x=293]C5 + [x=301]C51 + \ldots)$

Here, if x (the flight requested by the passenger) is 293, then the scenario proceeds as before. If, however, x is 301, then *C4* evolves as *C51* (whatever that may be). We have indicated any number of additional alternatives. If x has a value not covered by one of these alternatives, then the agent just stops.

The final feature of the n-calculus we mention – changing structure – is perhaps its most powerful feature. Suppose that, in our scenario, the clerk connects the passenger with the supervisor when he requests to double book. Assume that the "*C9*" in the definition of *C8* has been changed to "*C91*" and the "*P8*" in the definition of *P7* has been changed to "*P81*". Then the clerk agent will send the links connecting to the supervisor agent (links c and *d*) along link *b* to the passenger agent:

$C91 = b(c).\ b(d).\ C12$

The passenger agent will bind, say, variable *x* to link c and variable *y* to link *d*. It will then output its request on link *c* (= *x*) directly, now, to the supervisor. The passenger agent will now behave like *P91(y)*, who remembers the link *d* (= *y*) from the supervisor:

$P81 = b(x).b(y).\bar{x}(\text{request permission for double booking}).\ P91(y)$

The supervisor agent will behave as before – he is not concerned with who is at the other end of the links:

$S = c(x).S1$
$S1 = \bar{d}(\text{permission denied}).S2$

The passenger now waits for the message in the link it remembers, *y* (= d):

$P91(y) = y(z).P10$

The ability to restructure is of critical importance for a system of intelligent agents. It is generally assumed that the agents with which a given agent will communicate are not determined in advance.

5. *Defining Groups for Epistemic Logic*

In systems of intelligent agents, the knowledge that can be attributed to agents and groups of agents is particularly important. We follow [ ] in the syntax and definitions of epistemic logic, the logic of knowledge. Where φ is some proposition, $K_P\varphi$ means that agent *P knows* that φ. Where G is a group of agents, $E_G\varphi$ means that every agent in group G knows that φ (that is, $K_P\varphi$ holds for all $P \in G$). We can iterate the E_G operator:

$$E_G^i \quad . \quad \begin{cases} E_G\varphi & \text{if } i = 1 \\ E_G E_G^{i-1}\varphi & \text{if } i > 1 \end{cases}$$

Now, $C_G\varphi$ means that it is common knowledge in group G that φ:

$$C_G \varphi \quad \text{iff} \quad E_G \varphi \wedge E_G^2 \varphi \wedge E_G^3 \varphi \wedge \ldots$$

Common knowledge is particularly important for showing properties of communications protocols. Finally, $D_G\varphi$ means that it is distributed knowledge in group G that φ – that is, pooling their knowledge, the agents in G can infer φ.

The n-calculus in particular allows us to identify groups of agents, possibly hierarchically structured, by supplying for agent identifiers defining equations involving parallel composition. The n-calculus also lets us isolate groups using restriction and restructure groups by communicating links.

6. Conclusion

We have indicated how the n-calculus may be used to model communication, memory, selection of alternatives, grouping, and restructuring of groups in multi-agent systems. We have also discussed how the n-calculus can help define group structures that can be exploited by epistemic logic. Considerable work remains, however, before many of the fundamental notions of agent theory can be captured in the n-calculus. Notions, such as intention, commitment, and belief, that involve persisting attitudes require more work with the detailed structures of n-calculus agent expressions. Also, the relation between the semantic structures of epistemic logics (Kripke structures) and the n-calculus remain to be investigated. Indeed, the overlap in the domains of application of epistemic logics and the π-calculus remains to be investigated even though both are applied to communicating systems.

To get a better understanding of how the n-calculus can model a multi-agent system, we are using the Mobility Workbench [VM94]. This tool can be used to manipulate, animate, and analyze concurrent systems described in the n-calculus. Through various notions of bisimulation, it can determine when two agents have the same behavior, and it can determine when an agent satisfies a modal logic expression that specifies behavior at a high level.

References

[FHMV95] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi, *Reasoning about Knowledge.* Cambridge, MA: The MIT Press, 1995.

[MPW92] R. Milner, J. Parrow, and D. Walker, "A Calculus of Mobile Processes," Parts I and II. *Journal of Information and Computation,* Vol. 100, Sept. 1992, pp. 1-77.

[Sh93] Y. Shoham, "Agent-Oriented Programming." *Artificial Intelligence,* Vol. 60, No. 1 (1993), pp. 51-92.

[VM94] B. Victor and F. Moller, The *Mobility Workbench – A Tool for the n-Calculus.* Technical Report DoCS 94/45, Department of Computer Science, Uppsala University, Uppsala, Sweden, Feb. 1994.

[WJ95] M. Wooldridge and N.R. Jennings, "Intelligent Agents: Theory and Practice." *The Knowledge Engineering Review,* Vol. 10, No. 2 (1995), pp. 115-152.

Improving Student Achievement in Math and Science

Ms. Nancy G. Sullivan, NASA, Stennis Space Center, Mississippi
Dr. Irene Schulz Hamsa, Xavier University, New Orleans, Louisiana
Dr. Panagiota (Penny) Heath, Southern University, New Orleans, Louisiana
Dr. Robert Perry, Southern University, New Orleans, Louisiana
Ms. Stacy J. White, Mississippi Valley State University, Itta Bena, MS

Introduction

As the new millennium approaches, a long anticipated reckoning for the education system of the United States is forthcoming, Years of school reform initiatives have not yielded the anticipated results. A particularly perplexing problem involves the lack of significant improvement of student achievement in math and science. Three "Partnership" projects represent collaborative efforts between Xavier University (XU) of Louisiana, Southern University of New Orleans (SUNO), Mississippi Valley State University (MVSU), and the National Aeronautics and Space Administration (NASA), Stennis Space Center (SSC), to enhance student achievement in math and science. These "Partnerships" are focused on students and teachers in federally designated rural and urban empowerment zones and enterprise communities. The major goals of the "Partnerships" include:

1. **The identification and dissemination of key indices of success that account for high performance in math and science.**
2. **The education of pre-service and in-service secondary teachers in knowledge, skills, and competencies that enhance the instruction of high school math and science.**
3. **The development of faculty to enhance the quality of math and science courses in institutions of higher education.**
4. **The incorporation of technology-based instruction in institutions of higher education.**

These goals will be achieved by the accomplishment of the following objectives:

1. **Delineate significant "best practices" that are responsible for enhancing student outcomes in math and science.**
2. **Recruit and retain pre-service teachers with undergraduate degrees in Biology, Math, Chemistry, or Physics in a graduate program, culminating with a Master of Arts in Curriculum and Instruction.**
3. **Provide faculty workshops and opportunities for travel to professional meetings for dissemination of NASA resources information.**
4. **Implement methodologies and assessment procedures utilizing performance-based applications of higher order thinking via the incorporation of Global Learning Observations To Benefit the Environment (GLOBE), Mission to Planet Earth and the use of Geographic Imaging Systems into the K-12th grade curriculum.**

Critical personnel shortages of secondary math and science teachers pose a formidable problem in the delivery of educational services throughout the nation. Universities must vigorously accept the task of preparing teachers to work effectively with the learners of the next century. Teachers can become the dinosaurs of the educational process, unable to thrive in an environment in which they have not been prepared to even exist, much less prevail. These three university partnerships are addressing the above stated objectives as described in the following brief partnership descriptions.

Partnership 1

Project Kaizen (PK) (Japanese meaning improvement) is involved with an ongoing process-oriented improvement strategy encompassing everyone in an educational environment-teachers, administrators and students.

PK will contribute to a better understanding of what policies, procedures and specific teaching methodologies should be instituted into high school math and science curricula to enhance student achievement. PK also addresses the role of institutions of higher education in educating future math and science teachers in methodologies that yield high student achievement.

Partnership 2

The NASA/SUNO Partnership for Excellence in Education is a program of reform at both the institutional and academic levels. An electronic classroom will be established to improve the way mathematics and physics are taught by incorporating technology as a main tool for learning and teaching. The Faculty Development Program at SUNO will be expanded to enhance the quality of teaching in the Mathematics/Physics Department. The underlying premise and goal of this training is to equip the faculty with the ability to teach according to the recommendations made by the National Council of Teachers of Mathematics (NCTM)-to implement technology and writing in the classroom as an essential tool for learning.

The faculty will be trained to use questioning techniques that promote critical and analytical thinking, and to use alternative assessment to evaluate student progress and enhance learning. The faculty is also being trained to provide opportunities for the students to learn mathematics and science through discovery learning and hands-on experiments. Teachers are encouraged to engage students in research and provide assignments that require students to work collaboratively—thus producing life long learners.

Partnership 3

MVSU/NASA Information Technology Partnership is an innovative two-year program using Global Learning and Observations to Benefit the Environment (GLOBE), Mission-to-Planet-Earth curriculum support materials and Geographical

Information Systems to enhance the skills of in-service and pre-service teachers and the quality of instruction in mathematics and science in the Mississippi Delta. The major goal of the program is to educate in-service secondary teachers from public schools in the Mississippi Delta and pre-service secondary teachers from MVSU on the incorporation of technology-based instruction in secondary public school systems. In-service teachers and pre-service teachers receive rigorous academic experiences that integrate NASA resource materials, hands-on instructions, inquiry learning, lectures, and discovery activities during the intensive four-week program. The criteria for selection in the workshop is in-service participants must instruct in Biology, Chemistry, Physics, Mathematics, and Computer Science on the eleventh or twelfth grade level. Pre-service teachers must be enrolled in Biology or Mathematics Education and have completed the second semester of their junior year. Pre-service participants must also have a cumulative undergraduate GPA of 3,0/4.0.

These three "Partnership" Universities are located in federally designated urban and rural enterprise communities and empowerment zones. The population served by these universities is largely socially and economically disadvantaged.

XU and SUNO are located in New Orleans, Louisiana which is designated as an urban enterprise community. The potential for reaching at-risk students is almost unlimited, These partnerships provide this student population with needed opportunities to encourage self-improvement and assist them in discovering their talents, recognizing their potential and challenging their ability for achievement at a very early stage.

MSVU is located in the Mississippi delta which is designated as a Rural Empowerment Zone. The Delta is primarily a rural agricultural farming region, producing crops such as catfish, cotton, soybeans, and corn. Students in this region historically perform poorly on standardized math and science tests. Teachers in this area have limited classroom resources but have been expected to face the difficult task of developing innovative methods for reaching these students.

Although each "Partnership" is unique, they all have common characteristics. They will each build electronic classrooms at the university, they each introduce technology into the local high schools, participants are introduced to the internet, NASA's Mission-to-Planet-Earth Enterprise and Geographical Information Systems Technology. The Global Learning and Observations to Benefit the Environment (GLOBE) Program will be used by each "Partnership" as a means for teachers and students alike to continue using the skills they have acquired and stimulate interest in math, science, and their environment.

Teachers need to incorporate methodology that stresses hands-on learning, concept development, cooperative learning, higher-order thinking skills, and learner-centered activities that transmit information holistically, instead of discrete non-connected facts, formulas and procedures (Math, Sciences Education Board 1993, Loucks-Horsley et al 1990, National Center for Improving Science Education 1989 and Tobias 1992).

The fundamental ideas, principles, and concepts of mathematics and the biological, physical, and environmental sciences can be taught by increasing the subject content information of the teacher. An emphasis on discovery and inquiry techniques is essential. Student outcomes in math and science can improve by: integrating essential content across disciplines; fostering problem-solving, promoting critical and analytical thinking; and, encouraging learners to construct understandings of math and science processes for themselves. The goal is not to implement pedagogical strategies, but rather to structure, monitor, and adjust activities in which students engage (Koehler and Grouws 1992).

Although most secondary math and science teachers may not be recipients of a Nobel Prize, they can be winners of a different kind. They can strive to be the best in their profession by educating and inspiring our most prized American possession: the children of today who will be the teachers and leaders of tomorrow.

The classrooms of the future will be vastly different. Education leaders in universities must vigorously accept the task of preparing teachers to work effectively with the learners of the next century in mastering math and science principles and concepts. Teachers cannot become the dinosaurs of the education process, unable to thrive in an environment in which they have not been prepared to even exist, much less prevail.

Acknowledgement:
Special thanks to the National Aeronautics and Space Administration for funding all projects.

References:

Math Sciences Education Board. 1993. Reshaping school mathematics: A philosophy and framework for curriculum. Washington: National Academy Press.

Koehler, M., and Grouws, D. 1992. Mathematics Teaching Practices and Their Effects. In grouws (Ed.), Handbook of research on mathematics teaching and learning. New York: Macmillan.

Tobias, S. 1992. Science education reform. Journal of Science Education and Technology, (6), 86-93.

Loucks-Horsley, S., Kapitan, R., Carlson, M., Kuerbis, P., Clark, R., Nelle, G., Sachse, T., and Walter, E. 1990. Elementary School Science for the 90's. Andover: The Network

National center for Improving Science Education. 1989. Getting started in science: A blueprint for elementary school science education. Andover: The Network.

The HSCaRS Summer Enrichment Program: Research opportunities for Minority and Women Undergraduates in Global Change Science

Mr. Maurice G. Estes, Jr., Universities Space Research Association
Dr. Donald J. Perkey, Institute for Global Change Research and Education
Dr. Tommy L. Coleman, Alabama A&M University

Background and Program Description

The Center for Hydrology, Soil Climatology, and Remote Sensing (HSCaRS) was established July 1, 1995, through a cooperative agreement between the National Aeronautics and Space Administration (NASA) and Alabama A&M University. One challenge in the Center was to develop an educational component that would increase participation by students in mainstream research and increase the production of underrepresented minorities who are U. S. citizens or permanent residents in NASA-related fields. This goal was strongly supported by a number of educational research publications. The Task Force on Women, Minorities, and the Handicapped in Science and Technology reported in the 1988 Interim Report entitled, *Changing America: The New Face of Science and Engineering,* that one of America's most urgent tasks is to strengthen the science and engineering work force. The Task Force asserts that, "by the year 2000, 85 percent of new entrants to the Nation's work force will be members of minority groups and women." According to the *Wall Street Journal,* summer internships are flourishing, with an emphasis on minorities. Bausch and Lomb, Inc., and Chevron Corporation are increasing summer internship opportunities and are focusing recruiting efforts on minority candidates. The Aluminum Company of America is focusing internship opportunities on minorities and women with engineering and technical training. The challenge is clear: More women and minority scientists must be educated to meet the needs of America's technical work force.

Based on this data, a summer research internship program with an emphasis on minority and women students was designed. Undergraduates were selected, as opposed to graduate students, We hoped to recruit outstanding students in the physical sciences and mathematics and give them a positive research experience that would encourage them to attend graduate school and pursue research careers. Also, a survey of current NASA programs for students indicated that more opportunities were available for K- 12 and graduate students than undergraduates.

The primary objective of the HSCaRS Summer Enrichment Program (SEP) is to make significant contributions to the NASA Mission to Planet Earth and HSCaRS research missions by providing undergraduate student research internships, with an emphasis on minority and women students. Additional objectives are to encourage more minority and women students to pursue advanced degrees in Earth system science and to increase the participation of minority institutions in the U.S. Global Change Research Program. Also, the SEP strives to make students in the traditional science disciplines more aware of the opportunities in Earth system science.

In designing the SEP, it was acknowledged that HSCaRS was a new research effort and Center. Consequently, students were not expected to recognize the Center immediately as one would older, more established research laboratories with national reputations, such as Los Alamos, Battelle, National Center for Atmospheric Research (NCAR), etc. Yet we still wanted to compete nationally for the *best* students. Therefore, we designed the program with a competitive financial package that includes a stipend of $400 per week, round-trip transportation from home to the summer research site, and free campus housing and meal plans provided by Alabama A&M University. The internship program was 10 weeks in residence at Alabama A&M University or IGCRE in 1996 and adjusted to eight weeks in 1997. Students had the opportunity to select from general research areas such as: micro-meteorology, soil data analysis, hydrologic modeling, instrumentation, geographic information systems, soil hydrology, computer science, etc. Student participants also enrolled in an introductory global change science course as part of the summer program (a copy of the course outline is in the appendix). The program included participation in a field program for approximately two weeks in 1996, but did not have a significant field work component in the second year. We hope to be able to offer student researchers a field experience again in 1998 as a learning experience, regardless of the relationship of the field program to their majors or particular research projects.

Recruiting and Applicant Selection

Students were recruited by distributing posters throughout the university community. All the HBCU's, Other Minority Universities (OMU's), and a mixture of other types of institutions were targeted for the distribution of program information. Other institutions included large schools with established research programs, as well as smaller institutions with a traditional emphasis on teaching, as opposed to research. Students were required to have completed the sophomore year prior to the summer internship period.

The second year the total number of applicants dropped considerably, from 102 to 34, even though more posters were distributed. We think the timing of the distribution from the late fall in the second year versus early

January in the first year may have contributed to the decline. Recruiting materials in the second year also emphasized the need for at least one course in physics prior to the internship, which may have reduced our potential applicant pool. In year two, we also relied on our website as a primary source of information on the program. Possibly, we did not encourage enough personal interaction with potential students. As we recruit students for the third year, we are taking a more targeted and personal approach with potential applicants who show interest in the SEP. A database of faculty, students, and administrators who have contacted us about the program has been established, and this group will be a priority as information is distributed. Follow-up telephone or email contact will also be done to encourage student applications from this group. Presentation of the program has been made at the National Conference on Undergraduate Research (NCUR), the Council of Undergraduate Research (CUR) and other professional conferences such as the American Geophysical Union (AGU) as appropriate.

A two-phase process was used in the evaluation of the students' applications. The first phase was an evaluation by the SEP staff to identify the strongest applications by research group. This was done by considering the applicant's Grade Point Average (GPA), statement of interest in global change science, letters of recommendation, and resume. A grading scale was developed, that awarded two points for a GPA over 3.5, one point for a GPA between 3.0 and 3.5, and zero points for a GPA less than 3.0. An additional point or fraction of a point was awarded based on the staffs evaluation of the statement of interest in global change science, letters of recommendation, and resume.

This evaluation plan allowed us to consider all major relevant factors and was weighted toward students with the best grades. However, the plan did provide an opportunity for a student with a lower GPA to still score high enough to be selected. This was considered important, since talented undergraduates sometimes do not achieve high grades early in college due to immaturity, changing interests, etc. Some of these students may still be very interested in research and have the talent to succeed in a research environment. The statement of interest in global change science offered insight into some outstanding candidates who did not have high GPA's. Finally, relevant work experience, as shown in the resume, was considered to be of value.

Based on the phase one analysis, the students were ranked by score. Phase two provided an opportunity for a potential mentoring scientist to rank, in priority order, the student applicants interested in that mentor's area of research. This data was used, with program goals in mind, to make final selections. The program goals that most prominently influenced the evaluation process related to the identification of students who would likely be able to make research contributions and who would be interested in graduate school.

Results

Overall, the second year was very successful. The SEP milestones for year two were as follows:

- Recruit outstanding minority male and female students and other female students as a priority
- Recruit outstanding students from both minority and majority institutions
- Successfully encourage several outstanding students from the 1997 class to return for a second summer of research
- Have at least two undergraduate students co-author papers with mentor scientists
- Have two undergraduate students enroll in graduate programs relating to Earth system and global change science.

The program has been successful in attracting high quality minority and female students from both minority and non-minority institutions. The average grade point average (GPA) for all applicants has been above average in each of the first two years: 3.30 in 1997, and 3.07 in 1996. Minority institutions, especially HBCU's, have been a primary source of student applicants in each year. In 1997, 44% of the applicants were enrolled in minority institutions, and 79% in 1996.

The class of 1997 was comprised of 18 students, who were selected from an applicant pool of 34. A statistical comparison of the 1996 and 1997 student intern classes is provided in the Table 1.

At the end of the summer term, each student was required to prepare a written paper on his/her research and present the results in a seminar setting. Based on this data and feedback from mentoring scientists, all of the students made contributions to the research effort. Our plans are to invite back for a second summer of research several of the sophomore- and junior-level students from the 1997 class. Given our success in attracting students back for a second summer last year, we anticipate about five of the 1997 students returning for a second summer of research, A total of 20 students will be selected for the class of 1998.

Table 1.

Undergraduate Student Interns	1997	1996
Number	18	10
Average GPA	3.32	3.73
Number of Institutions Represented	17	9
HBCU's	10	6
OMU's	0	0
Others	7	3
Sophomores	1	2
Juniors	9	7
Seniors	8	1
Males	6	1
Females	12	9
Minority	12	7
Non-Minority	6	3

All SEP students have been encouraged to attend professional conferences and present HSCaRS research results. In particular, students have been encouraged to present their HSCaRS research at the NASA University Research Centers (URC) Technical Conference and the National Conference on Undergraduate Research (NCUR). HSCaRS has committed to paying travel expenses for SEP students attending these conferences. From the class of 1997, 12 students are making presentations at URC TC '98; two other students co-authored papers but are not presenting, and one student has applied for acceptance to present at NCUR. A listing of the 1996 student research project titles is included in the appendix.

Individual profiles of the 1997 undergraduate interns are in Table 2.

Table 2. Profiles of 1997 SEP Students

College/University	Major	Research Area	Class	Gender	Ethnic Group
Lincoln University (PA)	Physics	Soil Moisture Instr.	Senior	Male	African-American
Stillman College (AL)	Computer Science	Computer Science	Junior	Female	African-American
Auburn University (AL)	Math & Physics Education	Dataset Analysis	Senior	Female	Caucasian
Susquehanna University (PA)	Physics & Env. Science	Hydrological	Junior	Female	Caucasian
University of Maryland	Cartography	GIS Analysis	Junior	Female	Asian-American
Central State University (OH)	Water Resources Mgmt.	Hydrological	Junior	Male	African
Dillard University (LA)	Computer Science	GIS Analysis	Junior	Female	African-American
Univ. of Arkansas at Pine Bluff	Math	Dataset Analysis	Soph.	Female	African-American
Jackson State University (MS)	Physics	Soil Moisture Instr.	Senior	Male	African-American
Michigan Tech	Civil & Env. Engineering	Hydrological	Senior	Female	Caucasian
Marquette University (WI)	Electrical Engineering	Soil Moisture Instr.	Senior	Male	African-American
Penn State University	Earth Science	GIS Analysis	Junior	Female	Caucasian
Michigan Tech	Environmental Engineering	Soil Data Analysis	Senior	Female	Caucasian
Salisbury State University (MD)	Math	Dataset Analysis	Senior	Female	Caucasian
Norfolk State University (VA)	Pre-Chemical Engineering	Soil Data Analysis	Junior	Male	African-American
California St. Univ. at Monterey	Earth Systems Sci. & Policy	Dataset Analysis	Senior	Female	Hispanic-American
Oakwood College (AL)	Math	Computer Science	Junior	Male	African-American
Alabama A & M University	Environmental Science	Soil Data Analysis	Senior	Female	African-American

Source: SEP Program Office Records

Program Evaluation and Student Tracking

The administration of the SEP includes a strong commitment to program evaluation. Feedback on the program is requested from each class at the end of the summer, and additional information on their career paths and continuing research contributions are requested on an annual basis. Maintaining contact with students for several years after the summer internship is necessary to do a full evaluation of the program's impact. Student tracking is expected to give us insight into the students' career paths, decisions on graduate school and continued HSCaRS research contributions by way of papers and/or presentations. To gather the data needed, a form has been designed

that is mailed to the students on an annual basis for a minimum of three years. A copy of the form is included in the appendix.

Due to the short history of the program, we have thus far done a one-year tracking survey of only the 1996 SEP class of 10 students. Of this group, six are planning to attend graduate school; two are still considering graduate school, and two are unknown. These initial results indicate that the program is attracting students with an interest in a graduate-school path and is providing positive reinforcement in that direction. Other insights into how the program has influenced the students' career paths are found in comments on the tracking form:

> "It gave me [the] confidence I needed to work toward my desired field of software engineering and designing" – Fayetteville State University student.
>
> "I narrowed down the area of study I'm planning on pursuing for my graduate thesis" – University of Maryland student now in graduate school.
>
> "I am more interested in research. I am looking for M.D./Ph.D. programs" – Norfolk State University student.

The results from the program evaluation form administered at the end of the summer to the class of 1997 indicated that this group found the internship very beneficial. Eighty-seven percent of the students said they would like to participate in the program again and that they would recommend the program to others. These students found the most beneficial aspects of the internship to be interaction with faculty mentors and other students, as well as the opportunity to learn new software/programming languages. Finally, 67°/0 of the students responded that the program had an impact on future educational or career plans. Details on the nature of these impacts will be gathered when the student tracking form is administered to the class of 1997 is the spring of 1998.

Observations and Conclusions

After two classes of students, the HSCaRS SEP is accomplishing its major objectives of effectively involving women and minority undergraduates in global change research. Some background in physics continues to be positively correlated with successful undergraduate student research contributions. Project identification as early as possible is also very beneficial. Toward this end, faculty are contacting incoming students shortly after selection to begin discussions about potential research projects. This has enabled students to begin making immediate progress on research projects after their arrival at Alabama A&M University.

It is also evident that we need to recruit more effectively in the future than we did in year two if we are to maintain the high quality of undergraduates who have participated in the SEP thus far. We anticipate that the revised recruiting strategy, with a more targeted approach toward priority pools of potential applicants will accomplish this objective. Thus far, our process of evaluating and selecting applicants, appears to be effective. While the GPA remains a very useful criterion for ranking applicants, we continue to struggle with the weight that the GPA should receive, versus other factors. Several students with lower GPA's in comparison to other members of their SEP class have made significant research contributions.

Finally, it is essential that we be able to effectively measure and document the success of the program. This is necessary to ensure that we recognize the strengths and weaknesses of the program and be able to make the appropriate modifications. Student tracking will be the key to obtaining the data needed for an in-depth program evaluation, and consequently is an area of future emphasis.

Acknowledgments

Many thanks to the HSCaRS mentor scientists, Phyllis Campbell, and Kate Hinke for assistance with this project.

References

The Task Force on Women, Minority, and the Handicapped in Science and Technology, "Changing America: The New Face of Science and Engineering", Interim Report, Washington D. C., September 1988, p. 3.

Feinstein, Selwyn, 'Sumner Internships Flourish, with Emphasis on Minorities", *The Wall Street Journal,* April 24, 1990, p. 1.

Office of Space Science and Applications in Partnership with the Educational Affairs Division, "Looking to the Future: 1991 Catalog of Educational Programs and Activities", National Aeronautics and Space Administration, Washington, DC, 1991.

Appendix

HSCaRS Summer Enrichment Program Tracking Survey

To help us evaluate the Summer Enrichment Program for Undergraduate Minority and Women Students in which you were a participant in 1996, we ask that you complete and return this form to us in the enclosed self-addressed, stamped envelope. Please feel free to add any comments, positive or negative, that could assist us in improving this program in years to come. We request your response no later than June 1, 1997. Thank you for your assistance!

Name _____
Permanent Mailing Address _____
City _____ **State** _____ **Zip** _____
Phone Number _____ **Fax Number** _____
Email Address _____

At the end of this semester or quarter, what year will you be in school? _____
When did you graduate, or when do you expect to graduate? _____
What is your current major? _____
Do you plan to attend graduate school? Yes No (Please circle one.)
If no, what are your plans? _____
_____
_____

Did your time as a HSCaRS summer intern have any influence on your current area of study, or on your plans after graduation? Yes No (Please circle one.)
If yes, how did your time here influence your plans? _____
_____
_____
_____

Do you have plans to continue in the area of research in which you were involved during your time herein 1996? Yes No (Please circle one.)
If yes, please explain how. _____
_____

Have you made any presentations or prepared any papers for publication regarding your research here since your return to your home campus? Yes No (Please circle one.)
If yes, please give details. _____
_____
_____

Other Comments: _____
_____

1997 SEP Student Research Project Titles

"Artificial Neural Computations for Soil Moisture Data Classification," Birvid D. Atkins, Stillman College, Tuscaloosa, AL, (Mentor: Dr. Chih C. Hung)

"Testing a Coherent Wave Radiative Transfer Model Using Field Measurements and a Simple Hydrology," Christine Bowman, Susquehanna University, Selinsgrove, PA (Mentor: Dr. William Crosson)

"Effect of Fresh Poultry Litter and Compost on Soil Physical and Chemical Properties," Stacy Carr, Michigan Technological University, Houghton, MI (Mentor: Dr. Teferi Tsegaye)

"Screening Extractants to Evaluate Phosphorus Availability in Some Alabama Soils," Terrence Carl-James Ferguson, Norfolk State University, Norfolk, VA (Mentor: Dr. Zachary Senwo)

"Correlation Between Precipitation and Crop Yield for Corn and Cotton Produced in Alabama," Carol Hayes, Auburn University, Auburn, AL (Mentor: Dr. Donald J. Perkey)

"Spectroscopic Evaluation of Soil Moisture and Organic Matter as a Function of Depth," Andrè Jackson, Lincoln University, Lincoln University, PA (Mentors: Drs. B. R. Reddy and W. Belisle)

"A Novel Sensor for Measuring Soil Moisture Based on Total Internal Reflection of Light," Rohan Kennedy, Marquette University, Milwaukee, WI (Mentor: Dr. Anup Sharma)

"Change Detection Analysis in Urban and Suburban Area Using Landsat Thematic Mapper Data: Case of Huntsville, AL," Dana Kuan, Univ. of Maryland at College Park, College Park, MD (Mentor: Dr. Ahmed Fahsi)

"Application of GIS Technology to Evaluate the Spatial Distribution of Soil Moisture Content Under Different Vegetative Covers," Jennifer McClain, Dillard University, New Orleans, LA (Mentors: Dr. Teferi Tsegaye and Mr. Wubishet Tadesse)

"Assessment of Atrazine Movement in a Corn Plot with Respect to the Hydrologic Responses of Decatur Soil," Elvis Niba, Central State University, Wilberforce, OH (Mentor: Dr. Narayan Rajbhandari)

"Comparison of Micronutrient Distribution in Row Crop Field Versus Pastureland," Tomeka Kenyatta Prioleau, Alabama A&M University, Normal, AL (Mentor: Dr. Andrew Manu)

"Passive Microwave Data Analysis: Advanced Microwave Precipitation Radiometer," Susan Rains, Salisbury State University, Salisbury, MD (Mentor: Dr. Steve Wu)

"Effects of Grass Vegetation on the Brightness Temperature of Soil Surface," Morgan Ruark, Michigan Technological University, Houghton, MI (Mentor: Dr. Yaping Liu)

"Correlation Between Precipitation and Crop Yields: Peanuts and Soybeans in Alabama," Theresa Soliz, California State University Monterey Bay, Seaside, CA (Mentor: Dr. Donald J. Perkey)

"Correlation Analysis of Soil Moisture and Soil Temperature Under Vegetated Plots," Stacy Steinfeld, The Pennsylvania State University, University Park, PA (Mentor: Dr. Teferi Tsegaye)

"Microwave Data Analysis at L, S, and C Bands," LaToya Suber, University of Arkansas at Pine Bluff, Pine Bluff, AR (Mentor: Dr. Steve Wu)

"The Use of Ultrasound to Measure Soil Moisture Content," Mario Thomas, 'Jackson State University, Jackson, MS (Mentors: Drs. M. D. Aggarwal, R. Metzl, W. R. Belisle and T. Coleman)

"Data Aggregation of the Numerical Hydrological Model," Thomas A. Yates, Oakwood College, Huntsville, AL (Mentor: Dr. Marius Schamshula)

SPS 366- Climate and Global Change

4 Semester Credits; Lecture -3 hours weekly; Laboratory - SEP Research Project

Introduction to climate and global change, including the relationships between the sun and the Earth that drive the climate system, the global structure and variations of the atmosphere and oceans, and the influence of humans and natural processes on the climate system and its variability. Additional topics include the greenhouse effect, ozone depletion, air pollution, acid rain, biodiversity, paleoclimatology and volcanism.

Topics include Introduction to Climate and Global Change, Atmospheric Variables and Measurements, Surface Variables and Measurements, The Earth's Radiation and Energy Budget, Remote Sensing Measurements, Atmospheric Motion and Global Circulations, Climate and Ocean Currents, The Earth's Hydrologic Cycle, Climate and Global Change, The Earth's Carbon Cycle , Interactions and System Dynamics, and Climate and Global Change. Source: Dr. Donald J. Perkey, Instructor

98 URC158

BRIGHTNESS CHANGES IN SUN-LIKE STARS

Stephen M. Henry[1] and Gregory W. Henry

Center for Automated Space Science and Center of Excellence *in* Information Systems, *Tennessee State University, Nashville, TN 37203*

Abstract. Does the Sun's energy output vary with time? Are observable climatic changes on the earth caused by changes in the Sun? Can we gain greater insight into this relationship by studying other stars with properties similar to the Sun's? In recent years, satellite observations have shown that the solar irradiance varies in phase with the 1 l-year sunspot cycle. The Sun is brighter by about O. 1% at the peak of the sunspot cycle when solar magnet ic activity is at its maximum. Over longer intervals, changes in the cart h's climate and solar magnetic activity seem to be correlated. We are using automatic photoelectric telescopes to measure brightness changes in a sample of 150 Sun-like stars. Lowell Observatory astronomers have also observed about 30 of these same stars with a manual telescope in a program that began 10 years before ours. Since these two data sets were acquired with different instruments and so have significant systematic differences, we developed software to combine them accurately and, therefore, extend our observational time coverage. We show sample results of brightness variations over 14 years in several Sun-like stars with different ages. Longitudinal studies like these, combined with cross-sectional studies of the larger sample of stars, may eventually allow us to infer with confidence the Sun's long-term brightness history and its impact on the earth's climate.

1. Introduction

Astronomers refer to the total radiant energy received per unit area from the Sun at the top of the earth's atmosphere as the *solar constant* or *solar irradiance.* This solar energy warms the earth and drives atmospheric circulation patterns that determine climate. Throughout most of the last two centuries, scientists worked to refine measurements of the solar constant. For example, Charles Abbot of the Smithsonian Institution and his associates made daily radiometer measurements of the solar constant between 1923 and 1952 from mount aintop locations in both northern and sout hem hemispheres (Abbot, Aldrich, & Hoover, 1942; Aldrich& Hoover, 1954). Abbot was convinced that he saw variations in the solar constant of up to 2% associated with solar magnetic (sunspot) activity. Further, he claimed these variations were manifest in earthly temperature and precipitation records. Abbot's findings generated considerable controversy among his contemporaries. Sterne & Dieter (1958), for example, published their own analysis of Abbot's data and concluded instead that the solar constant was *constant* to better than O. 17%; Abbot (1958) responded with a spirited defense of his belief in solar variations.

Since 1978, radiometers on earth-orbiting satellites have made extremely precise measurements of the solar irradiance from above the complicating effects of the earth's atmosphere. As a result,

[1] Also at Henry Home School, Nashville, TN

we know today that the solar irradiance *does* vary on timescales from minutes to decades but wit h amplitudes too small to see in earlier ground-based observations. In part icular, the Sun is brighter by about O. 1% at the peak of the 1 l-year sunspot cycle when solar magnetic activity is at its maximum (Willson & Hudson 1991). There also may be a long-term upward trend of approximately 0.04% in solar irradiance over the past two sunspot cycles (Willson 1997). Thus, satellite observations of the Sun's irradiance have revealed a definite relationship between the Sun's energy output and its level of magnetic activity. Our knowledge of solar brightness changes before 1978, however, is much more uncertain due to lack of precise, direct measurements.

In a landmark study published just before the onset of the space-based radiometer measurements, Eddy (1976, 1977) document ed long-term changes in solar magnetic activity over the past 5000 years. Since direct measurements of magnetic fields in the Sun have not been possible until this century, he analyzed indirect proxy indicators including historical sunspot records, the occurrence of aurorae, the ratio of carbon-14 to carbon-12 in tree rings, and the appearance of the Sun's corona during tot al eclipses. He found at least twelve extended episodes of decreased solar magnet ic activity, and each corresponded to unusually cold climatic conditions. He also found periods of enhanced magnetic activity that corresponded to warmer climatic conditions. He concluded that changes in the Sun were the dominate source of climate change over the past several millennia.

These and other more recent findings, such as the close correlation between solar-cycle length and nort hem hemisphere temperatures over the past 100 years found by Friis-Christensen & Lassen (1991), make it clear that close attention must be paid to the role of solar influences in the study of climate change on earth. While studies such as Eddy's tell us something about the long-term magnetic history oft he Sun, precise observations of solar *brightness* changes span only two decades. To improve our knowledge of the Sun's long-term brightness changes, programs to monitor brightness changes in Sun-like stars are being conducted at Lowell Observatory and Tennessee State University (TSU). Sun-like stars are stars whose mass, temperature, and composition are similar to the Sun's. Careful observations of a large sample of these stars with widely differing ages over a relatively short period of time will reveal changes the Sun is likely to exhibit over a much longer time. A parallel program to measure the magnetic variations in these same stars, the HK Project, is being carried out by Baliunas et al. (1995) at the Mount Wilson Observatory. In this paper, we briefly describe photometric observations of a sample of Sun-like stars and discuss what these observations can tell us about the Sun's long-term brightness history.

2. Photometric Observations

Lowell Observatory began their long-term program to monitor 40 Sun-like stars in 1984 with a 0.5 m reflecting telescope on Mars Hill in Flagstaff, Arizona. Details of the observing program, analysis of data, and results of the first 12 years have recently appeared in Lockwood et al. (1997). The TSU program began in 1993 with the commencement of precise brightness observations of 75 Sun-like stars with a 0.75 m automatic photoelectric telescope (APT) located at Fairborn Observatory in the mount ains of sout hem Arizona (see Henry, Figures 1 & 2, this volume). This telescope operates completely automatically without the presence of a human operator. A site-control computer, interfaced to an automated weather station, opens the observatory and signals the telescope-control computer when it is time to begin operations. The telescope-control computer is programmed over the Internet from TSU and observations are returned there automatically each morning. A second APT of 0.80 m aperture began measuring an additional sample of 75 Sun-like stars from the same location in 1996. Details of the data acquisition, reduction, and analysis can be found in Henry (1995a,b).

3. Calibration of the Lowell and APT Observations

Both observing programs measure changes in the brightness of a program star relative to a nearby constant comparison star through intermediate-bandwidth ($\sim$200Å) blue and yellow filters, and the observations are reduced to differential magnitude units. The Lowell observations are reported on their instrumental system, i.e., corrected for atmospheric extinction effects and for any slow changes in instrumental response, but not transformed to any standard photometric system. Our APT observations *are* transformed, with coefficients determined from extensive nightly observations of standard stars. Therefore, the Lowell and TSU differential magnitudes of a particular program star will differ systematically by an amount depending primarily on the color difference between the program and comparison stars.

There are 30 program/comparison star pairs in common between the Lowell and TSU programs. In order to verify the subtle brightness changes reported in Lockwood et al. (1997), to validate our own automatic observations, and to extend our light curves back in time to 1984, we developed a series of computer programs to determine the offsets required to bring the Lowell measurements into the best agreement with the standardized APT observations. This calibration process was applied separately to each individual observing season for each program/comparison star pair for which we had overlapping observations. With four years of overlap, we could determine four separate offsets for each star pair; these were then averaged to determine the best offset for each pair. The precision of the offsets was typically between 0.0005 and 0.0010 magnitudes, or 0.05% to O. 1%.

The method we used to determine the offsets for a particular program/comparison star pair depended on the characteristics of the individual light curves. If the light curves were essentially constant within each observing season, we simply slid the Lowell data with respect to the APT data until the standard deviation of the combined data was minimized. If slow light variations were seen, we slid the Lowell data until the string length connecting the chronologically-sorted observations was minimized. If more rapid (and therefore undersampled) variations were seen within an observing season, sinusoidal curve-fits were used to find the mean value for each data set, and the offset applied was simply the difference in the mean values. Figure 1 shows our results for the star HD 10476. Because the observations were essentially constant within each of the four overlapping seasons, we used our standard deviation minimization technique. HD 10476 is a solar-aged star roughly 4.3 G yr old but somewhat cooler than the Sun (spectral type K1V compared to G2V for the Sun). The yearly means of the matched data sets are shown in the bottom panel. There is a clear indication of a luminosity cycle in the star with a cycle length of roughly eleven years, very similar to the Sun's cycle length but with an amplitude several times larger. Essentially the same cycle length is seen in the star's magnetic activity (Baliunas et al. 1995).

4. Discussion

In Figure 2, we present our results for three additional Sun-like stars. The young (0.8 Gyr) G0V star HD 115383 (top panel) has rather large brightness variations ($\sim$1.4%) but no clear cycle length. The somewhat older (2. 1 Gyr) F7V star HD 120136 = τ Boo (middle panel) appears to have a luminosity cycle of around 9 or 10 years but with a smaller amplitude of about 0.6%. The 10.7 Gyr-old G0IV star HD 13421 (bottom panel) has little or no long-term brightness variability. We see the pattern of decreasing brightness variability with age in Sun-like stars that is discussed in more detail by Radick et al. (1990). This is consistent with what we understand about stellar dynamos that generate the magnetic activity in these stars. As a star ages, its rotation gradually slows over the billions of years it spends on the main sequence, and so its internal dynamo weakens.

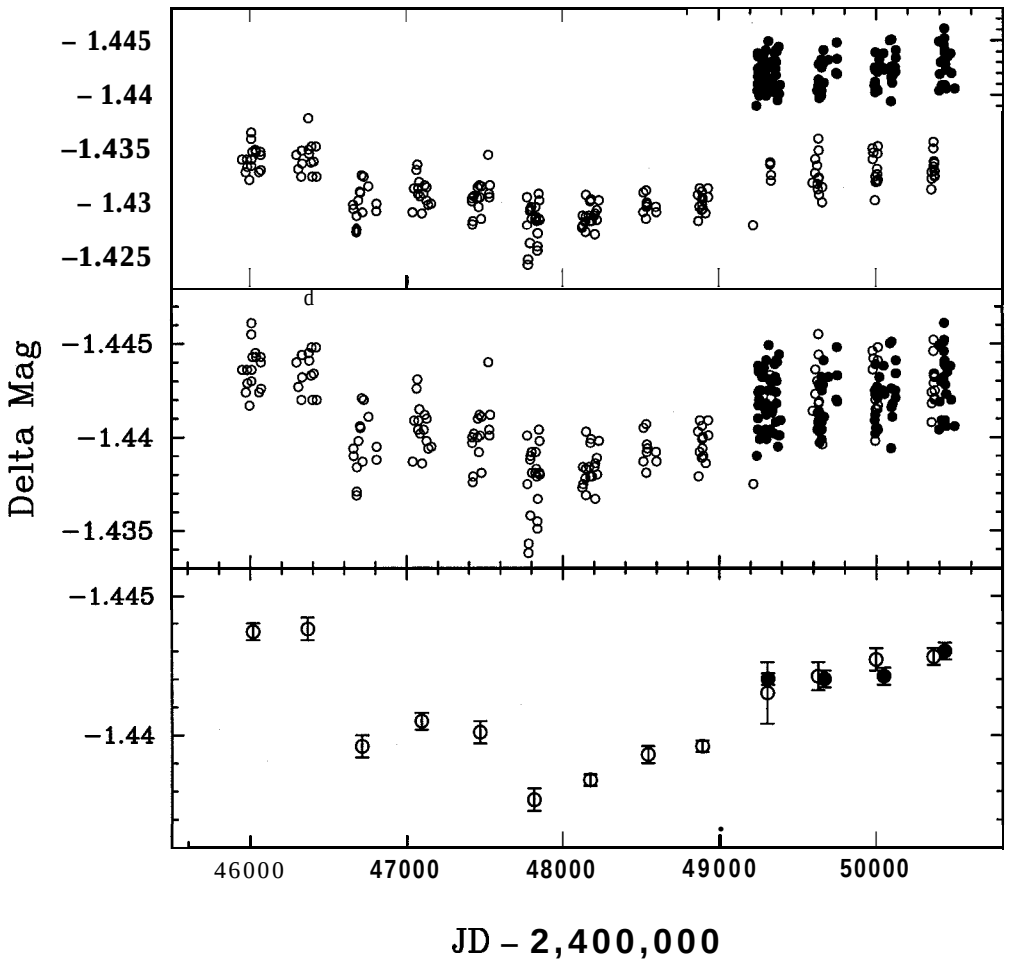


Figure 1. Results of our calibration of the 4.3 Gyr-old K1V star HD 10476. Closed circles represent observations from the 0.75 m APT. Open circles are. from Lowell Observatory. The top panel shows the two data sets before calibration. The middle panel plots the observations after the Lowell data has been offset to match the APT data. The bottom panel shows the yearly means of the matched data sets. Error bars represent the standard deviations of the means. A clear 11-year luminosity cycle is seen.

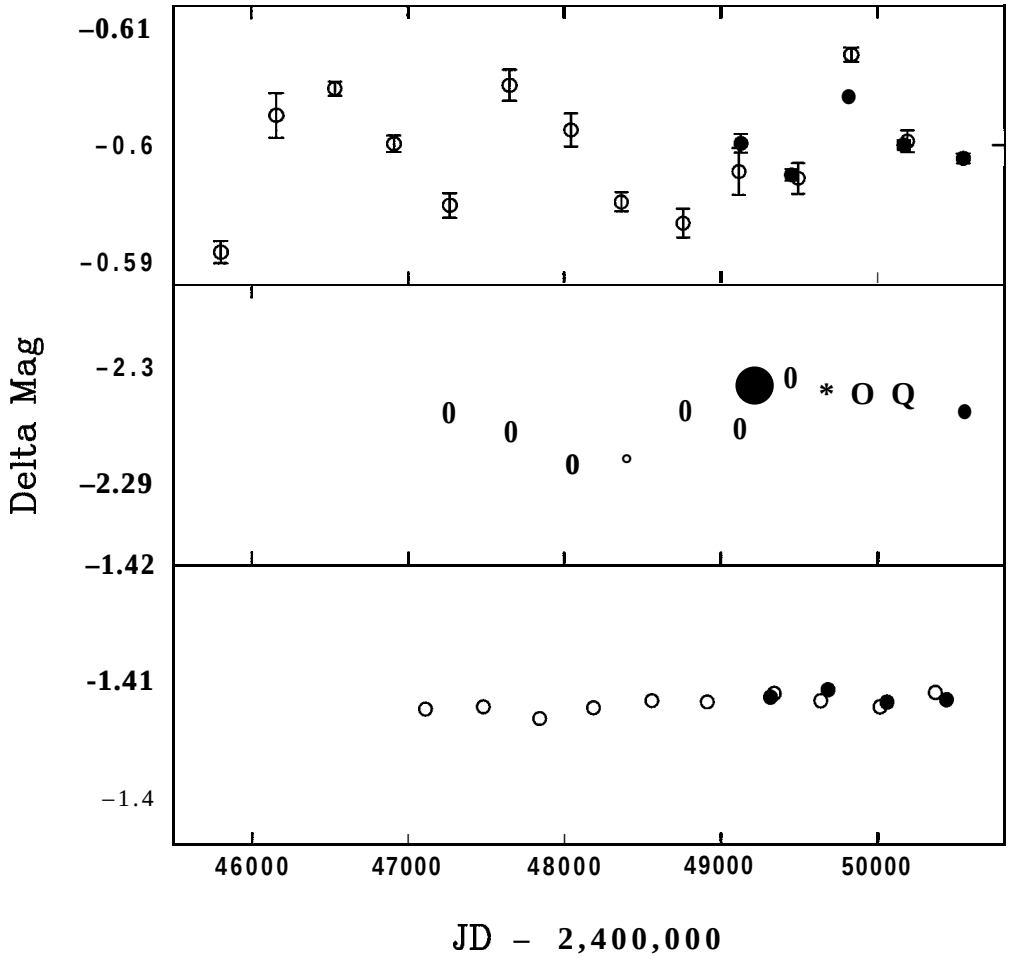


Figure 2. Luminosity variations in the young G0V star HD 115383 (top panel), the middle-aged F7V star HD 120136 (middle panel), and the old G0IV star HD 13421 (bottom panel), plotted to the same scale. Symbols and error bars are the same as in Figure 1. Error bars for the lower two panels are roughly the same as the symbol size. These observations show how long-term brightness variations decrease in amplitude as Sun-like stars age.

As a result, its overall levels of magnetic activity and brightness variability gradually decrease (see, e.g., Baliunas et al. 1997).

Observations such as these, combined with the contemporaneous measurements of magnetic act ivit y from the HK Project, have much to tell us about the past and future of the Sun and its influence on the earth. Zhang et al. (1994) have shown how these data can be used to convert the Sun's long-term magnetic activity record into a long-term brightness history. Their preliminary result suggests that the Sun brightened by 0.2% – 0.6% over the past 300 years in response to increasing solar magnet ic activity. Soon et al. (1996) used this result to infer that the terrestrial global warming observed over the past century is more likely due to an increase in solar irradiance rather than an enhanced greenhouse effect caused by the accumulation of greenhouse gases in the atmosphere.

Acknowledgments. We thank Wes Lockwood and Brian Skiff for providing the individual observations on which their recent paper was based. Astronomy with automated telescopes at Tennessee State University has been supported by the National Aeronautics and Space Administration, most recently through NASA grants NAG8-1014, NCC2-883, and NCC5-228 (which funds TSU's Center for Automated Space Science), and the National Science Foundation, most recently through NSF grant HRD-9550561 and HRD-9706268 (which funds TSU's Center for Systems Science Research).

References

Abbot, C. G. 1958, Smithsonian Contributions to Astrophysics, 3, 13

Abbot, C. G., Aldrich, L. B., & Hoover, W. H. 1942, Ann. Astrophys. Ohs. Smithsonian Inst., 6, 85

Aldrich, L. B., & Hoover, W. H. 1954, Ann. Astrophys. Ohs. Smithsonian Inst., 7, 25

Baliunas, S. L., et al. 1995, ApJ, 438, 269

Baliunas, S. L., Donahue, R. A., Soon, W., & Henry, G. W. 1997, in The Tenth Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. R. A. Donahue & J. A. Bookbinder (San Francisco: ASP), in press

Eddy, J. A. 1976, Science, 192, 1189

Eddy, J. A. 1977, Scientific American, May, 80

Friis-Christensen, E., & Lassen, K. 1991, Science, 254, 698

Henry, G. W. 1995a, in ASP Conf. Ser. 79, Robotic Telescopes: Current Capabilities, Present Developments, and Future Prospects for Automated Astronomy, ed. G. W. Henry & J. A. Eaton (San Francisco: ASP), 37

Henry, G. W. 1995b, in ASP Conf. Ser. 79, Robotic Telescopes: Current Capabilities, Present Developments, and Future Prospects for Automated Astronomy, ed. G. W. Henry & J. A. Eaton (San Francisco: ASP), 44

Lockwood, G. W., Skiff, B. A., & Radick, R. R 1997, ApJ, 485, 789

Radick, R. R., Lockwood, G. W., & Baliunas, S. L. 1990, Science, 247, 39

Soon, W. H., Posmentier, E. S., & Baliunas, S. L. 1996, ApJ, 472, 891

Sterne, T. E., & Dieter, N. 1958, Smithsonian Contributions to Astrophysics, 3, 9

Willson, R. C. 1997, Science, 277, 1963

Willson, R. C., & Hudson, H. S. 1991, Nature, 351, 42

Zhang, Q., et al. 1994, ApJ, 427, L111

MECHANISM OF NITRIC ACID DETECTION IN A POLYMER-COATED QUARTZ CRYSTAL MICROBALANCE.

Kahsay Habte and S. K. Pollack
Department of Chemistry
The Center for the Study of Terrestrial and extraterrestrial Atmospheres
Howard University
Washington, DC 20059.

Coated piezoelectric quartz crystal microbalance (QCM) have been studied as potential selective and sensitive detectors for a wide range of toxic vapors. There is a current need to develop sensitive and chemically specific sensors for the detection of nitric acid for in-situ measurements in the atmosphere. The resonance frequency of the quartz crystal is determined by physica properties such as density, thickness, and shear modulus and is therefore intrinsically very stable. However, if any of its physical parameters change, the resonance frequency of the resonator will also change and a trace amount in picogram range can be detected. In order to detect trace amounts of nitric acid in the polymer-coated quartz crystal, we developed a home made QCM apparatus. The QCM apparatus, showed promising results in changing frequency when a certain mass was loaded on to 10MHz AT-Cut quartz crystal, but failed to maintain the stability and dramatic frequency jump was observed. To fulfill this requirement, a QCM apparatus (EQCN -402) with a Faraday Cage built in it was employed to avoid any noise or dramatic second or third harmonic oscillation. This apparatus is capable of measuring any picogram mass change on the Quartz Crystal and also can be used in liquid phase. Polymer coatings such as polystyrene, showed promising results in detecting nitric acid gas selectively. In order to prove the selectivity of polystyrene, we introduced gases such as hydrochloric acid and sulfur dioxide and to date no interference has been observed. Contact angle and ellipsometry are some of the methods currently used to detect any changes in surface chemistry of the films.

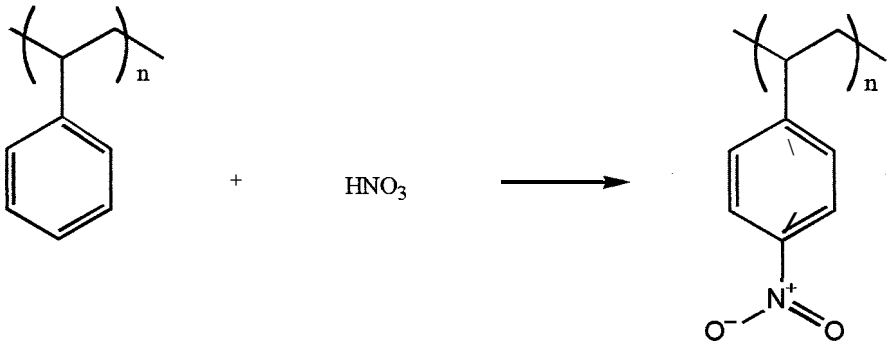
One of the most important issues in stratospheric aviation are to understand the chemical and physical process in the atmosphere. Nitric acid produced by high temperature combustion is a reservoir species, and indirectly destroys the ozone layer by reactin gwith various radicals which are responsible for photolysis of ozone layer in the stratosphere. Measurement using remote sensing is more accurate and expensive compared to in-situ measurement of nitric acid, but it requires that the molecule of interest should have to have a molar absorption and unique spectroscopic characteristics to avoid any interference's from other atmospheric species.

The phenomenon of piezoelectricity was first observed in 1880 by Pierre and Jacques Curie in a number of substances including natural quartz (l); they discovered that

a pressure exerted on small piece of quartz caused an electrical potential between deformed surfaces. In 1959, Sauerbrey (2,3) developed a relationship between the weight of metal films deposited on quartz crystal and the change in frequency. The quartz-crystal microbalance (QCM) is rapidly becoming one of the most important and widely applicable analytical techniques. Amongst the well-known sensors, the quartz crystal microbalance, which is compromised of a thin vibrating AT-cut quartz wafer sandwiched between two metal excitation electrodes, has received increasing attention because of its potential applications in many fields and the simplicity of its structure.

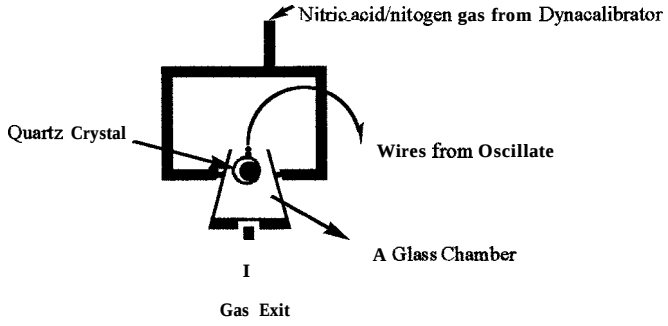
Recently in our group, there has been some progress in using polymers as coatings of QCMs for detecting nitric acid for in-situ measurements. Several polymer coatings had been used for detection of hydrochloric acid, sulfur dioxide, nitric acid and other trace gases. When a film coating of the copolymer of 3,4-dimethoxystyrene with styrene (PS/PDMS) was exposed to nitric acid, a frequency decrease was observed indicating mass loading. The dimetoxystyrene unit was incorporated as a reactive group for aromatic nitration. This film gave no response when exposed to ozone, sulfur dioxide and hydrogen chloride vapor. However, when polystyrene (PS) was exposed to nitric acid gas it too gave a positive response. This indicted that the more costly material is not necessary. Polystyrene alone was utilized as potential polymer coatings for the detection of nitric acid gas. It is assumed that the polystyrene reacts with nitric acid to give nitro polystyrene in a simple nitration mechanism, where nitronium acts as electrophile and the phenyl group (electron rich) acts as nucleophile. The reaction of polystyrene with nitric acid is shown in figure 1.

Figure 1. Reaction of HNO_3 with Polystyrene



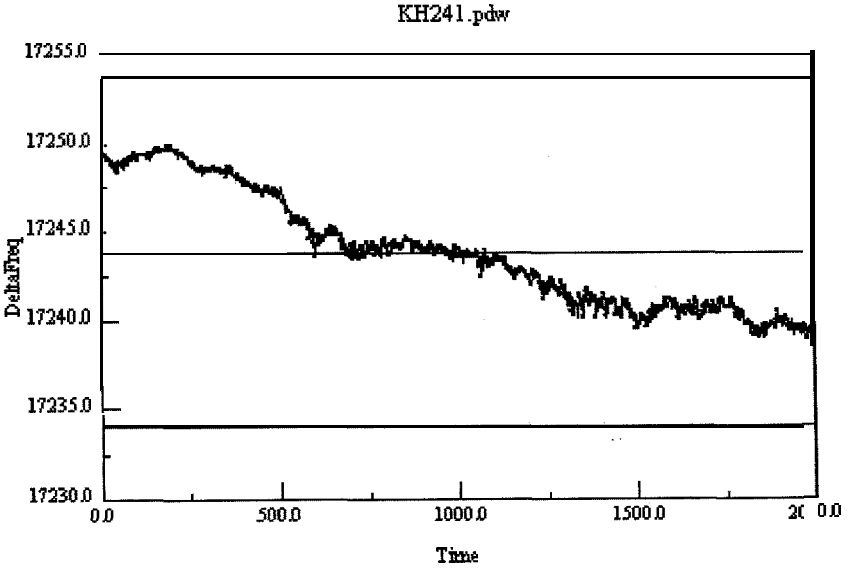
The polystyrene (granulated) was purchased from Aldrich and used as supplied. The polymer (polystyrene) is dissolved in toluene for coating. The AT-cut crystal, which is typically cut an angle of +35° 15' and has a zero frequency/temperature coefficient at or near room temperature, was dipped in toluene/polystyrene solution and dried in oven 5-10 min before it was placed in the chamber for detection of nitric acid gas or other gases. Nitric acid gas is generated in Dynacalibrator (VICI Metronics). After the nitric acid in permeation tube was heated and the flow was set to the desired concentration and temperature, nitric acid gas flows into the chamber where the AT-cut quartz crystals are placed In order to increase the sensitivity, a small 50 mL flow chamber was designed to hold the crystals.

Figure 2.



Using the oscillator (EQCN 402), control program (Microsoft Visual Basic), the frequency counter (HP 53131) and the computer (PC), any mass deposited will change the frequency of the crystal and the data can be saved for later analysis. Frequency change versus time can be graphed and a decreasing graph will be observed if mass deposits on the crystal.

Figure 3. A sample file (KH241.pdw). Mass loading when nitric acid (250ppb) gas was allowed to flow into the chamber for 2000 seconds.



Polymers can be applied as thin films on the quartz crystal, but they are not very selective. To create selectivity, we can attach different functional groups to the polymeric chain. This can interfere with the polymerization process and low molecular weight polymers may result. To avoid this, we are trying to use self-assembled monolayer (SAMs), in which the gold cathodes of crystals will be coated with the thiols. The thiol

groups tightly bind to the gold and the hydrocarbon species will form self assembled monolayer. Once the thiol group is tightly bound to the gold crystal, various functional groups can be attached (4). Recently, self assembled monolayer were used as coating material because of their structure (5). Smooth reaction surface can be prepared using self assembled monolayer, which is very useful in increasing the sensitivity of the crystal. These surfaces also allow for subsequent spectroscopic analysis of the reacted crystals.

Phenylhexadecyl thiol will be synthesized in a multi step synthesis and will be tried as a sensor for nitric acid. We will also explore the use of other polymer or copolymer coatings for nitric acid detection, study the kinetics of nitric acid absorption onto quartz crystal microbalance, and spectroscopic changes to the coatings.

REFERENCES

(1) F. W. Maarsen, M. C. Smit, and J. Matze, *Rec. Trav. Chim. Pays-Bas, 76,713 (1957)*

(2) R. A. Heising, "Quartz Crystal for Electrical Circuts", Van Nosrtand, New York, N. Y.> 1946, p 24

(3) G. Z. Sauerbrey, Z. Phys., 155,206 (1959)

(4) Bain, C.D.; Evall, J.; Whitesides, G.M. *J. Am. Chem Sot. 1989,111,7155-7164*

(5) Nuzzo, R. G.; Dubois, L. H.; Allara, D. L. *J. Am. Chem.* Sot. 112,558 (1990)

Meteorological and Watershed Factors Affect the Runoff and Peak Flow Reduction Capacities of Forested Filter Zones

Narayan B. Rajbhanadri , Alabama A&M University, Department of Plant, Soil, and Animal Science
P. O.Box 1208, Normal, AL-35762
and
James Gregory, North Carolina State University, Department of Forestry
P. O.Box 8008, Raleigh, NC-27695-8008

INTRODUCTION

Runoff is the key hydrologic process affecting the stability of land in the riparian watershed system. Maintenance of a forested filter zone (FFZ) is recognized as an effective way of reducing the velocity of surface runoff from an upland area (Barker and Gregory, 1989, and Shultz, 1995). Because the sediment carrying capacity of surface runoff is proportional to the fifth power of velocity, a small decrease in flow velocity can cause a significant increase in sediment deposition (Alberts et al., 1981). Rajbhandari et al., (1996) reported that sediment deposition was significantly increased where drainage density of small ephemeral channels in a FFZ was low. They further reported that the deposition was remarkable when the infiltration of field runoff in the FFZ was increased by dispersing flow across contour.

Dispersal of surface runoff utilizes all the hydrologic attributes of a filter zone or riparian zone. In a FFZ, litter fall increases hydraulic roughness, the deposition processes, and infiltration capacity. Forest debris along the slope contours significantly reduces flow velocity and enhances water detention storage. In addition, dense ground cover and understory vegetation increase turbulent flow to substantially reduce flow velocity and enhance infiltration and detention storage. Extensive root development of trees enhances soil structure and porosity for deep percolation of water. Nevertheless, these attributes change with seasons. In this paper, we describe differences in total surface stormflow (including A horizon interflow) and peak flow rate across FFZ's adjacent to agricultural fields with regards to meteorological, runoff, and watershed factors. The meteorological factors tested were the two seasons - growing season and dormant season. Comparing two modes of runoff, dispersed flow and concentrated flow, were the hydrologic factors. The watershed factors included drainage density, vegetation cover, and soil moisture content.

MATERIALS AND METHODS

We conducted this study on the Oxford Tobacco Research Station, located in Granville County, North Carolina on two small agricultural watersheds. Each watershed consisted of several small fields from which runoff drained via grass-vegetated waterways into an adjacent forest (Figure 1). The two watersheds were nearly equal in size but differed in the ratio of field area to adjoining forested filter area, topography, and drainage density (ratio of the total hydraulic length of channels to the total area of the FFZ (Table 1).

Soils in the fields of WS-1 were Vance sandy loam (Typic hapludult, clayey, mixed, thermic) and Durham sandy loam (Typic hapludult, fine loamy, siliceous, thermic). In FFZ-1, the soil consisted of Helena loamy sand (Aquic hapludult, clayey, mixed, thermic). Soils in the fields of WS-2 were Helena loamy sand and Vance sandy loam and in FFZ-2 were Vance sandy loam with some Durham sandy loam inclusion.

The agricultural fields of the watersheds were managed by the research station in a rotation of tobacco (*Nicotina tabacum L.),* soybean (*Glycine max L),* and wheat *(Triticum aestivum L.)* experiments. The forest stands of the FFZs of both watersheds consisted of a natural stand of loblolly pine *(Pinus taeda L.)* mixed with several hardwood species.

After surface storm flow drained from the fields through the grass-vegetated waterways, it was channeled into the FFZ by berms constructed on either side of the natural topographic drainage. The berms contained plywood cutoff walls that extended through the A horizon. As the runoff entered a FFZ, it passed through a 61 cm H-flume (upper flume) for flow measurement. Runoff treatments consisted of: (1) allowing concentrated flow in the natural ephemeral channel or (2) dispersing the flow on contour at the up-slope edge of the FFZ through a level spreader. Rainfall events were alternated between concentrated and dispersed flow.

Table 1. Physical Dimensions of the two watersheds at the Oxford Tobacco Research Station, Granville County, NC.

Dimension		Watersheds WS-1	WS-2
1.	Field area above upper flume (hectare)	1.34	1.22
2.	Forested area of watershed (hectare)	0.16	0.25
3.	Forested filter zone area (hectare)	0.12	0.21
4.	Total area of watershed (hectare)	1.50	1.47
5.	Ratio of field to forested filter zone	11.17	5.80
6.	Average slope of field (percentage)	5.00	6,00
7.	Average slope of the filter zone (percentage)	4.00	9.00
8.	Average slope of the grass path way (percentage)	4.00	4.00
9.	Maximum length of filter zone (meter)	30.50	39.60
10.	Drainage density (meter per hectare)	5820.00	2120.00
11.	Maximum width of filter zone (meter)	54.90	67.10
12.	Average width of grass pathway (meter)	17.00	7.00

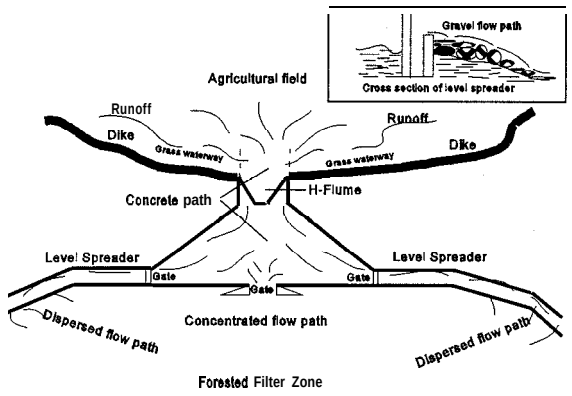


Figure 1. Pattern of flow path from field to forested filter zone (not to scale).

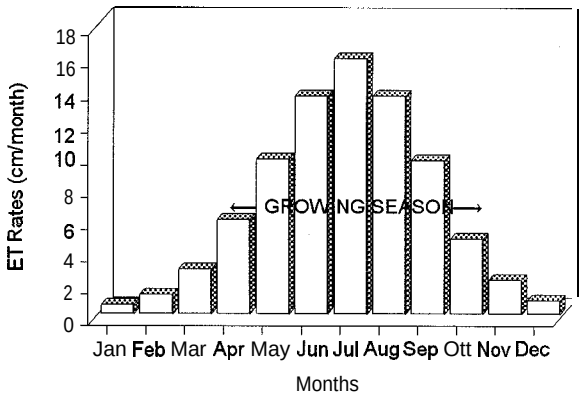


Figure 2. Monthly evapotranspiration rate and growing season

The level spreaders were made of wooden troughs extended on contour across the FFZs. A distribution box connecting the spreaders was constructed over the natural ephemeral channel in each watershed. The runoff was concentrated in the natural channel by discharging it through the central outlet of the distribution box. Dispersed flow was produced by discharging the runoff through the entrance doors of the level spreaders on either sides of the distribution box.

A plywood wing wall (extended through the A horizon) across the lower end of each FFZ directed the runoff into a second H-flume (lower flume) centered on the natural drainage-way at the down-slope edge of the FFZ. At the lower flume, the flow was also measured. Runoff measured at both upper and lower flumes consisted of overland flow and A-horizon interflow. The data discussed here covers a period of 27 months, January, 1992 through March, 1994.

Mean annual rainfall at the research site was about 1121 mm and mean annual temperature was about 150 C. Monthly total potential evapotranspiration (PET) was estimated by using Thornthwaite's method (Thornthwaite, 1948). The monthly PET was highest during the months of May - September when it ranged from nearly 10 cm/month to 17 cm/month (Figure 2). During the remaining portions of the year, the monthly PET ranged from nearly 1 cm/month to 6 cm/month.

The growing season was identified with regard to monthly potential evapotranspiration rates of the two watersheds. The monthly PET rates started to gradually increase about mid-March and gradually decreased about mid-November. The period between these two months (April - October) was considered as the growing period. The remaining period (November - March) was considered as the dormant period.

Stormflow events entering and leaving the FFZ's were classified for analysis by size, dormant or growing seasons, and concentrated or dispersed flow. The effectiveness of flow dispersal in reducing the volume and peak rate of stormflow leaving the FFZ's was evaluated by comparing the relative percentage of runoff reduction during each mode of dispersion. A similar evaluation procedure was performed for the peak flow rates. Runoff reduction refers to the ratio of flow leaving the FFZ at the lower flume compared to that entering the FFZ at the upper flume.

RESULTS AND DISCUSSION

Flow dispersal was very effective in enhancing the reduction of total storm runoff in the FFZ's and had a greater effect in FFZ 2 than in FFZ 1 (Table 2). The ratio of FFZ area to watershed area was lower and storm runoff and peak flow from the fields were lower in FFZ 2 than in FFZ 1. Runoff reduction was nil for concentrated flow in FFZ 1 and was 18% for dispersed flow. In FFZ 2, runoff reduction was 30 % for concentrated flow and 62 % for dispersed flow.

There was also a seasonal difference in the impact of flow dispersal in reducing total storm runoff across the FFZ's and that seasonal effect was more pronounced in FFZ 1 than in FFZ 2 (Figure 3A). In FFZ 1, there was little effect of dispersal in the dormant season when high water table conditions resulted in more total surface runoff (some of which was exfiltrated base flow) discharging from the FFZ than entered it. However, during the growing season runoff reduction was 27 % for dispersed flow but only 4 % for concentrated flow. Runoff reduction in

Table 2. Effect of flow dispersal on percent runoff reduction in the two FFZ's.

FFZ	Concentrated Flow	Dispersed Flow
1	0	18

FFZ 2 was much higher than in FFZ 1 for both concentrated and dispersed flow in both seasons (Figure 3A). However there was still a significant season interaction with flow dispersal. Average runoff reduction was 33 % for concentrated flow in the dormant season and flow dispersal increased the runoff reduction percentage by a factor of 1.8 to 58 %. In the growing season, average runoff reduction was 29 % for concentrated flow and was increased by a factor of 2.3 to 67 % by flow dispersal.

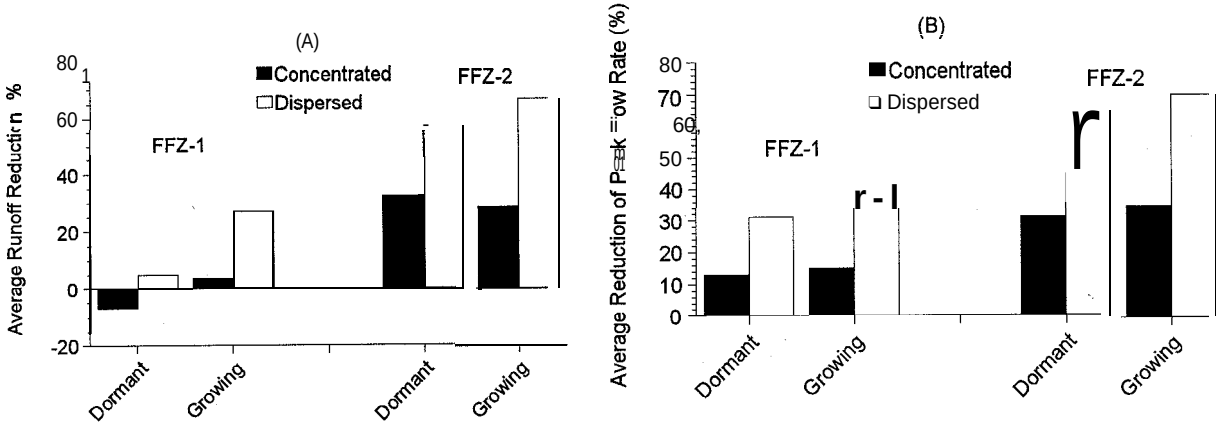


Figure 3. Average percentage of reduction of(A) runoff and (B) peak flow rate at the lower flumes of forested filter zone.

These results indicate some combination of lower storage capacity, infiltration capacity, and deep interflow capacity in the soils of the FFZ's during the dormant season when soil moisture was higher than during the growing

season. Schwer and Clausen (1 989) reported a similar seasonal effect upon a vegetated buffer strip in a dairy milk-house waste water study.

Seasonal differences in drainage density and surface roughness manifested significant effects on dispersed flow. During the dormant period when ground vegetation was sparse, average Manning's roughness coefficient (n) was estimated at 0.07 for main ephemeral channels and 0.02 for side channels (Rueda, 1993). The roughness coefficient (n) increased to 0.15 for main channels and 0.1 for side channels during the growing season which was attributable to the dense ground cover. The increase in n means a decrease in runoff discharge rate and greater opportunity for infiltration, storage, and deep interflow and vise versa.

The hydrography of two dormant season dispersed flow events (Julian day 5, 1993) and the hydrography of two dormant season concentrated flow events (Julian day 86, 1993) are compared in Figure 4. These events show the slight impact of FFZ-2 on flow reduction during concentrated flow and the large impact during dispersed flow, with soil moisture content at field capacity,

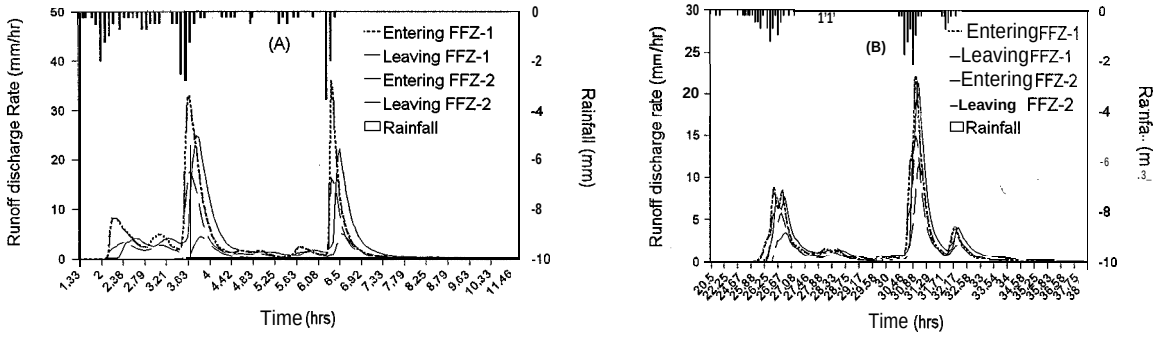


Figure 4. (A). Two dispersed flow hydrography for storm events which occurred on Julian day 5, 1993. Antecedent soil moisture condition was field capacity. (B). Two concentrated flow storm events that occurred on Julian day 86, 1993. Antecedent moisture content was field capacity.

Further, a large dispersed flow event during the growing season (Julian day 249, 1992) demonstrated substantial amounts of infiltration in FFZ-2 even when the surface soil moisture content was greater than field capacity (Figure 5A). This event was preceded by a large concentrated flow event on Julian day 248, 1992 (Figure 5B) when the moisture status of the surface soil was near field capacity. However, due to dense vegetation cover on the surface soil, the peak flow rate from the FFZ was lower than that from the field. In addition, due to high ET rates and a relatively long dry period (below normal rainfall), the water table remained low and the moisture content of the B and C soil horizons remained low. As a result, there was still ample infiltration, storage, and deep interflow capacity to attenuate runoff from the FFZ's.

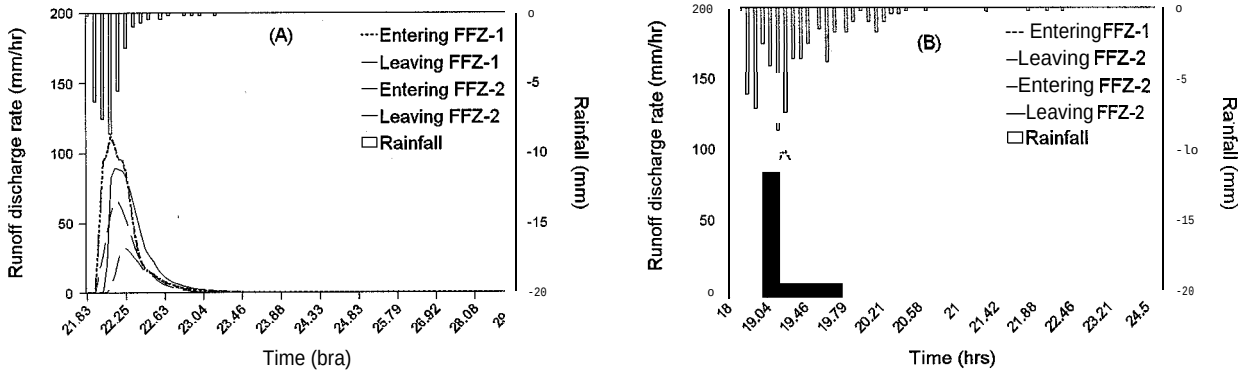


Figure 5. (A) The large dispersed flow eventoccurred on Julian day 249, 1992. Antecedent soil moisture in the FFZ was greater than field capacity. (B) The large concentrated flow event occurred on Julian day 248, 1992. Antecedent soil moisture was near filed capacity.

Since Manning's n was higher for the channels and the water table remained quite low during the growing season, flow dispersal was expected to reduce peak flow rates more efficiently during the growing season. Comparing concentrated flow by season shows that there was slightly more peak flow reduction through the FFZ's in the growing season than in the dormant season (Figure 3B). The percent peak flow reduction in FFZ-2 was nearly double that for FFZ- 1 in both seasons. However, the relative effective of dispersal was greater in the growing season than in the dormant season, particularly in FFZ-1 where the peak flows were higher than in FFZ-2. In FFZ-1, dispersal in the dormant season improved peak flow reduction compared to concentrated flow by a factor of 2.38 (from 13°/0 to 3 1°/0) and in the growing season by a factor of 2.80 (from 15°/0 to 42%). In FFZ-2, dispersal in the dormant and growing seasons improved peak flow reduction by factors of 2.10 and 2.28 respectively.

CONCLUSIONS

Attenuation of surface stormflow volume across FFZ's was greatly enhanced by dispersal of agricultural stormflow across contour in the FFZ's. Runoff reduction was somewhat higher in the growing season than in the dormant season. The most important factors in runoff reduction across the FFZ's were maximizing the ratio of FFZ area to watershed area and minimizing drainage density of surface rivulets in the FFZ. Surface stormflow attenuation in the FFZ's was attributed to a combination of detention storage and relatively deep interflow, both of which would be expected to enhance water quality amelioration.

REFERENCES

Alberts, E. E., W. H. Neibling, and W. C. Moldenhauer. 1981. Transport of sediment, nitrogen, and phosphorus on runoff through corn stalk residue strips. J. Soil Science Society of America, 46 (6): 1177-1184.

Barker, S. E. and J. D. Gregory. 1983. Effect of stream side buffer zone configuration on sedimentation. Small Woodlot Forestry R&D Program, North Carolina State University, Raleigh, Research note serial no. 17.

Rajbhandari, N.B., J.D. Gregory, C. Franklin, and J.E. Parsons. 1996. Dispersal of agricultural runoff through a level spreader in a forested filter zone enhances sediment and nutrient removal. Proceedings, Water Quality Conference. March 19-20, 1996, North Carolina State University, Raleigh, USA.

Rueda, J.P. 1993. Evaluation and application of ANSFOR model in riparian buffer areas. MS Thesis. NC State University, Raleigh NC.

Schultz, R. C., T.M. Isenhart, and J. P. Colletti. 1995. Riparian buffer systems in crops and rangelands. In Agroforestry and sustainable systems: Symposium proceeding, ed. W. J. Rietveld, 13-27. USDA For. Serv. Gen. Tech. Rep. RM_GTR-261.

Thornthwaite, C. W. 1948. An approach toward a rational classification of climate. The Geographical Review, Vol. 38, No. 1. Pp 55-94.

Schwer, C. B. and J. C. Clausen. 1982. Vegetative filter treatment of dairy milk house wastewater. J. Environmentalal Quality, 18:446-451.

ACKNOWLEDGMENT

This research work was supported by the Woodlot Forestry Program, Forestry Department, North Carolina State University, Raleigh.

Assesment of Atrazine Movement in a Corn Plot with Respect to the Hydrologic Responses of Decatur Soil

Elvis Nibs
Central State University
International Center for Water resources Management
Box 5, Wilberforce, OH 45384
and
Narayan Rajbhandari
Center for Hydrology, Soil Climatology, and Remote Sensing
Plant and Soil Science department
Alabama A&M University
Normal, AL 35806

Introduction

Application of Pesticides is common in agricultural practices because of their ability to control disease vectors and pests that adversely affect the production and quality of food. One harmful factor of such a practice has been the severe environmental stress to many river basins. Some recent studies have shown that subsurface flow can carry detectable levels of pesticides to contaminate river basin and ground water (Buhler et al., 1994).

Atrazine [*2-chloro-4ethyl amino-6 isoproilamino-5-triazine*] *is* one of the widely used pesticide for corn (*Zea mays L.*) *culture* and is one of the most commonly detected pesticides in stream water and ground water in the USA (USDA, 1990). Rajbhandari (1 996) identified that Atrazine was sensitive to most of the soil types in the coastal plain during corn planting period in Norh Carolina. In this paper, we try to assess Atrazine movement in the corn field at the upland region in Alabama, USA, with respect to some hydrologic parameters and to find out if substantial amounts can contaminate ground water.

Methodology

In the first week of June, 1997, we established a 100 m^2 study plot, at Alabama A&M University's Winfred Thomas Agricultural Research Station (WTARS) located in Hazel Green, Alabama. The plot was planted with corn in a disk bed on April 26, 1997. The Atrazine was applied to the bed at the rate of 2.80 kg/ha prior to the plantation.

Surface flow and sub-surface flow from this plot drain down to a nearby irrigation pond. The slope of the edge between the plot and the pond is about 3°/0. The soil types as classified by USDA for the research site is upland Decatur soil. The upper soil profile (1-1 8 cm) contains 4 1.7% sand and 20. 9% clay and the lower soil profile (19-103 cm) contains about 47% clay and less than 20% sand.

Soil Sample Analysis

In the second week of June, 1997, soil samples were collected using a soil probe at the upper soil profile at 20 m intervals at the lower edge of the corn plot to estimate the amount of Atrazine leaching down to the ground water and to the nearby pond. The soil samples collected at the five sites of the plot were labeled A_1 to A_5 for the first 10 cm soil depth and B_1 to B_5 for the remaining soil depth. The soil samples were air-dried for 24 hours and ground using a mortal and a pestle. They were extracted by mixing 25 grams of soil in 100-ml of methanol and water (4:1 ratio) in a 250-ml erlenmeyer flask. The flask was closed with cellophane and was shaken with a wrist action shaker for 6 hours. The solution was centrifuged at 100 revolutions per minute to obtain a clear supernatant. This process lasted for about 60 seconds. The collected supernatant was then transferred to vials in preparation for concentration. The solution was concentrated using a constant temperature water bath of 50 °C for approximately 30 hours.

The samples were concentrated to 15-ml. The pesticides in the concentrated solution were extracted using Solid Phase Extraction. This was done by injecting 10-ml of the concentrated sample through the C_{18} column, using a vacuum box. The pesticides were removed from the column by passing 2-ml of ethyl acetate through the column. One ml of the eluded pesticides were taken to the gas chromatography to determine the amount of Atrazine found in the samples.

Water sample analysis

Water samples were collected from five different locations along the edge between the corn plot and the pond, on June 13, 1997 (before rainfall) and June 18, 1997 (after rainfall). The June 13 samples were combined and labeled (D_1) and the June 18 samples were also combined and labeled (C_1). The water samples were collected in vials and immediately taken to the lab and stored in the refrigerator for 2 days. Ten ml of water were transferred into a 250-ml erlenmeyer flask. The soluble pesticides were extracted using solid phase extraction. The water samples were passed through the C_{18} column. The samples were extracted through the column using a vacuum box. Like in the soil sample analysis process, the pesticides were removed from the column by passing 2-ml of ethyl acetate. One ml of the extracted pesticide was taken to the gas chromatography to determine the amount of Atrazine found in the samples.

A model run

The Erosion Productivity Impact Calculator (EPIC) model was used to estimate the average amount of Atrazine in the pond water and the ground water for each month. Information regarding farm practices such as dates and amounts of pesticide application, soil tillage practices, types of crops planted, harvesting procedure and harvesting time were obtained by interviewing the WTARS farm manager. Additional information on the hydrologic parameters such as temperature, precipitation, relative humidity, and varied velocity were obtained from the Huntsville Weather Station. Information on the pesticide parameters of the cornfield was obtained from USDA data files. The EPIC model was ran for five years using the weather generator component of the model to estimate average amounts of Atrazine carried through sub-surface flow and water percolation through the root zone. Figure 1 demonstrates the average monthly temperature and evapotranspiration for the study plot. Figure 2 demonstrates the monthly rainfall, percolation and sub-surface flow.

Results and discussion

The Atrazine level was considerably higher at the upper soil depth than at the lower soil depth (Figure 3). The results indicate that a substantial amount of Atrazine was carried by sub-surface flow down to nearby irrigation pond. This fact is further supported by our water sample analysis. The water samples collected on June 13 (before rainfall) showed that the concentration level of Atrazine in the pond was 0.0138 ppm (Figure 4). Interestingly, the results from the samples collected on June 18, after 5 days of consecutive rainfall, showed that the concentration level was increased to 0.158 ppm. Since, the upper layer of the soil profile contained relatively high sand percentage and the lower layer of the soil profile contained relatively high clay percentage, infiltration was expected at the upper portion and sub-surface flow was expected at the lower portion. Consequently, the rainfall event on June 18 accelerated sub-surface flow and, thereby, increased the concentration of Atrazine level in the pond. However, the Atrazine level was not high enough to contaminate the ground water.

The EPIC model suggests that the pond would receive Atrazine at the rate of about 1.17 gin/ha during June (Figure 5). The transportation of Atrazine to the pond was considerably high during May. This is due to the fact that pesticides were applied in late April. Therefore, during this time of the year, the slightest amount of precipitation would initiate leaching. Similarly, higher rate of ground water leaching was observed during May (Figure 5).

We infer that there was a direct relationship between evapotranspiration and temperature (Figure 1). As the temperature increased, evapotranspiration also increased. This relationship was observed between May and

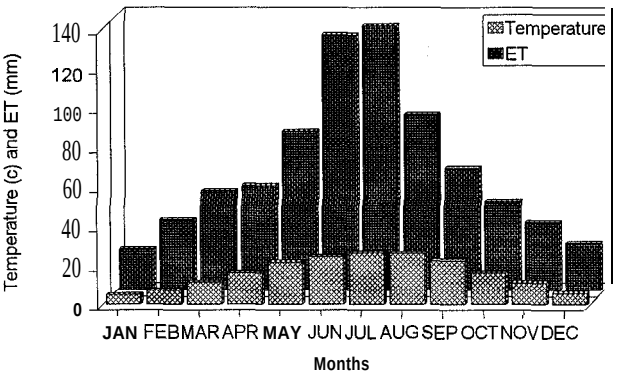


Figure 1. Monthly air temperature and evapotranspiration (ET) rate for the Winfred Thomas Agricultural Research Station, Alabama A&M University, Alabama. (EPIC output).

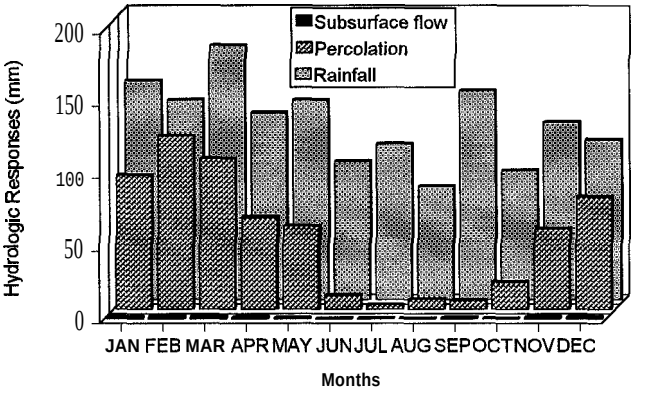


Figure 2. Monthly distribution of *rainfall* and its impact upon the *amount of water* percolation from root zone and subsurface flow from the corn plot, (EPIC output)..

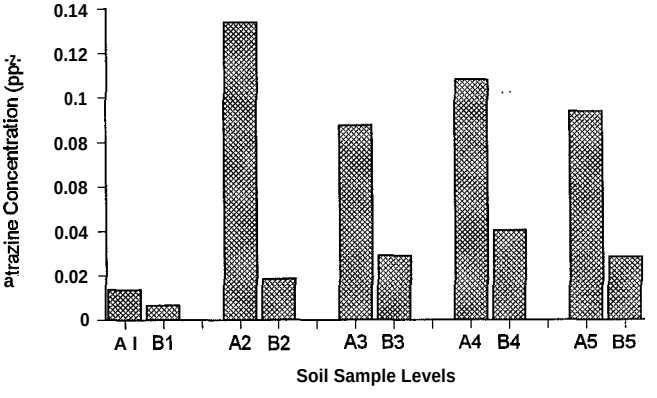


Figure 3. Atrazine concentration in the soil samples collected at the lower edge of the corn plot. The soil samples collected at the five sites of the plot were labeled A_1 to A_5 for the first 10 cm soil depth and B_1 to B_5 for the remaining soil depth.

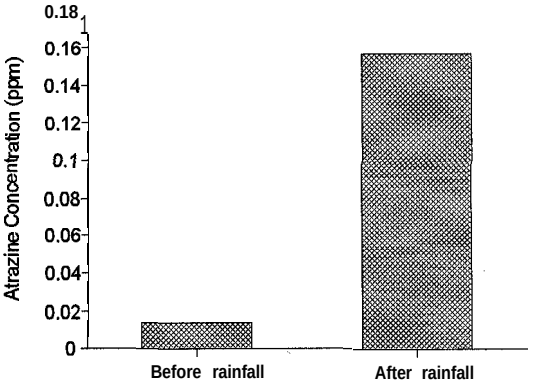


Figure 4. Atrazine concentration in the water samples collected along the edge between the corn plot and the pond before (June 13, 1997) and after (June 18, 1997) rainfall events.

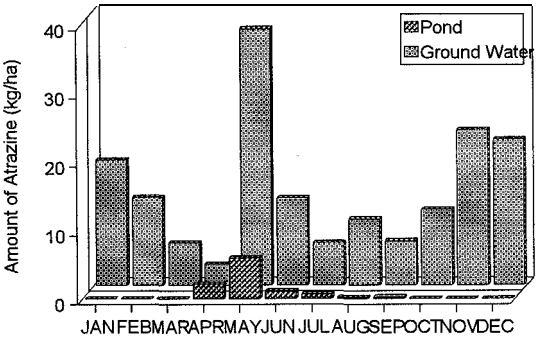


Figure 5. Monthly amount of Atrazine leaching to the irrigation pond and to the ground water as predicted by EPIC.

August. This explains why percolation was lowest during this time of the year despite the fact that precipitation was relatively high. Therefore, the rate of Atrazine leaching to ground water was relatively low during these months except during May. For the rest of the periods, the percolation from the root zone remained high for the high rainfall events (Figure 2). As a result, the amount of Atrazine leaching to ground water was also accordingly increased (Figure 5).

Conclusions

Despite the importance of Atrazine in pest control and improvement in food quality, it is possible that sub-surface flow can carry substantial amounts of Atrazine that contaminate freshwater resources such as streams, rivers, lakes, and estuaries. The amount of Atrazine was remarkably high after its application in the study field for a few months. Also, the increase in the amount of Atrazine found in the pond after five consecutive days of rainfall is consistent with other findings which show a direct relationship between precipitation and pesticide leaching. As a result, ground water contamination was considerably high when percolation from the root zone was increased. However, the amount of Atrazine found in the pond water samples was not significant enough to contaminate

ground water. Due to time constrains detailed study of the area was not done. Therefore, further analysis is required for more reliable results.

Acknowledgments

We would like to thank all the faculty and staff members of the department of Plant and Soil Science, Alabama A&M university, for their efforts in making this research possible. We would like to extend out thanks to Mr. Null for providing the corn field for this study. Special thanks is extended to Belvitt for his support and assistance in pesticide analysis. Finally, we would like to extend our acknowledgments to all the graduate students for their help and support.

References

BUHER, D.D., RANDALL, G. W., KOSKINEN, W.C. 1994. Atrazine and Alachlor losses from sub-surface Tile Drainage of a clay loam soil. J. Environ. Qual., 22:583-588.

GAMERDINGER, A. P., WAGNET, R. J. and VAN GENUCHEN. 1990. Application of Two-site/Two-Region models for studying simultaneous Non-equilibrium Transport and
Degradation of pesticides. Soil Sci. Sot. Am. J. 54:957-968.

RAJBHANDARI, N. B. 1997. Evaluation of the EPIC model to predict groung water contamination potential of different pesticides for Duplin County, NC. A report to NCSU Water Quality Group, North Carolina State University, Raleigh, NC.

SEYBOLD, C. A., MESWEENEY, K. and LOWERY, B. 1994. Atrazine Adsorption in sandy soils of Wisconsin. J. Environ. Qual., 23:1291-1297.

TURIN, H. J., BOWMAN, R. S. 1993. A Solid Phase-Extraction Based soil Extraction method for pesticides of varying polarity. J. Environ. Qual., 22:332-335.

Integration of GIS, Geostatistics, and 3-D Technology to Assess the Spatial Distribution of Soil Moisture

M. Betts, T. Tsegaye, W. Tadesse, T. L. Coleman, and A. Fahsi
Center for Hydrology, Soil Climatology, and Remote Sensing (HSCaRS)
Alabama A&M University, Normal, AL 35762

ABSTRACT

The spatial and temporal distribution of near surface soil moisture is of fundamental importance to many physical, biological, biogeochemical, and hydrological processes. However, knowledge of these space-time dynamics and the processes which control them remains unclear. The integration of geographic information systems (GIS) and geostatistics together promise a simple mechanism to evaluate and display the spatial and temporal distribution of this vital hydrologic and physical variable. Therefore, this research demonstrates the use of geostatistics and GIS to predict and display soil moisture distribution under vegetated and non-vegetated plots. The research was conducted at the Winfred Thomas Agricultural Experiment Station (WTAES), Hazel Green, Alabama. Soil moisture measurement were done on a 10 by 10 m grid from tall fescue grass (GR), alfalfa (AA), bare rough (BR), and bare smooth (BS) plots. Results indicated that variance associated with soil moisture was higher for vegetated plots than non-vegetated plots. The presence of vegetation in general contributed to the spatial variability of soil moisture. Integration of geostatistics and GIS can improve the productivity of farm lands and the precision of farming.

INTRODUCTION

Spatially distributed estimates of near surface soil moisture data are becoming increasingly important as inputs to spatially explicit landscape, regional and global hydrology models, and remotely-sensed data calibration (Tsegaye et al. 1997, Houser and Shuttleworth 1997). This variable is also a critical model parameter in climate studies. Studies have shown that a proper estimation of local and regional scale soil moisture distribution is an essential prerequisite for obtaining realistic results from atmospheric simulation models (Zhang and Randall 1997).

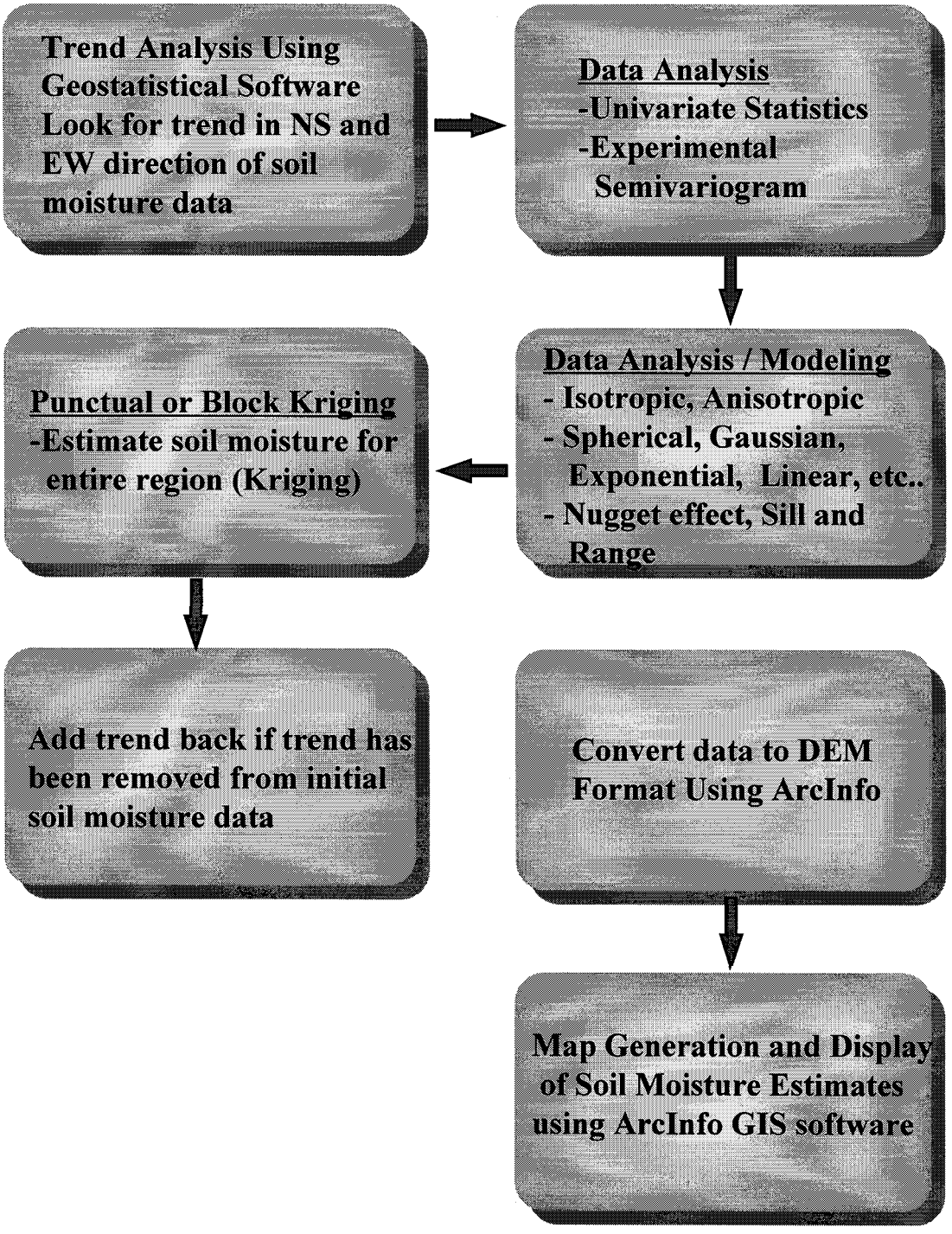
Soil moisture is a dynamic process and vary over space and time. Such variability directly affects the growth of plants and crop yields, and also impacts the water and energy cycle of the earth. Understanding land surface soil moisture status is essential for adequate management of natural resources and the environment. Soil moisture measurement technology has been growing rapidly. However, soil moisture estimation and data displaying techniques such as geostatistics and GIS have not been sufficiently utilized in soil moisture variability assessment for local and regional hydrological processes. Thus, the goal of this research is to improve the understanding and estimation of the space-time structure of soil moisture through integration of geostatistical interpolation techniques and displaying estimates of soil moisture using a geographical information system (GIS). The objectives of this study are: (1) to demonstrate the use of geostatistics and GIS to estimate and display the spatial and temporal distribution of soil moisture for vegetated and non-vegetated plots and (2) to compare the variance associated with soil moisture for vegetated and non-vegetated plots.

MATERIALS AND METHODS

The experiment was conducted at Alabama A&M University Winfred Thomas Agricultural Research Station at Hazel Green, Alabama. The soil at this site is a Decatur silt loam. Four 60 X 50 meter plots were used to conduct this study. Two plots were covered with tall fescue grass (GR) and alfalfa (AA), while the other two were left bare smooth (BS) and bare rough (BR) surface. Soil samples were taken from the four plots on a 10 by 10 m grid from O to 5-cm depth on three sampling periods to determine the moisture status of the soil. The traditional gravimetric soil moisture measurement technique was used to determine moisture content.

Figure 1 is a flow chart that shows the geostatistical and GIS steps followed to estimate the spatial distribution of soil moisture within each plot. The theory of regionalized variables (Matheron, 1971) was used to investigate the spatial variability of soil moisture content. The semivariance calculation, semivariogram function model fitting and kriging were performed using commercially available software, GS^+, (Gamma Design, Inc.; Plainwell, MI). The linear, linear/sill, spherical, exponential, and gaussian models were explored as models to tit the semivariogram functions. The spherical model was found to be the best fitted model among all possible models. Punctual kriging, which is an exact interpolator (Delhome, 1978), was used to estimate soil moisture values for unsampled locations.

Flow Chart for Geostatistical and Geographic Information System (GIS) Integration to produce Soil Moisture distribution map



RESULTS AND DISCUSSION

The statistical parameters used for estimating soil moisture are given in Tables 1 and 2. The mean soil moisture decreased by almost 6 to 12% over the three sampling period. The moisture loss was higher in the non-vegetated plot as compared to the vegetated plots. Direct exposure of the soil surface to incoming solar radiation and high water loss to evaporation contributed to the mean soil moisture variation over time. Except for the bare smooth plot, variance associated with soil moisture content increased over time (Table 1).

Table 1. Mean and standard deviation of gravimetric soil moisture content (%) for grass (GR), alfalfa (AA), bare rough (BR), and bare smooth (BS) for three sampling periods.

Plot	Sampling Period					
	1		2		3	
	Mean	SD	Mean	SD	Mean	SD
GR	16.37	1.84	16.13	2.35	15.27	2.18
AA	17.34	1.93	17.70	1.67	16.00	2.60
BR	18.69	1.39	18.53	2.03	16.92	2.21
BS	17.73	1.05	17.02	1.03	15.59	0.97

Geostatistical analysis of directional semivariograms suggested that the GSM varied isotropically for all sampling periods. The nugget variance was higher for the vegetated plots indicating that there is high spatial and temporal variability of soil moisture within the vegetated plots. It is also clear that there is strong spatial correlation within the vegetated plots as compared to the non-vegetated (Table 2).

Table 2. Geostatistical parameters used for kriging soil moisture content at unsampled location (Spherical model was used for all plots and sampling period).

Plot	Sampling Period								
	1			2			3		
	Nugget	Sill	Range (m)	Nugget	Sill	Range (m)	Nugget	Sill	Range (m)
GR	1.34	3.90	23.20	2.06	10.07	78.30	3.01	4.61	15.00
AA	2.32	4.07	15.00	3.01	3.01	11.80	4.69	7.13	15.00
BR	0.29	2.53	39.50	0.04	4.62	20.80	0.01	5.69	23.30
BS	0.59	0.95	15.00	0.00	1.18	24.10	0.45	1.85	102.40

The spatial and temporal distribution of soil moisture for period one and two of each plot are given in Figures 2 to 5. These 3-D figures show the continuous distribution of soil moisture estimates in each plot. The distribution of soil moisture is more uniform in the non-vegetated plot than in the vegetated plot for sampling period one and three. Erratic changes of soil moisture distribution were observed in the alfalfa plot (Figure 3a). This could be attributed to variation in soil moisture uptake by the plants and root distribution as well as the degree of plant foliage cover on the ground. Comparing the two sampling periods, more erratic soil moisture distribution was observed in the vegetated plots (Figures 2 and 3). High evapotranspiration demand, the change in soil physical properties such as pore size distribution, and distribution of plant roots could influence such variability over time.

Figure 2a. The spatial distribution of gravimetric soil moisture content (%) for tall fescue grass plot (Period 1)

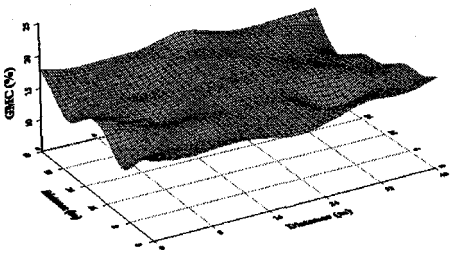


Figure 2b. The spatial distribution of gravimetric soil moisture content (%) for tall fescue grass plot (Period 3)

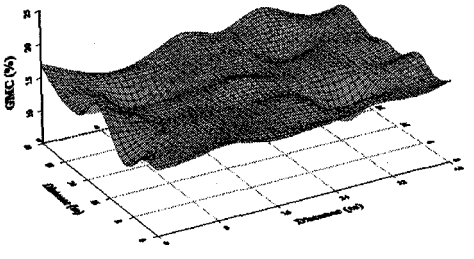


Figure 3a. The spatial distribution of gravimetric soil moisture content (%) for alfalfa plot (Period 1)

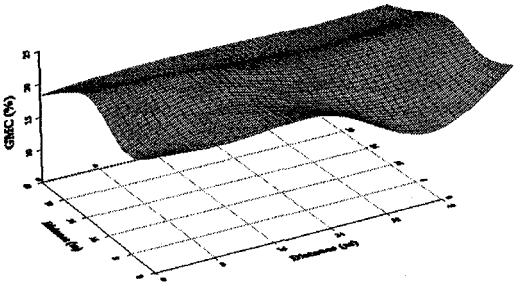


Figure 3b. The spatial distribution of gravimetric soil moisture content (%) for alfalfa plot (Period 3)

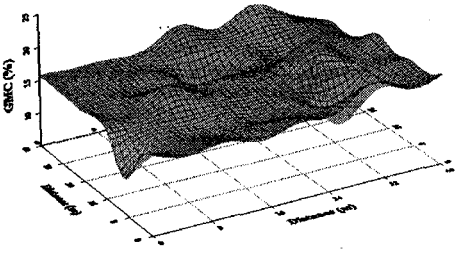


Figure 4a. The spatial distribution of gravimetric soil moisture content (%) for bare rough plot (Period 1)

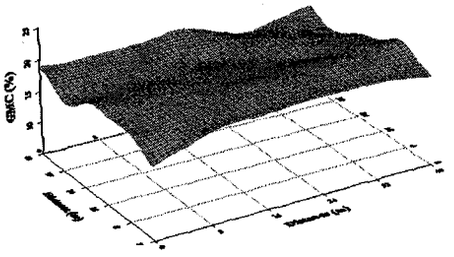


Figure 4b. The spatial distribution of gravimetric soil moisture content (%) for bare rough plot (Period 3)

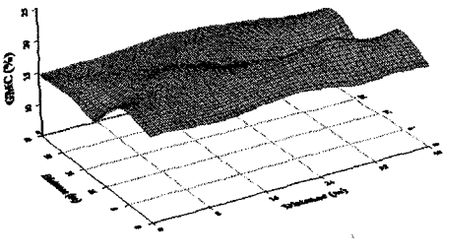


Figure 5a. The spatial distribution of gravimetric soil moisture content (%) for bare smooth plot (Period 1,

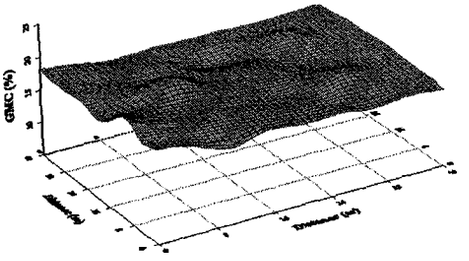
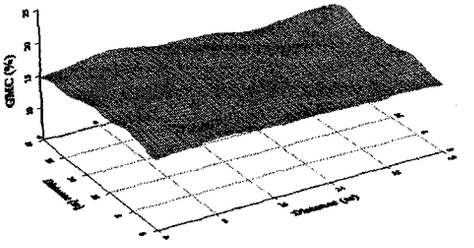


Figure 5b. The spatial distribution of gravimetric soil moisture content (%) for bare smooth plot (Period 3)



Geostatistical techniques were used to estimate soil moisture distribution **over the** entire sampling area. The use of GIS, on the other hand, enabled us to visually observe the spatial distribution of the soil moisture and understand the data better than looking at a tabular data. Both techniques were utilized simultaneously to facilitate our data manipulation and interpretation. The spatial and temporal differences observed from the data analysis may not be statistically significant. However, these visual differences may have tremendous effect on many agricultural practices that require high precision and accuracy for proper management of our farm lands.

CONCLUSION

The vegetated plots exhibited higher random variance as compared to non-vegetated plots. The variance also generally increase as the soil dries down. In our study, the integration of geostatistics and GIS assisted us to estimate and display the spatial and temporal distribution of soil moisture within vegetated and non-vegetated plots. The two techniques could be used widely to improve the precision of farming and increase the productivity of our farm lands by correcting deficient areas in a given field.

-ACKNOWLEDGMENTS

The authors gratefully acknowledge the Center for Hydrology, Soil Climatology, and Remote Sensing (HSCaRS) support staff, Mr. Johnny Boggs and Mr. Garland Robertson for their valuable time in assisting with the field and laboratory work. Contribution from the HSCaRS Center and Agricultural Experiment Station, Alabama A&M Univiversity . "This work was supported by Grant No. NCCW-0084 and NCC8- 140 from the National Aeronautics and Space Administration (NASA), Washington, DC. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government.

REFERENCES

Delhome, J.P. 1978. Kriging in the hydrosciences. Advances in Water Research. 1:251-266.

Houser, P. and Shuttleworth, W. 1997. Integration of remote sensing and hydrologic modeling using data assimilation. *13th Conference on Hydrology, American Meteorological Society*, p. 176-178.

Matheron, G. 1971. The theory of regionalized variables and its applications: Ecole de Mines, Fantainebleau, France.

Tsegaye, T., Coleman, T., Manu, A., Senwo, Z., Fahsi, A., Belisle, W.,. Tedesse, W., Robertson, G., Boggs, J., Archer, F., Surrency, J., Birgan, L., Laymen, C., Crosson, W., and Miller, J. 1997. The spatial and temporal variability of soil moisture with and without cover crop and its impact on remote sensing, *13th Conference on Hydrology, American Meteorological Society,* p. *349-351*

Zhang, C. and Randall, D. 1997. The simulated soil moisture budget and its sensitivity to variations of land vegetation parameters.*13th Conference on Hydrology, American Meteorological Society,* p. 167-168.

Long Term Effects of Poultry Litter on Soil Physical and Chemical Properties in Cotton Plots

J. Surrency, T. Tsegaye, T. Coleman, A. Fahsi, and C. Reddy*
CENTER FOR HYDROLOGY, SOIL CLIMATOLOGY, AND REMOTE SENSING
ALABAMA A&M UNIVERSITY, NORMAL, AL. 35762

ABSTRACT

Poultry litter and compost can alter the moisture holding capacity of a soil. These organic materials can also increase the nutrient status of a soil during the decomposition process by microbial actions. The objective of this study was to evaluate the effect of poultry litter and compost on the dielectric constant and moisture holding capacity of soil. The Delta-T theta-probe was used to measure volumetric soil water content and the apparent dielectric constant of the upper 6-cm of the soil profile. Soil texture, pH, and organic matter were also determined for each plot. Results of these analyses indicated that the pH of the soil ranged from 6.4 to 7.7 and the volumetric soil moisture content ranged from 0.06 to 0.18 m^3/m^3 for the upper 6-cm of the soil profile. The effect of poultry litter and compost on soil properties resulted in an increase in the volumetric moisture content and dielectric constant of the soil due to the improvement of the soil structure.

INTRODUCTION

Poultry litter, as a crop fertilizer is an economically and environmentally sound method of utilizing poultry production by-products and their nutrients. Millions of chickens are produced each year for consumption. These chickens produce several metric tons of litter per year. Litter is a mixture of feces, bedding, and water containing several nutrients of great importance for crop growth. The practice of applying litter, composted or fresh, as a substitute of traditional fertilizer is gaining popularity. The nitrogen availability from litter is a main focus due to the high requirement of nitrogen by crops. Its availability to plants must be studied to minimize environmental concerns as a result of runoff and leaching.

The use of fresh litter and compost usually increases organic matter content in the soil. This increase leads to an increase in the water holding capacity, less susceptibility to erosion, improved soil structure, and increased infiltration rates reducing runoff. The use of poultry litter, as a substitute to traditional fertilizers, is an excellent way to minimize costs and to use unwanted waste. However, surface and groundwater standards may be violated due to runoff of litter. Therefore, the litter concentrations must be monitored to avoid such violations. The conservation practices that have being promoted to poultry producers include the distribution of excess poultry litter with applicable nutrients to areas and watersheds outside of the poultry-producing region. Even though there are considerable amounts of research on poultry litter, the relationship between soil physical properties and dielectric constant (e) of the soil is not fully investigated in the poultry amended field plots. Therefore, the objective of this study was to evaluate the effect of poultry litter and compost on the dielectric constant and moisture holding capacity of a soil.

MATERIALS AND METHODS

The research was conducted at the Belle Mina Research and Experimental Station in Northern Alabama. The research plots received fresh and compost poultry litter in the last three years. The Delta-T 6-cm theta Probe was used to measure the volumetric water content and the apparent dielectric constant of the upper 6- cm of the soil profile. Disturbed soil samples were taken from the top 15-cm soil depth using a two inch Giddings probes (Carter, 1993) to determine particle size distribution, pH, and organic matter. Particle size distribution was determined by the hydrometer method (Gee and Bauder, 1986). The pH was determined by using a 1:2 soil solution ratio. A pH meter was then placed in the solution to determine the soil pH (Klute, 1986). Organic matter was determined by the Walkley-Black method utilizing acid bichromate digestion (Allison, 1965).

RESULTS AND DISCUSSION

The statistical variables such as mean, standard deviation, and minimum and maximum values for the soil physical properties are given in Table 1. Soil dielectric constant (ϵ) was measured before (ϵ_1) and after (ϵ_2) the application of poultry litter to the experimental plots. The addition of poultry litter increased the mean of ϵ by 14%.

Table 1. Statistical analysis of selected soil physical properties.

Variable	N	Mean	Std. Dev.	Minimum	Maximum
1	80	6.15	0.198	4.37	8.47
2	80	8.29	0.261	5.81	12.11
PH	80	6.99	0.226	6.37	7.73
% OM	80	1.13	0.299	0.21	1.7
% Sand	80	36.97	7.12	24.4	53.12
% Silt	80	42.01	5.64	17.8	52.72
% Clay	80	21.02	6.8	10.16	30.52
(m^3/m^3)	80	0.118	0.031	0.059	0.185
(m^3/m^3)	80	0.18	0.043	0.105	0.278

1 = Soil dielectric constant before application of poultry litter.
2 = Soil dielectric constant after application of poultry litter.

Among the soil physical properties studied, the percentage of sand and clay had significantly higher correlation coefficient (r) with ϵ (Table 2). The highly significant correlation indicate that these two soil variables can better explain the variability associated with ϵ. The presence of high organic matter in the soil increases the soil moisture holding capacity and improves the structure of the soil. The mean organic matter content prior to application of the poultry litter was 1. 13°/0. Therefore, it is expected that the additional application of poultry litter substantially contributed to the increased mean value of ϵ by 14°/0. As a consequence, there seems to have significant correlation between OM and ϵ. However, this finding needs to be validated by measuring OM after the poultry litter application.

Table 2. Pearson Correlation analyses for soil physical properties.

	1	2	pH	%OM	'/i Sand	%Silt	%Clay
	1.00	0.06^{ns}	0.07ns	0.14ns	0.44***	-0.09ns	-0.39***
		1.00	0.06ns	-O.16ns	0.42***	0.09ns	-0.53***
PH			1.00	0.15ns	0.13ns	-0.02ns	-0.12^{ns}
%OM				1.00	0.08ns	-0.03ns	-0.05^{ns}
%Sand					1.00	.().45 ***	.().67 ***
0/0 Silt						1.00	-0.36**

ns = not significantly different (p=0. 10)
*, **, *** = Significant at the 0.1, 0.01, and 0.001 probability levels, respectively

The above finding is further supported by comparing the different levels of application means rate (Table 3). There was no significant difference between the means of ϵ for the different application rates at ϵ_1. Interestingly, significant differences were observed between the different levels of application rates at ϵ_2. The highest level of application rate (120 kg N ha^{-1}) increased the ϵ by 6°/0.

Table 3. Comparing means of soil dielectric constant under different poultry litter amended field plots with four application rates.

	Rate (kg N ha-])	1	2
Level 1	0	6.00'	7.67^{b}
Level 2	40	6.10'	8.24^{ab}
Level 3	80	6.40'	8.12^{ab}
Level 4	120	6 . 0 5 '	8.70'

1 = soil dielectric constant before application of poultry litter.
2 = soil dielectric constant after application of poultry litter.

Comparison between treatments indicated no significant difference among the means of ϵ_1 (Table 4). On the contrary, during ϵ_2 significant differences were observed among the means of ϵ_2. One of the beneficiary effects of poultry litter amendment to our crop land was that it would increase the organic matter content of the soil over time and improve the structure of the soil. In turn, this will improve the soil moisture holding capacity and increase the soil ϵ.

Table 4. Comparison of Soil dielectric constant means by treatment.

Treatment	Treatment Name	1	2
1	Control	5.71'	7.67[ab]
2	CMP (Inhibitor)	6.25[a]	7.62[ab]
3	40 – Urea	6.00[a]	7.73[ab]
4	40- Urea+ CMP	6.40'	8.94[a]
5	40 – FL	5.66'	8.58[ab]
6	40 -FL+ CMP	6.35'	7.90[ab]
7	40 – CL	5.81[a]	8.07[ab]
8	40 -CL+ CMP	6.25[a]	8.24[ab]
9	80 – Urea	6.81[a]	7.29[b]
10	80- Urea+ CMP	6.76'	7.95[ab]
11	80 – FL	5.90[a]	8.58[ab]
12	80 -FL+ CMP	6.76[a]	8.29[ab]
13	80 – CL	6.10[a]	8.76[ab]
14	80 -CL+ CMP	6.10'	7.90[ab]
15	120 – Urea	5.57[a]	8.12[ab]
16	120- Urea + CMP	5.71[a]	8.41[ab]
17	120 – FL	6.10[a]	9.18[a]
18	120- FL + CMP	6.25[a]	8.47[ab]
19	120 – CL	6.66'	9.24[a]
20	120- CL + CMP	6.05'	8.94[a]

*=Means followed by the same letter are not significantly different from each other (p=0.05), $_1$=soil dielectric constant before application of poultry litter, $_2$=soil dielectric constant after application of poultry litter.

CONCLUSION

Selected soil physical properties including particle size distribution, soil pH, organic matter, and dielectric properties of a soil were evaluated in plots where fresh and poultry litters have been applied for more than three years. Results of this field study indicated that addition of poultry litter increased the soil dielectric property and soil water holding capacity. It is therefore advantageous to use poultry litter in areas where there is low level of organic matter content.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Center for Hydrology, Soil Climatology, and Remote Sensing (HSCaRS) support staff, Mr. Johnny Boggs for their valuable time in assisting with the field and laboratory work. Contribution from the HSCaRS Center and Agricultural Experiment Station, Alabama A&M Univ., Journal no. "This work was supported by Grant No. NCCW-0084 and NCC8- 140 from the National Aeronautics and Space Administration (NASA), Washington, DC. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government

REFERENCES

Allison, L.E. 1965. Organic carbon. p.1 367-1378. In C.A. Black et al. (Eds.) Methods of soil analysis part 2. Am. Sot. Agron., Madison, WI.

Carter, M.R. 1993. Soil Sampling and Methods of Analysis, Canadian Society of Soil Science, Lewis Publisher, CRC Press, Inc., Florida.

Gee, G. W., and J.W. Bauder. 1986. Particle size analysis. p.383-412. In A. Klute (cd.) Methods of soil analysis part 2nd ed. Agron. Monogr. 9 ASA and SSSA, Madison, WI.

Klute, A. 1986. Methods of soil analysis. Part 1. Physical and mineralogical methods. 2nd ed. Agron. Monogr. 9 ASA and SSSA, Madison, WT.

"Fabrication and Atomic Force Microscopy Characterization of Molecular Composites of Fullerenes in Aerogel Matrix for Optical Limiting"

W.J. Lu*, H. B. Sunkara***, D. Shi*, S. H. Morgan*, B. Penn**, D. Frazier**, W. E. Collins*

*Center for Photonic Materials and Devices
Department of Physics
Fisk University
Nashville, TN 37208

** ES76, Marshall Space Flight Center
Huntsville, AL 35812

*** DuPont R & D Center
Wilmington, DE 19898

Abstract

An optical limiter is a device which exhibits a decrease in the transmittance in a material with an increase in intensity of light. Sol-gel techniques offer many advantages in the fabrication of materials. These materials possess many desirable properties for nonlinear optical (NLO) device applications which include transparency, high thermal and chemical stabilities, very low refractive index and dielectric constants. C60 shows a higher excited state absorption cross section than the ground state absorption cross section over the complete visible spectrum, and the spectrum of the excited state absorption of C60 has the same general shape as the ground state absorption. This fact suggests that fullerenes are ideal optical limiting materials.

Aerogels are fabricated by sol-gel processing. One of the key issues is the dispersion of fullerenes into small and uniform pores of silica aerogel host matrices. The aerogel network was characterized by Raman spectroscopy. Atomic force microscopy is a technique with many advantages to characterize the aerogel materials. The morphology of the cleaved surface for a C60/aerogel sample shows that there are long paralleled shaped stripes with 20-30 nm in width and about 500 nm in length on the cleaved surface. The cleaved surface also was etched b y5°/0 HF solution for one minutes, and it became smoother after HF etching. The main feature in on the surface is the spherical particles with the size of few nanometers, and no aggregated fullerenes appear. The fullerenes are well dispersed in the aerogel matrices.

Key words: aerogel, fullerene, optical limiter materials, and atomic force microscopy,

1: Introduction.

An optical limiter is a device which exhibits a decrease in the transmittance of a material with an increase in intensity of light. All photonic sensors, including the human eye, have an intensity level above which damage occurs. Using an appropriate optical limiter in the system prior to the sensor extends the dynamic range of the sensor and allows the sensor to continue to operate under harsher conditions than otherwise possible,

The discovery of the carbon-caged fullerenes [1 ,2] has had enormous effect in this field. C60 shows a higher excited-state absorption cross section than the ground-state absorption cross

section over the complete visible spectrum, in other words, reverse saturable absorption (RSA) [3]. The spectrum of the excited state absorption has the same general shape as the ground state absorption, which implies the induced absorption will not vary widely over the visible spectrum. This fact suggests that the fullerenes are ideal optical limiting materials, and have a promising application in optical limiting for sensor protection. In recent years, optical limiting with fullerene (C60, C70 and their derivatives) solutions [4,5] has been well established. However, environmental instability and difficulties in using liquid phases are major obstacles for device applications.

Therefore, several research groups have attempted to improve the environmental stability by doping fullerenes in solid organic and inorganic polymers which include polymethyl methacrylate, polystyrene, silica xerogels, sonogels, and glasses [6- 12]. The optical properties of fullerenes have been retained in such solid matrices, nonetheless, the C60 solid materials showed much reduced performances in poly(methylmethacrylate) when compared to their liquid analogs [12]. One of the major reasons for the reduced performance is the lack of molecular dispersion in these solid matrices.

Sol-gel techniques offer many advantages in the fabrication of materials. For example, using the technique, more uniform phases can be distributed in a multi-component system, size and morphological control is easy, and lower temperature processing and consolidation is possible. A few reports have shown that sol-gel techniques can enormously improve the dispersion of fullerenes in solid matrices, and affect the optical/electrical properties [6,9,10,13]. Zhu *et al.* **[9]** reported that fullerenes can be uniformly encased in silica aerogel by the sol-gel process together with the supercritical drying technique. The optical characterization of the materials confirms that the C60 and its derivative are embedded in the solid matrix. Bentivegna *et al.* **[6] also** reported that the nonlinear optical properties of fullerenes are retained when they are enclosed in the solid xerogel matrix derived from tetraethoxysilane (TEOS), methyltriethoxysilane (MTEOS), and vinyltriethoxysilane (VTEOS).

Low density aerogels are promising matrix materials for nonlinear optical materials because of their unique properties: low density, small pore size, transparency, high thermal stability and tunable dielectric constant [14, 15]. The doping of fullerenes into such matrices would result in the benefits of a low optical loss from the host and desired optical nonlinearities of guest species. These materials possess many desirable properties for nonlinear optical (NLO) device applications which include transparency, high thermal and chemical stabilities, very low refractive index and dielectric constant. Aerogels have the advantages of low cost, and ease in manufacture.

The goal of the work is to overcome the problems of fullerenes as optical limiting materials, to improve the optical limiting properties, and to develop the practical applications. We are presenting a preliminary study of developing solid fullerene materials that have enhanced optical and environmental properties *via* sol-gel techniques. The fullerenes in aerogel composites sample was prepared. The atomic force microscopy was used to characterize the dispersion of fullerenes.

2: Experimental methods.

a: The sample:

The sol-gel technique has been well established [14,1 5]. The aerogel-fullerene sample was prepared at Marshall Space Flight Center. The preparation conditions of aerogel-fullerene is adjusted to achieve the optimal properties for optical limiting. The process is divided into the

following steps: (1) preparation of the solution, (2) gelation, (3) aging, and (4) drying.

b. Atomic Force Microscopy:

A NanoScope E (Digital Instruments, Inc.) scanning probe microscopy is used in the study. The sample was cleaved, and the AFM images were measured on the cross-section.

3.Results and Discussions.

A number of characterization techniques can be used to establish the chemistry-structure-property relationship of the aerogel/xerogel-fullerene composites. AFM is a technique with unique advantages in studying aerogel/xerogel materials. Compared to other microscopy techniques, such as TEM and SEM, AFM has a high resolution, a very simple sample preparation procedure, and more importantly, it does not require the sample to be conductive. AFM (atomic force microscopy) is a very effective technique to study fullerenes dispersed at the silica aerogel surface.

Figures 1 and 2 show the AFM images of the cross-section for the aerogel fullerene composite sample. As shown, the main feature is the long straight stripe shapes on the surface, The size of these stripes is at least 500 nm in length and about 20-30 nm in width. The section analysis of Figure 2 is shown in Figure 3. The surface roughness is in the range of 20 nm. From Figures 1-3, we are unable to see individual fullerenes on aerogel composite. Therefore, the fullerenes are well dispersed in the aerogel networks, not aggregated, and the surface features are the result of the structure of aerogel composite.

In order to expose more fullerenes on the surface, the cleaved sample was etched in 5% HF solution for one minute. Figure 4 shows the morphology of the etched sample. As shown, the surface became smoother. The section analysis of Figure 4 is shown in Figure 5; the roughness is about 5 nm after etching. Figure 6 shows the morphology of the etched sample at 100 nm scale. The main feature in Figure 6 is the spherical particles with the size of few nanometers on the surface, and no aggregated fullerenes appear. This feature represents the fullerene caged aerogel composite, although we can not identify the fine structures and chemical composition by AFM.

Acknowledgment:

This work was partially supported by NASA under Grant No. NAGW-2925 and NAG3-1430.

References:

[1] H.W. Kroto, S.C. Heath, R.F. Curl, R.E. Smalley, *Nature, 318, 162(1985).*

[2] W. Kratschmer, L.D. Lamb, K. Fostiropoulos, D.R. Huffman, *Nature, 347, 354 (1990),*

[3] R.J. Sension, C.M. Phillips, A.Z. Szarka, W.J. Romanow, A.R. McGhie, J.P. McCauley, Jr., A.B. Smith, III, and R.M. Hochstrasser, *J. Phys. Chem.,* 95, 6075(1991).

[4] L. Yang, E. Royer, A.D. Walser, and R. Dorsinville, *Chemical Physics Letters, 239, 399(1995).*

[5] J.L. Bredas, C. Adant, P. Tackx, A. Persoons, B.M. Pierce, *Chemical Review, 94, 243(1994).*

[6] F. Bentivegna, M. Canva, P. Gerorges, A. Brun, F. Chaput, L. Malier, and J.-P. Boilot, *Applied Physics Letters, 62(15), 1721(1993).*

[7] J.W. Perry, K. Mansour, I.-Y. S. Lee, W.-L. Wu, P.V. Bedworth, C.-T. Chen, D. Ng, S.R. Marder, P, Miles, T. Wada, M. Tian, and H. Sasabe, *Science, 273,* 1533(1996).

[8] Y. Kojima, "1'. Matsuoka, H. Takahashi, and T. Kurauchi, *Macromolecules,* 28, 8868 (1995).
[9] L. Zhu, Y. Li, J. Wang, and J. Shen, *J. Appl. Phys., 77(6), 2801 (1995).*
[1 O] M. Prato, M. Maggini, G. Scorrano, G. Brusatin, P. Innocenzi, M. Guglielmi, M, Meneghetti, and R. Bozio, *"Science and Technology of Fullerence Materials",* MRS Symposium Proceedings, **Vol.** *39,* Edited by P. Bernier *et al.* Boston, MA. USA, p351,1994.
[11] M. Brunei, M. Canva, A. Brun, F. Chaput, L. Malier, and J.-P. Boilot, *"Materials for Optical Limiting",* MRS Symposium Proceedings, **Vol. 374,** Edited by R. Crane *et al.* Boston, MA. USA, p281,1995.
[12] A. Kost, L. Tutt, M.B. Klein, T.K. Dougherty, W.E. Elias, *Opt. Lett., 18, 334(1993).*
[13] S. Dai, R.N. Compton, J.P. Young, and G. Mamantov, *J. Am. Ceram. Sot., 75(10), 2865(1992).*
[14] L.L. Hench, and J.K. West, *Chemical Review, 90, 33(1990).*
[15] C.J. Brinker, and G. W. Scherer, "Sol-gel *Science",* Academic Press, Inc., 1990.

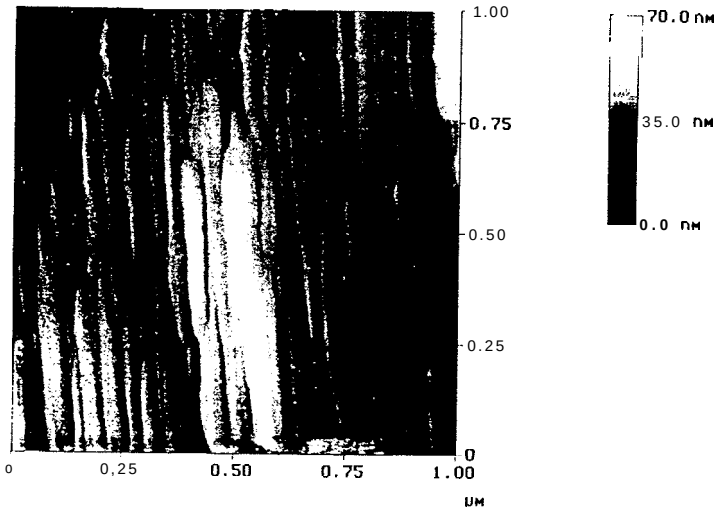


Figure 1. AFM image of the fullerene/aerogel composite.

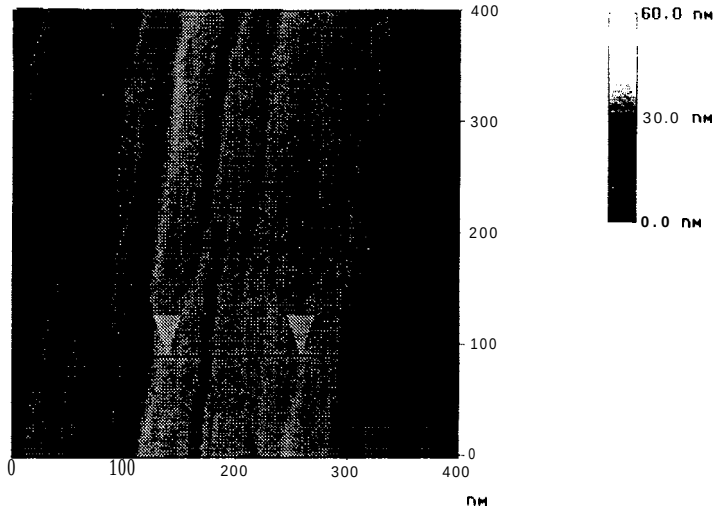


Figure *2.* AFM image of the fullerene/aerogel composite at small scale.

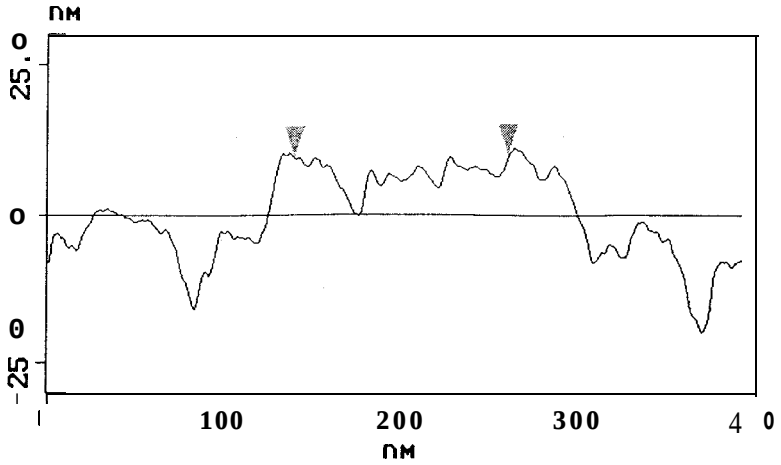


Figure 3. The section analysis of Figure 2.

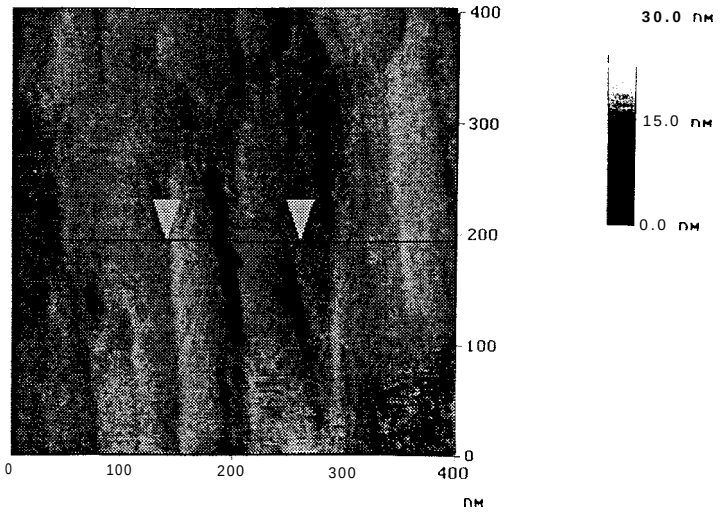


Figure 4. AFM image of the etched fullerene/aerogel composite.

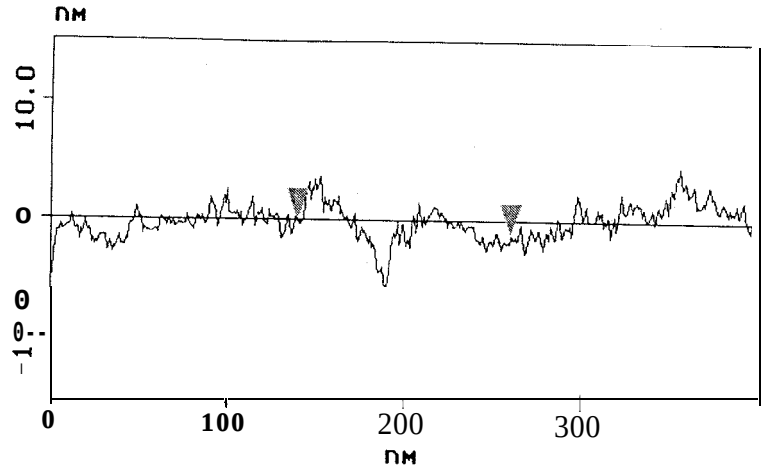


Figure 5. The section analysis of Figure 4.

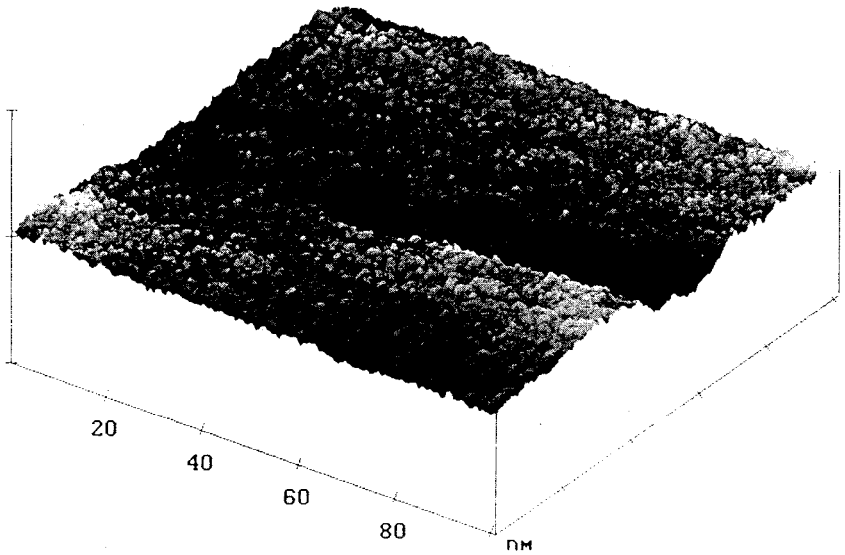


Figure 6. AFM image of the fullerene/aerogel composite at small scale.

Research and Interpretation of the Camino Real from Taos, New Mexico, to Parral and Chihuahua, Chihuahua

by
Alfredo Enriquez, PACES, UTEP, Dr. John A, Peterson, Anthropology Research Center, UTEP, and Dr. Nicholas E, Pingitore, Jr., Department of Geological Sciences, UTEP

The Camino Real from Chihuahua to Taos was the major thoroughfare for Spanish exploration and settlement of the region of northern New Spain. Missions, presidios, and haciendas, the three vital links in the Spanish Colonial system, are all still to be found in the region, dating from as early as the late 17th century. The Camino Real was the connective fiber for these sites during historical times. The archaeological traces of the Camino Real arid its associated settlements unfortunately are poorly documented and have never been integrated into models for the region.

Human settlement in what was northern New Spain has been contingent on water resources. The nodes of settlement and the routes of the Camino Real were constrained by water availability throughout the arid expanses of the Chihuahua Desert. Prehistoric settlement, which has extended at least 12,000 years before present in the region, likewise was limited by water resources, many of which have changed through the Holocene. However, the correlation of known archaeological sites with existing and fossil water resources has never been evaluated.

The goal of this project is to use satellite imaging to delineate remaining traces of the Camino Real and associated water resources. These findings are being compared with historical records pertaining to that route . Historical and prehistoric archaeological settlement data and geohydrological data for northern Mexico and the American Southwest are integrated with the NASA data in order to develop predictive models for human settlement. This project correlates investigations of the Camino Real and colonial era settlements, prehistoric settlement systems in the region, and geohydrological investigations to model surface and near-surface water availability throughout the settlement history of the region . A predictive model for human settlement based on geohydrological resources is developed and applied to contemporary as well as historical settlement.

The proposed project has three objectives:

1. Remote sensing data for the region is being compiled and integrated into a GIS in order to search for traces of the Camino Real and other road systems in the region, clusters of settlement, arid to model hydrological resources in the region. A variety of data sources has been shown to be effective for these purposes, ranging from vegetation analysis of LANDSAT data to SIR geophysical prospection strategies. Established ground sections of the Camino Real have been visited for correlation with remote sensing interpretations.
2. Archival historical as well as site data at repositories such as the ARMS database of the Historic Preservation Division, State of New Mexico, and the records of the *Instituto National de Antropología e Historia in* Chihuahua are being compiled in order to develop GIS layers of prehistoric and historical period settlement for the region. Field visits to sites and site clusters have been undertaken where necessary to clarify or ground-truth data.
3. A predictive model is being developed using previous studies of human settlement, remote sensing data of transport systems and settlement clusters, and geohydrological resources. The model is being

applied to contemporary settlement patterns and potential for sustainability of hydrological resources and settlement systems in the region.

Environmental Background

The study area is located in the Chihuahua Desertscrub biogeographic region which surrounds the El Paso area and extends south along the Rio Grande to around Langtry, Texas, and into northern Chihuahua and Coahuila in northern Mexico. The topography of the area is basin and range characterized by anticlines dissected by steep canyons with an absence of permanent streams except for the Rio Grande. The elevation of the El Paso area ranges from 3,600' msl in the area of Clint in the Lower Valley of El Paso to as high as 8,500' msl in the Franklin Mountains.

Vegetation communities of the Chihuahua Desertscrub include Chihuahuan riparian scrub, consisting of mesquite and acacia scrub along watercourses and floodplain; creotebush (Larrea spp.) and saltbush communities along lower slopes and the bajada, mixed scrub (succulent) communities along higher slopes; and semidesert grassland, interior chaparral, and madrean montane forest on higher slopes. The range of vegetation communities conforms to the diversity of the landscape and the opportunities and constraints provided by variation of slope, aspect, and altitude (Brown 1982). The diversity of plants in these communities provides numerous microenvironmental opportunities within a relatively small catchment area for human harvesting.

The climate of the El Paso area is typical of the arid to semiarid region of the southwestern U.S. and northern Mexico. The days are warm and the nights are cool. Summers are long and hot. Winters, though short, may include brief periods of below freezing temperatures, In late winter and early spring, high winds and blowing sand are common. The humidity is low and the evaporation rate is high and greatly exceeds the average annual rainfall of about 8.0 inches (Texas Department of Water Resources 1984). The desert washes, dry most of the year, become filled with water after a rain. Some basins, without drainages to the sea, lose their water to shallow playa lakes that go completely dry during rainless periods (Dice 1943).

The Rio Grande, while usually depicted prior to the 20th century as a perennial and reliable stream was subject to severe flooding and intense drought. These periods of maximal flow and drought sequences characterize the minimal conditions for settlement and subsistence along the river. Historic accounts of drought so severe that the river channel dried up indicate that the river was not a reliable source of water, and frequent flooding often destroyed what had survived the droughts. Since the Rio Grande is an exotic stream where snowpack from the Rocky Mountains makes a considerable contribution to stream flow even in the El Paso area, local, regional, and extra-regional climatic patterns must all be considered for local interpretations. Furthermore, cultural patterns such as increased upstream irrigation may have severely affected streamflow in Historic periods of occupation. Other than the throughflow of the Rio Grande in the northern part of the Chihuahua Desert and the Rio Conchos in the central and eastern part, where is flows from near Parral, Chihuahua northeast to its confluence with the Rio Grande near Presidio, Texas, the region is characterized by vast basins which are internally drained. These basins consist of Plio-Pleistocene deposits formed from playas and have potentially vast groundwater reserves, much like the Hueco Bolson and Mesilla aquifers in the El Paso, Texas area which constitute one of the major water supplies even where the Rio Grande supply is available, These groundwater resources are relatively shallow in many areas, and occasionally emerge as springs in areas such as the Samalayuca Mountains.

Historic Cultural Background

Spanish Period

The first Spanish presence to the area may well have been that of Cabeza de Vaca, whose route may have taken him through the Lower Valley on his journey from the Texas Gulf Coast to the Sea of Cortez where he returned to Spanish territory. The Lower Valley was not visited again until the expeditions of Chamuscado and Rodriguez in 1581 and of Espéjo in 1582.

The colonizing expedition of Juan de Oñate in 1598 led to the first permanent settlement along the Rio Grande, but it was several decades before the Lower Valley was settled by the Spanish. While there had been some earlier attempts to colonize the area, the first mission to the Manses was not established until 1659. The forced re-location of Piro, Tompiro, and Tigua from the northern Rio Grande following the Pueblo Revolt in 1680 led to the establishment of large communities in the area. These communities, which evolved into the present towns of Ysleta, Socorro, and San Elizario, began as missions to the transplanted puebloans and, in the latter case, as a presidio that was built for protection from Apache raiding.

Jornada de Cantarrecio, or Camino de Chihuahua

The route used by Juan de Onate which probably intersected the river 50 km southeast from Nuestra Señora de Guadalupe where the Arroyo de las Bandejas enters the floodplain. Casa de huesos, excavated by Gerald *et al.* but never reported, is a Spanish Colonial site located at the mouth of the Arroyo. Also, the "fort" of Cantarrecio is located 25 M up the Arroyo near the present-day Rancho de las Tinajas (Gerald 1990). This site may also have been occupied during Spanish Colonial times. This route was pioneered by the Juan de Onate expedition in 1598 as the most direct route to the river from Villa Ahumada. This route largely avoided the Samalayuca dunes, which were treacherous for travel by oxcart and mule wagon, However, a more direct route to Paso del Norte was eventually developed in the late 18th century as that part of the Rio Grande valley eclipsed the importance of more southern settlements. The present imagery depicts this late route of the Camino Real.

Results

The use of LANDSAT data has been instrumental in the research. By utilizing the LANDSAT data and manipulating it into a digital format, it is possible to create images based on light reflected off of the earth's surface. The data is capable of showing such things as visible light, infrared, and other wavelengths on the geography of the image. By utilizing this technology it is possible to view landscapes in graphic detail and with more potential for information than is available by use of maps or aerial photography alone. The geography of the image shows detailed environmental impacts not easily visible to ground surveyors because of the magnitude of the impacts as a whole. Such an impact is clearly shown on the LANDSAT-based imagery for much of the northern Chihuahua area, where it is possible to view trails that maybe remnants of the Camino Real del Tierra Adentro. An image of the Samalayuca area in northern Chihuahua, Mexico clearly displays a dark linear pattern that lies in the area that contained the Camino Real. This linear pattern is also shown as a trail on the Samalayuca quad topographic map, The linear pattern appeared clearly different from the straight lines that delineate the paved roads and the lighter colored linear patterns that delineate dirt roads. Thus, it was hypothesized that this dark linear pattern, different from other known patterns and lying in the historically correct area for the Camino Real, maybe visible remnants of the Camino Real picked up by the satellite digital imagery.

With this hypothesis in mind, a field visit to the area was initiated in order to discover whether the linear image showing was possibly some form of trail used by the locals or perhaps a country road that simply varied from the norm in coloring when picked up by the LANDSAT data. A ranch named Rancho Lobo was used as the site from which the field research was launched. The ranch owner

claimed not to know of any trails or roads lying in the area that the LANDSAT image showed the trail, but only knew of the previous Mexican highway that lay behind his ranch, close by. This old highway was clearly shown on the LANDSAT image as a straight light-colored line, and lay just over 200 meters away from the ranch. The field crew walked the sandy dual area east of Rancho Lobo in search of some physical evidence of the trail image. A series of linear depressions in the dunes was found that were noticeably differing from the surrounding landscape. Lengthwise, the depressions ran north/south and contained larger and more abundant vegetation than the surrounding sandy dunes, which may serve as an explanation for the coloring of the linear pattern on the LANDSAT imagery. The distance from Rancho Lobo to the linear depressions was paced at 2.7 km. east of Rancho Lobo. No rural roads or dirt trails were found aside form the linear depressions, indicating that the linear pattern picked up by the LANDSAT image was not a road or trail currently in use by locals. According to the imagery, the trail seems to head west just before reaching the dunes and run closer to the Sierra Samalayuca mountains, which would be reasonable for the Camino Real to do as there are springs located at the base of the mountains mainly along their northern edge. The imagery may be further manipulated to detect water sources and internally drained basis by using different bands to show the wavelengths of light being reflected, such as infrared, ultraviolet, or even thermal readings. With such LANDSAT data, it is possible to highlight areas of more abundant vegetation and water sources. Possible parajes were also logged along the highway for further possible inspection. These areas were noted by more abundant vegetation and larger types of vegetation.

In many areas, the Camino Real may have been a number of roughly parallel routes, such as across the grasslands near Chihuahua, but in dunes it was probably a single route which may have been maintained and was probably compacted and along higher ground in order to avoid effort of travel through the loose dual sands. With this in mind, it seems probable that the Camino Real would still be visible to larger and broader images such as the LANDSAT imagery.

Future efforts of the research project include using gar with a 225 Megahertz antenna to characterize shallow aquifers near Carrizal and Montezuma prehistoric sites (both near Villa Ahumada). Also, further LANDSAT data will be utilized to attempt to examine the effects of loose sands, which are in abundance in the Samalayuca area, against compacted sands. Vegetation changes will also be researched, as well as an examination of groundwater sources through changes in terrain as well as vegetation. Also, the gar studies on aquifer sites may contribute a new tool for analysis of groundwater resources in an arid region which is currently experiencing a crisis of water supply.

Conclusions

Early analysis indicates potential for the use of LANDSAT data for this type of research. Further use of images to detect earth and sand compaction, as well as the use of long-wave bands to discover water sources and vegetation communities are ongoing and may further aid the research. The imagery indicates a trail that gives positive indication that it may be the Camino Real, and further research may help resolve this question of trail identity.

References Cited

Brown, David E.

1982 Biotic Communities of the American Southwest-United States and Mexico. In *Desert Plants.* Published by the University of Arizona for the Boyce Thompson Southwestern Arboretum.

Dice, Lee R.

1943 *The Biotic Provinces of North America.* University of Michigan Press, Ann

Arbor.

Gerald, Rex

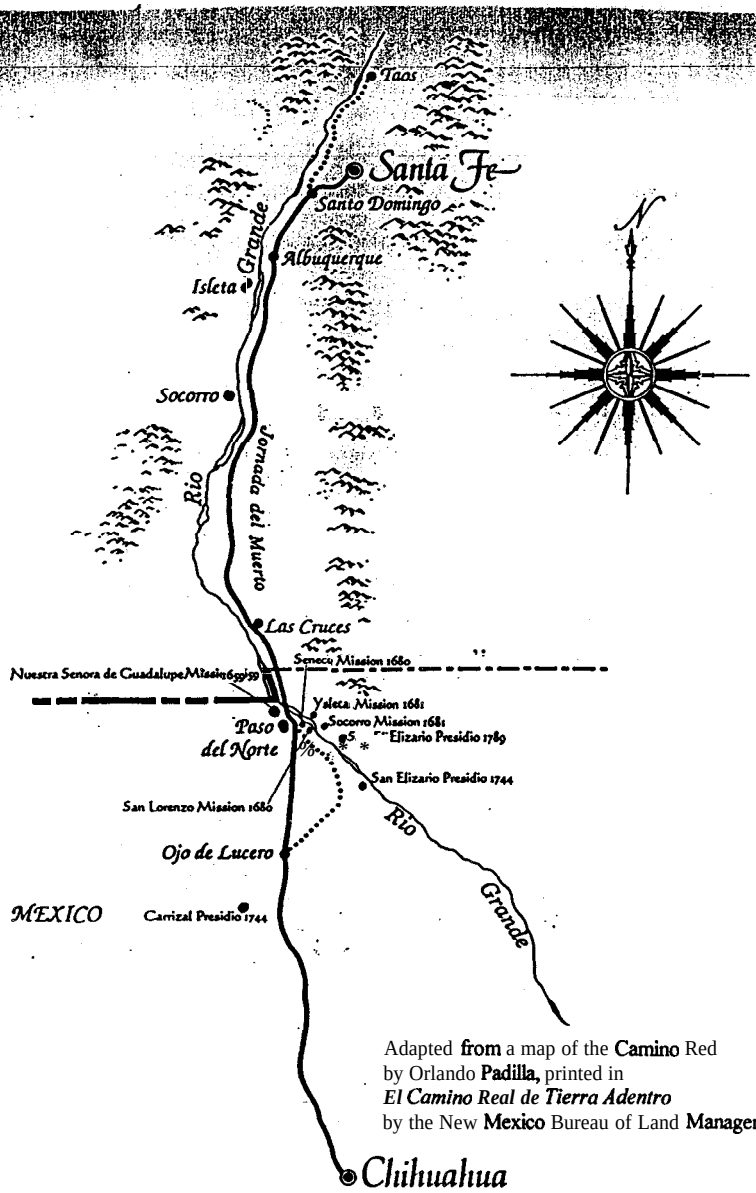
1990 Collected Papers. *The Artifact*, Volume *28,* Number 3. El Paso Archaeological Society, Inc., El Paso.

New Mexico Bureau of Land Management

1989 *El Camino Real de Tierra Adentro.* Bureau of Land Management New Mexico State Office, Santa Fe.

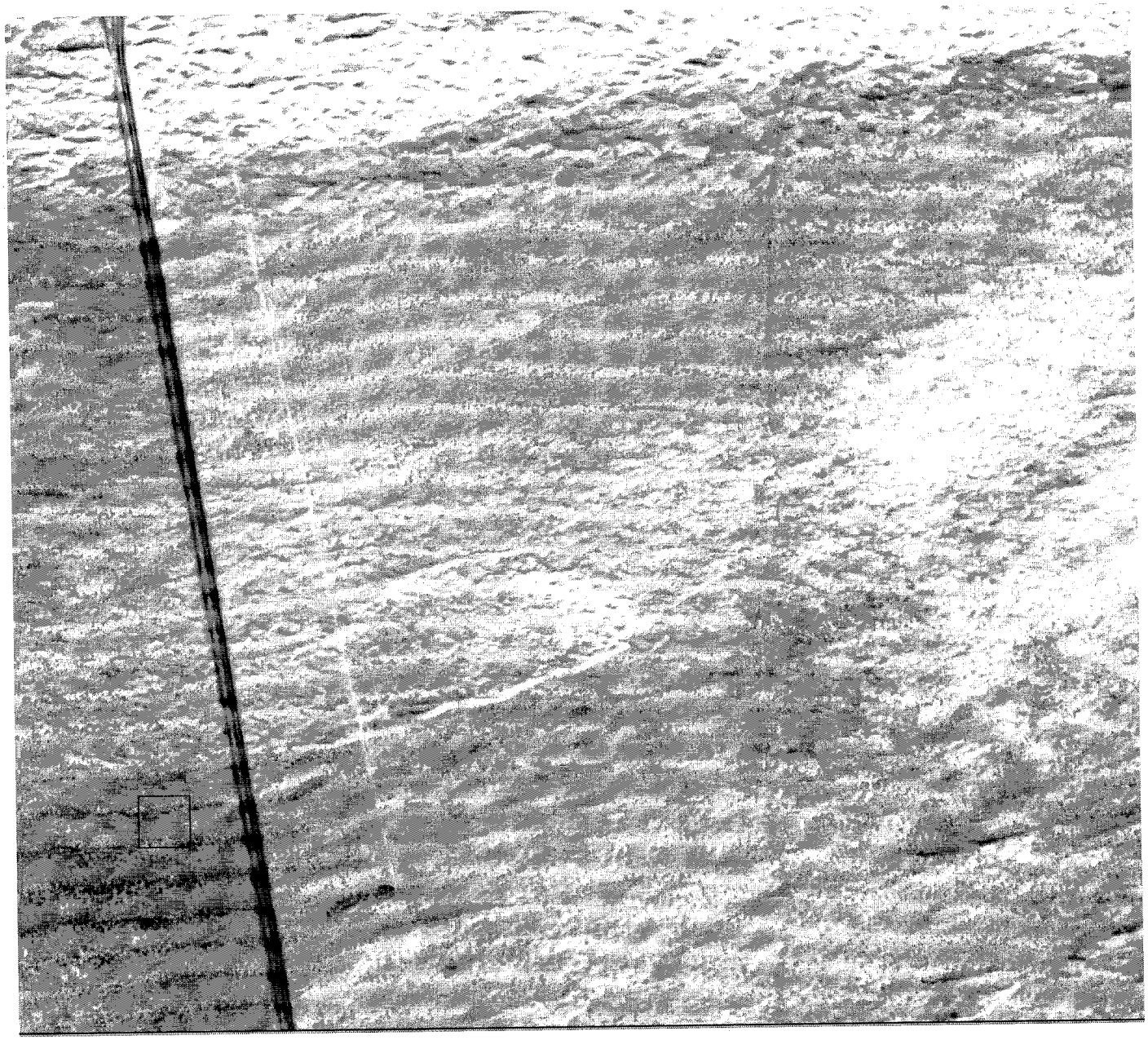
Texas Department of Water Resources

1984 *Water for Texas Technical Appendix,* Volume Two. Austin, Texas. Texas Department of Water Resources.,



Adapted from a map of the Camino Red by Orlando Padilla, printed in *El Camino Real de Tierra Adentro* by the New Mexico Bureau of Land Management

LANDSAT image showing an area of Chihuahua, Mexico south of the Samalayuca dunes, along Highway 45, The Highway 45 is the dark straight line on the left side of the image, the abandoned highway is the white straight line running parrallel to Highway 45, and the Camino Real is the north/south running line on the right side of the image.



A New Image Processing and GIS Package

98 URC166

D. Rickman[1], J. C. Luvall[1],and T. Cheng[2],

1-Global Hydrology and Climate Center/NASA, 2-private consultant

The image processing and GIS package ELAS was developed during the 1980's by NASA. It proved to be a popular, influential and powerful in the manipulation of digital imagery. Before the advent of PC's it was used by hundreds of institutions, mostly schools. It is the unquestioned, direct progenitor or two commercial GIS remote sensing packages, ERDAS and MapX and influenced others, such as PCI. Its power was demonstrated by its use for work far beyond its original purpose, having worked several different types of medical imagery, photomicrographs of rock, images of turtle flippers and numerous other esoteric imagery.

Although development largely stopped in the early 1990's the package still offers as much or more power and flexibility than any other roughly comparable package, public or commercial. It is a huge body or code, representing more than a decade of work by full time, professional programmers. The current versions all have several deficiencies compared to current software standards and usage, notably its strictly command line interface. In order to support their research needs the authors are in the process of fundamentally changing ELAS, and in the process greatly increasing its power, utility, and ease of use. The new software is called ELAS II.

ELAS was conceived and built with several key ideas. All data that can be expressed in raster form is stored in a single format read by all relevant modules. Intermediate information, process controls, statistics and all similar data are stored in a form available to the user. The software may check and prevent fundamental errors in input and output, such as attempting to write a 32bit value into an 8 bit storage, but for most operations the user is assumed to be responsible for the results of his choices. And there is no assumption of a normal processing flow. Instead, each module does a discrete operation or suite of operations independent of data source or future plans. This means that ELAS is an excellent choice for dealing with raster data when the user needs to do something other than what a programmer planned. Overcoming the limitations in the code while retaining and increasing its power has be our challenge.

Our approach is several fold with one key concept which appears to be unique. The existing command line interface is predicated on use of hard coded assignments taking input from the keyboard and writing output to the terminal. This will be changed to utilizing named pipes. This single seemingly simple change has multiple ramifications. Initially the most important will the addition of a GUI, with all the benefits implied by current GUI technology. This is illustrated by the new program, ELASHeader, for manipulating the information stored in each data file's header. This is illustrated in figures 1 and 2.

Where we depart from the ordinary is the ability to make the GUI *intelligent/* The control data and response information provided by ELAS are fairly simple. As long as the incoming pipe provides the necessary information, the module will execute.

An application of this concept has been developed for setting the area to be processed. Virtually all of the modules (the basic unit of ELAS seen by a user) have processing controls expressed in file coordinates, termed elements and lines from the original MSS data usage standard. Thus processing X-Y extents are given by parameters termed IE (Initial Element), LE (Last Element), IL (Initial Line), LL (Last Line). To set the extents the user would type the string: LE 538 IE **252** IL 63 LL 789. While order is not significant and use of short, two letter names is a saving for typing, it is not possible to do error checking and such strings are cryptic at best for a beginner to

learn. Further, while the basic ELAS file definition includes Earth geometry information, this is generally not available to the user to control processing. Therefore, if a certain geographic extent is to be worked the user must compute the equivalent element, line values and process based on the computed values. The new alternative using an intelligent GUI is shown in Figure 3.

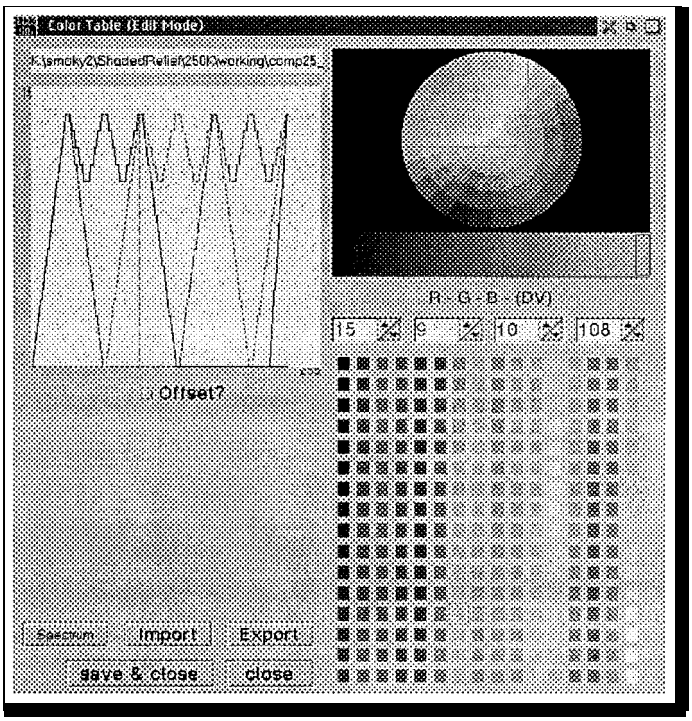


Figure 1. Color table editing.

The code behind this tool allows the user to change element and line value without typing the parameter names simply by tabbing between entry points. Thus it can actually be quicker than the older interface, which requires more keystrokes. Several advantages are also gained. Invalid entries are checked and corrected. The user can also select values using spinners. The area being processed is shown compared to the total file in a manner that maintains the aspect ratio of the area to process and the file extents. The user can see at any time the total possible range and the amount of processing being defined. The user can also define the area to be processed by point and click within the graphic window. When the code becomes available to generate thumbnails of the imagery, the graphic window will show the actual data to be processed. These are all within any good GUI added to a preexisting command line package. What is special is the addition of controls based on coordinates. The user can set limits based on projection coordinates and the intelligence in the interface will compute the appropriate element, line values. To do this requires more than just the information needed by the ELAS module and therefore is a separate body of code. In the standard software paradigm to add this capability to the ELAS modules each module would have to be changed. But by changing the module to used named pipes this can be implemented in one piece of code and applied across all modules.

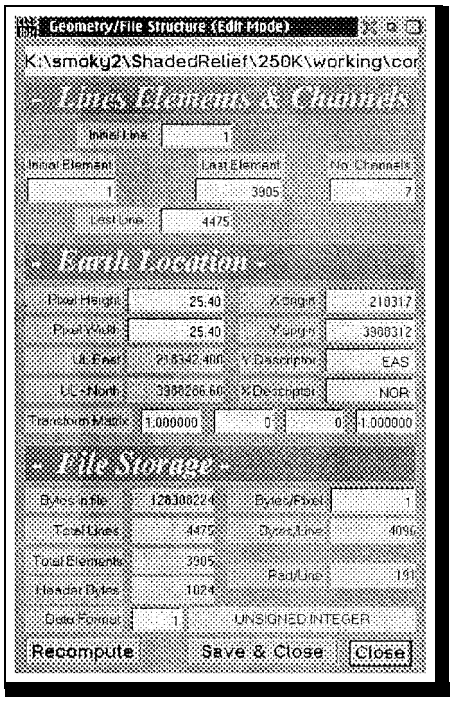


Figure 2. Geometry, file size and Earth location header information.

Another advantaged gained by use of named pipes is an ability to distribute tasks across a network. Named pipes can include a machine ID. Thus it becomes practical to set up a system where large tasks can utilize multiple machines.

Our work is being done under OS/2. This is being done in order take advantage of the operating system's advanced object oriented interface, the native scripting language, and its robust multitasking, multithreading, task preemption and prioritizing nature. Currently no other OS offers the functionality at the cost with ease of use of OS/2. At the user level this will have several obvious benefits. For example the user will be able to open multiple copies of ELAS II, or the modules or the controlling routines and maintain independence. And since ELAS is designed to have a very small memory "footprint" even modest machines will have plenty of RAM available for this.

OS/2 has adopted as a native "scripting language" REXX which we will use in several ways. While REXX is not compiled it is tokenized on execution and can run as fast as native C code in many cases. The senior author has demonstrated this with a real world reformatting operation he frequently uses. It is also a language designed for users that do not do a lot of programming but do want the ability to "batch" jobs while not being limited to just simple tasks. Finally, there are multiple vendors that offer tools that allow a user to build GUIs using REXX. The applications shown in the paper are all done using such a package, VisProRexx. These packages utilize the OS/2 object oriented interface, and several can be programmed by "drag and drop" techniques, a great benefit to a non-programmer. And as the language is not compiled, it is trivial to make the source available and editable to others.

One of our uses of REXX is to build the intelligent GUIs. By use of such inexpensive commercial tools as VisPro we can rapidly build interfaces that accomplish what we need, while letting other customize them to their needs. And for those who do not wish to pay even minimal fees the tools we build are license free and others can be built using various public domain tools.

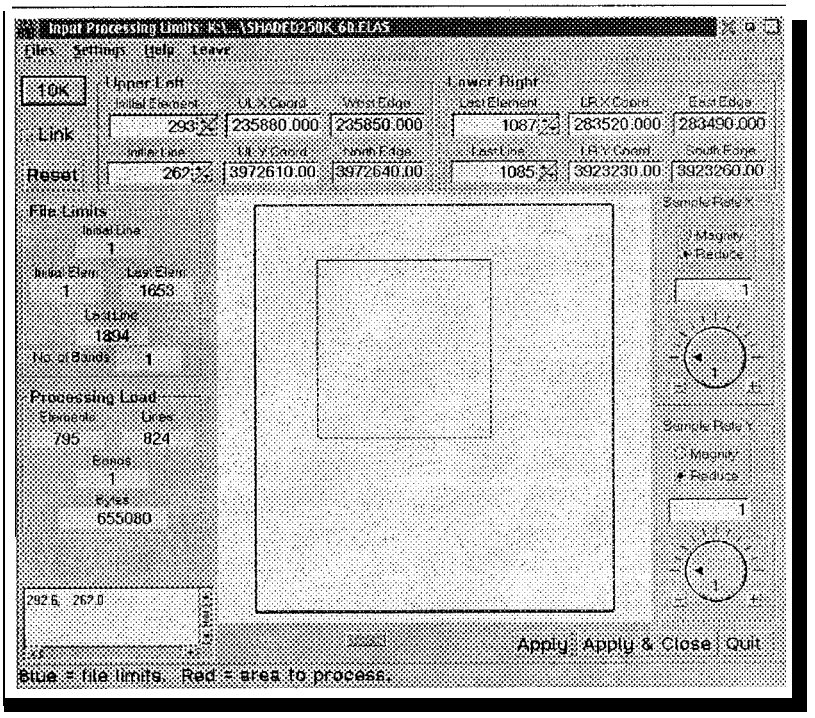


Figure 3. Panel for selection of spatial area to process.

We will also be using REXX as a completely new scripting language, replacing the older XTIO built into ELAS. REXX will allow us complete access to all the operating system functions, the functionality built into the basic REXX system and to all the many REXX extensions commercially available and in the public domain. In fact our approach will allow recursive execution of ELAS II, i.e. an ELAS II script can start either another script or an other ELAS II task.

Finally, we will use REXX to replace and expand on the existing user programmable module DBAS. This module was designed to let users write their own code which would operate on the values stored in a single data file. While other packages have now added similar functionality

DBAS is still more powerful than most. By replacing DBAS with a REXX based replacement we will greatly expand its power, ease the difficulty of coding or the user and probably increase its speed. Beyond replacing the existing code we will create a whole new operation that will allow the user to access multiple files and process arrays. The latter is something that has been requested by ELAS users for more than a decade, within our new design it is at last both doable and practicable.

Finally, the combination of named pipes and REXX will allow a user to create processing flows that call upon multiple modules and weld them into a single new process. This new process can in turn be used by others and as part of yet another process.

The existing massive printed manual will be placed within the standard OS/2 help environment. The programmers manual and the User guide will be included. Use of the OS/2 environment will allow use of color illustrations and cross references. It will also allow full text searching, rather than the much more limited key word searching used on other platforms. It has been the author's experience that when one is trying to learn a new package, trying to guess what the programmer considers a keyword is extremely frustrating and an impediment to learning.

Currently (mid Dec. 1997) the existing command line ELAS has successful y been transferred from the Silicon Graphics to OS/2. Referred to as OS/2 ELAS, this code has been undergoing testing for several months and will be publicly available from the authors shortly after the new year. The work on ELAS II has begun and prototypes can already be demonstrated as seen in the figures. It is our intention to also make it publicly available when appropriate. It is our hope that by presenting this paper that others will be explore the package, adopt it and assist in its continued development. The work has been designed such that most of it does not require deep programming knowledge or ELAS experience. If we are able to share the work load, others will get the pleasure of building a part of a major image processing package and will in turn assist us by distributing the development burden.

Preliminary Work for Examining the Scalability of Reinforcement Learning

Jeff Clouse

NASA Center for Autonomous Control Engineering
and
Department of Computer Science
North Carolina A&T State University
Greensboro, NC 27411

Introduction

Researchers began studying automated agents that learn to perform multiple-step tasks early in the history of artificial intelligence (Samuel, 1963; Samuel, 1967; Waterman, 1970; Fikes, Hart & Nilsonn, 1972). Multiple-step tasks are tasks that can only be solved via a sequence of decisions, such as control problems, robotics problems, classic problem-solving, and game-playing. The objective of agents attempting to learn such tasks is to use the resources they have available in order to become more proficient at the tasks. In particular, each agent attempts to develop a good policy, a mapping from states to actions, that allows it to select actions that optimize a measure of its performance on the task; for example, reducing the number of steps necessary to complete the task successfully.

Our study focuses on reinforcement learning, a set of learning techniques where the learner performs trial-and-error experiments in the task and adapts its policy based on the outcome of those experiments. Much of the work in reinforcement learning has focused on a particular, simple representation, where every problem state is represented explicitly in a table, and associated with each state are the actions that can be chosen in that state. A major advantage of this table lookup representation is that one can prove that certain reinforcement learning techniques will develop an optimal policy for the current task. The drawback is that the representation limits the application of reinforcement learning to multiple-step tasks with relatively small state-spaces.

There has been a little theoretical work that proves that convergence to optimal solutions can be obtained when using generalization structures, but the structures are quite simple. The theory says little about complex structures, such as multi-layer, feedforward artificial neural networks (Rumelhart & McClelland, 1986), but empirical results indicate that the use of reinforcement learning with such structures is promising. These empirical results make no theoretical claims, nor compare the policies produced to optimal policies.

A goal of our work is to be able to make the comparison between an optimal policy and one stored in an artificial neural network. A difficulty of performing such a study is finding a multiple-step task that is small enough that one can find an optimal policy using table lookup, yet large enough that, for practical purposes, an artificial neural network is really required. We have identified a limited form of the game **OTHELLO** as satisfying these requirements. The work we report here is in the very preliminary stages of research, but this paper provides background for the problem being studied and a description of our initial approach to examining the problem.

In the remainder of this paper, we first describe reinforcement learning in more detail. Next, we present the game OTHELLO. Finally we argue that a restricted form of the game meets the requirements of our study, and describe our preliminary approach to finding an optimal solution to the problem.

Reinforcement Learning

Reinforcement learning is a set of learning techniques in which the learner adapts its policy based on the consequences of its own actions. Figure 1 shows the major components of the reinforcement learning scenario: a *learner,* and a *task.* At each time step in learning to perform the task, the learner selects an action based on the current state, performs the action, and then receives a scalar reward that indicates how well the learner is performing. Negative valued rewards are assigned to undesirable actions, positive values to desirable actions, and zero is the reward for all other actions. As an example, consider a mobile robot that is facing a wall. After examining its options, the robot chooses to move forward, only to receive a negative reward for having run into the wall. Reinforcement learning allows the learner to develop an action-selection policy that optimizes a measure of the rewards received. In particular, the learner can learn to maximize the expected return, which is a weighted sum of the rewards received over the course of performing the task. For an in-depth look at reinforcement learning, see Kaelbling, Littman, & Moore (1996).

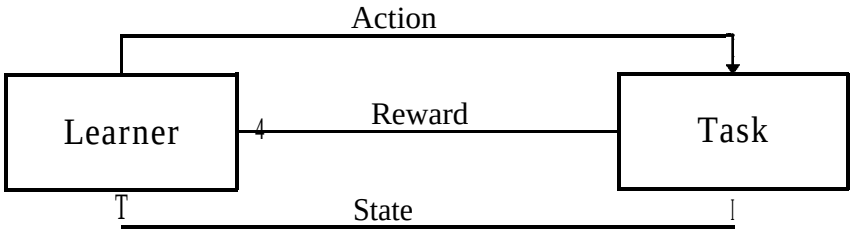


Figure 1: Reinforcement Learning

Reinforcement learning has received the attention of theorists and, thus, there exist many proofs that the application of certain reinforcement learning approaches will lead to optimal solutions (Watkins & Dayan, 1992; Dayan & Sejnowski, 1994; Jaakola, Jordan, & Singh, 1994; Tsitsiklis, 1994), due mostly to the links reinforcement learning has with dynamic programming (Bellman, 1957), the classical control engineering solution method. Given that the policy is represented in a table, and assuming that every action and every state is experienced an infinite number of times, there are several reinforcement learning techniques that allow the policy to converge to optimality. Although the assumptions of the proofs can be rather restrictive, in practice table lookup representations allow the optimal policies to be found.

Reinforcement learning has found success in several areas, including game-playing (Samuel, 1963; Tesauro, 1995), and robotics, control, and maze tasks (Barto, Sutton, & Anderson, 1983; Anderson, 1989; Mahadaven & Connell, 1992; Singh, 1994; Barto, Bradtke, & Singh, 1995; Crites & Barto, 1996). In fact, two of the most recent successes have used the artificial neural network representation for the policy. Tesauro's Neurogammon (1995), which plays backgammon, is considered the best computerized backgammon player in the world, and is

often compared to the top human players. Crites' and Barto's Elevator Control System (1996) performs better at certain types of elevator scheduling problems than solutions that people have been able to develop in the many decades of elevator use. Although both of these systems, and others, perform well, there is no indication of how close to optimal their control policies are. It is not known whether achieving optimal is possible for a complex artificial neural network.

OTHELLO[1]

The game of OTHELLO is a two player, perfect information, zero sum game played on a board marked off in an 8x8 grid. The game starts with an initial configuration of two white and two black discs in particular grid squares and progresses as each player takes turns placing their discs on the board. When a player places a disc on the board, some of the opponent's discs change ownership; for example, when the player playing white places a white disc on the board, a number of black discs are flipped over to their white side. The location and number of discs that flip depend on the location of the discs on the board, and on the placement of the new disc. The objective of the game is to be the player that owns the most discs at the end of the game.

Many researchers have employed Othello for studying various issues in search and learning (Buro, 1995; Lee & Mahajan, 1990, 1988; Rosenbloom 1982) because it is a difficult task to tackle. Each of the systems that were developed played quite well, reportedly at the "world-championship-level ." However, none of the researchers were able to make any claims about optimality, something we would like to see changed. It is also the case that none of the programs employed reinforcement learning in their development, but relied instead on statistical or ad hoc methods to achieve the advanced level of play.

Because OTHELLO is a perfect information, zero sum game, there exists an optimal strategy for placing the discs on the board. The difficulty is that the problem-space for the game is so large that the optimal policy cannot be represented explicitly. Another game for which optimal play exists is tic-tac-toe, which has a small enough problem-space that one can actually determine the optimal policy be developing the entire game tree, which is a representation of every possible move from every possible state. Even so, tic-tic-toe has over 19,000 states ($3^{3\times3}$) if one ignores symmetries. Even for a simple game like tic-tac-toe, developing the optimal solution can be impractical.

There are sixty-four possible locations for discs on the OTHELLO board, and three possible contents for each location (blank, white disc, and black disc); thus, there are at most $3^{8\times8} \approx 3.4 \times 10^{30}$ board configurations. Because many of these boards can not be reached by following the legal rules of the game (for example, there is no way to have a board with only a white disc in the top right corner) the actual number of legal boards is much smaller. Even if we assume that only one board out of a million possible boards is legal, we are still left with over 3 x 1024 boards, for which we would need over 3 x 1015 gigabytes of storage, assuming that one board can be represented in only one byte. So, even though we know that optimal play exists, we do not have the resources to develop the optimal policy via a table representation.

Our Approach

[1] OTHELLO is a registered trademark of Tsukuda Original, licensed by Anjar Company

We are attempting to determine whether the use of multi-layer, feedforward artificial neural networks will allow reinforcement learning to develop optimal policies. In order to do so, we need to find a problem that is large enough to require artificial neural networks to store the policy in practice, but small enough that an optimal policy can be developed through careful planning, and memory and disk usage. The full game of OTHELLO does not meet these requirements, however. The game is much too large to store the entire policy in a table. So, we are focusing instead on two simplifications of the game. Our first choice is to reduce the board size from an 8x8 board to a 6x6 board. This will reduce the number of board states by several orders of magnitude when compared to the original game: from at most $3^{8\times8} \approx 3.4 \times 1030$ to only $3^{6\times6} \approx 1.5 \times 1017$, which is still a rather large number of game boards. Reducing the game further, to only a 4x4 board reduces the problem too far. Furthermore, it may be that learning about the 6x6 game will not tell us anything about the larger game.

For this reason, our second choice is to examine the end-game of the full game: start with boards that have thirty to twenty-five moves left. This will reduce the size of the problem we are interested in while at the same time giving us information about how to play the original game well. The difficult here is in identifying legal boards that have the appropriate number of moves left, since one must begin at the beginning of the game and follow the progress of the legal moves in order to develop legal boards. Instead of attempting to produce all legal boards with the requisite number of moves left, we will produce a large number of them, say, a few million.

In addition to setting up the problem, our work has also been on examining techniques such as hash tables that may allow us to store and retrieve every game board and an associated value that is used in calculating the policy. The practical knowledge we expect to gain is a better understanding of the multiple-step task known as the game of OTHELLO. The scientific knowledge we expect to gain relates the use of reinforcement learning, which we know can produce optimal policies for small problems, to larger problems. We may learn that we can develop optimal, or near-optimal, solutions with the marriage of reinforcement learning and multi-layer feedforward neural networks. But we may find that the neural network approach, at least in our implementation, does not produce near-optimal policies, even though it performs well in practice.

Acknowledgments

This material is based upon work supported by the National Aeronautics and Space Administration, via grant ACE-48 146 to the NASA Center for Autonomous Control Engineering at North Carolina A&T State University, Greensboro, NC, and via grant NAG4- 132 to Jeff Clouse.

IV. References

Anderson, C. W. (1989). Learning to control an inverted pendulum using neural networks. *IEEE Control Systems Magazine, 9,31-37.*

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics, 13,835-846.*

Barto, A. G., Bradtke, S. J., & Singh, S. P. (1995). Learning to act using real-time dynamic programming. *Artificial Intelligence, 72, 81-138.*

Bellman, R.E. (1957). *Dynamic Programming.* Princeton, NJ: Princeton University Press.

Buro, M. (1995). Statistical feature combination for the evaluation of game positions. *Journal of Artificial Intelligence Research, 3,373-382.*

Crites, R. H., & Barto, A. G. (1996). Improving elevator performance using reinforcement learning. *Advances in New-al Information Processing Systems.* Cambridge, MA: MIT Press.

Dayan, P. D., & Sejnowski, T. J. (1994). TD(λ) converges with probability 1. *Machine Learning, 14, 295-301.*

Fikes, R. E., Hart, P. E., & Nilsson, N. J. (1972). Learning and executing generalized robot plans. *Artificial Intelligence, 3,* 251-288.

Jaakola, T., Jordan, M. I., & Singh, S. P. (1994). On the convergence of stochastic iterative dynamic programming algorithms. *Neural Computation, 6.*

Kaelbling, L.P., Littman, M.L., & Moore, A.W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4,237-285.*

Lee, K-F & Mahajan, S. (1988). A pattern classification approach to evaluation function learning. *Artificial Intelligence, 36, 1-25..*

Lee, K-F & Mahajan, S. (1990). The development of a world class Othello program. *Artificial Intelligence, 43,21-36.*

Mitchell, D. (1984) *Using features to evaluate positions in experts' and novices' Othello games,* (Masters Thesis), Evanston, IL: Department of Psychology, Northwestern University.

Rosenbloom, P. (1982). A world-championship-level Othello program. *Artificial Intelligence, 19, 279-320.*

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing.* Cambridge, MA: MIT Press.

Samuel, A. (1963). Some studies in machine learning using the game of Checkers. In Feigenbaum & Feldman (Eds.), *Computers and Thought. New* York McGraw- Hill.

Samuel, A. (1967). Some studies in machine learning using the game of Checkers II: Recent progress. *IBM Journal of Research and Development,* 11,601-617.

Singh, S. P. (1994). *Learning to solve Markovian decision processes.* Doctoral dissertation, Department of Computer Science, University of Massachusetts, Amherst, MA.

Tesauro, G. (1995). Temporal different learning and TD-Gammon. *Communications of the ACM, 38, 58-68.*

Tsitsiklis, J. N. (1994). Asynchronous stochastic approximation and Q-Learning. *Machine Learning, 16, 185-202.*

Waterman, D. A. (1970). Generalization learning techniques for automating the learning of heuristics. *Artificial Intelligence, 1, 121-170.*

Watkins, C.J.C.H., & Dayan, P. (1992). Q-Learning. *Machine Learning, 8, 279-292.*

DYNAMICS AND MECHANISMS OF $O(^3P)$ ATOM REACTIONS WITH SULFUR-CONTAINING COMPOUNDS

K. Ravichandran, Xirong Chen, Fei Wu, Yuchuan Gong, Yuying Cheng, Jiande Han, Idelisa Ayala and Brad R. Weiner*

Department of Chemistry, University of Puerto Rico, P. O. Box 23346, UPR Station, Rio Piedras, Puerto Rico 00931

1. INTRODUCTION

Reactions of atoms and radicals play a central role in controlling the overall chemistry of the atmosphere and in combustion systems. Our understanding of these complex chemical environments has resulted from detailed kinetics and dynamics studies of elementary reactions that are involved in these systems. Interest in the reactions of oxygen atoms (^1D and ^3P) with hydrocarbons and hydrogen containing molecules are two fold: (1) $O(^1D)$ reactions are known to produce hydroxyl (OH) radicals that determine in part, the ozone concentration level in the stratosphere and (2) $O(^3P)$ reactions with unsaturated hydrocarbons, both olefins and alkynes, play an important role in combustion chemistry. Oxidation of naturally occurring reduced sulfur compounds, such as hydrogen sulfide (H_2S), carbonyl sulfide (OCS), and carbon disulfide (CS_2) by oxygen atoms, initiate a chain of reactions in the troposphere that produce sulfur dioxide (SO_2), and ultimately result in acid precipitation [1,2] Previous kinetics measurements of these systems are in good agreement, but the mechanisms are complex, and perhaps even dependent upon the conditions of the experiments.[3] Therefore, considerable uncertainty still exists in the mechanistic interpretation of these elementary reactions.

Mechanistic details, such as the transition state structures, possible intermediates, and the nature of energy flow due to atomic rearrangements, can be revealed for a chemical reaction by measuring the nascent energy disposed to the products [3]. Energy partitioned to the products' vibrational degrees of freedom, particularly to the newly formed bonds, provides a wealth of information on the reaction mechanism. In addition, product rotational energy disposal studies can give a handle on the structure of reaction intermediates and transition states.

We report here our detailed dynamics study on the following $O(^3P)$ atom reactions with sulfur compounds from the nascent $SO(X^3\Sigma^-)$ product rovibrational energy distributions:

		ΔH_{rxn} (kcal/mol)	
$O(^3P) + OCS \longrightarrow SO(X^3\Sigma^-) + CO(X^1\Sigma^+)$		-50.5	(1)
$+ CS_2 \longrightarrow SO(X^3\Sigma^-) + CS(X^1\Sigma^+)$		-21	(2)
$+ C_2H_4S \longrightarrow SO(X^3\Sigma^-) + C_2H_4(^1A_g)$		-63	(3)
$+ CSCl_2 \longrightarrow SO(X^3\Sigma^-) + CCl_2(^1A_1)$		-19	(4)

Reaction (1) is the most studied of the above four cases, since carbonyl sulfide (OCS) constitutes the major component of the sulfur reservoir in the troposphere and is also a major greenhouse gas.[4,5] OCS is mostly unreactive in the troposphere and therefore reaches the stratosphere, where it undergoes UV photolysis and/or reacts with other transient species, including oxygen atoms. As a result, reaction (1) plays a major role in the global sulfur budget and radiation balance.

* Author to whom correspondence should be addressed

Both kinetics and dynamics studies have been carried out on reaction (1) by several groups in the past.'> [6-11] Kinetics studies reveal an activation barrier of 4.4 kcal/mol from rate constant measurements over the temperature range of 220-600 K. Several groups have studied the dynamics of reaction (1) by measuring the energy disposed to the CO product. Lin and co-workers found 11% of the available energy of the reaction deposited to the CO vibrational degree of freedom, whereas Nickolaisen *et al.* under better defined initial conditions account for only O .5°/0. Reactive scattering studies by Grice and coworkers reports the fraction of total available energy disposed into products' translation as a function of collision energy. None of the above dynamics studies investigated the energy disposed into the $SO(X^3\Sigma^-)$ product for reaction (1). Our recent $SO(X^3\Sigma^-)$ energy disposal studies from reaction (1) finds 26% of the available energy partitioned into the vibrational and 8.10/0 into the rotational degrees of freedom.[12] This large amount of energy disposed to the $SO(X^3\Sigma^-)$ internal degrees of freedom and an inverted vibrational state distribution for $SO(X^3\Sigma^-)$ suggests a direct S atom abstraction mechanism for this reaction which is consistent with the previous reports.

The reaction of $O(^3P)$ with CS_2 is important in the atmospheric sulfur cycle and in CS_2 chemical lasers.[13,14] Smith reported 18% and 8% of the available energy into the SO and CS vibrational degrees of freedom.[15] Previous studies on this reaction have focussed on the measurement of product branching ratios since CO + S_2 and OCS and S are the other reaction channels for these reactants[10,16-21] These previous studies report 110/0 for the other two channels, leaving the $SO(X^3\Sigma^-) + CS(X^1\Sigma^+)$ to be the dominant chemical pathway. Crossed molecular beam studies on reaction (2) show that this channel proceeds via a direct stripping type mechanism .22-24 Our current $SO(X^3\Sigma^-)$ and $CS(X^1\Sigma^+)$ energy disposal studies provide greater insight into the proposed mechanism for this reaction.

Lee *et al.*, measured the rate constant for reaction (3) between 268 and 424 K temperature range.[25] The measured rate constants at different temperatures were fit to the Arrhenius expression k = (2.87 ± O. 12) X 10^{-11} exp (-1880+ 24/1.987 T) cm^3 molecule-l s^{-1}. Due to the large rate constant and near zero activation energy, a direct abstraction of S from C_2H_4S was proposed for this reaction. Our recent $SO(X^3\Sigma^-)$ energy disposal studies on this reaction resulted in a non-Boltzmann vibrational distribution with a maximum population in v" =1 and measurable population detected up to v " = 8.[26] The non-statistical energy distribution in $SO(X^3\Sigma^-)$ is consistent with an abstraction mechanism for reaction (3). The only previous study on reaction (4) is a kinetic measurement by Slagle *et al.*, between 250 and 500 K which resulted in an Arrhenius expression, k = (3 .09 ± 0.54) X 10^{-11} exp (O. 115 ± 0.106 kcal mol^{-1}/RT) cm^3 molecule-l s^{-1}.[27] Due to the strongly polarized C - S bond in Cl_2CS and the electrophilic nature of $O(^3P)$ atom, Slagle *et al.*, proposed an addition mechanism to produce a triplet dichloro-sulfine (Cl_2CSO) adduct as the primary step in this reaction, followed by C-S bond cleavage to give SO and CCl_2 products. This electrophilic addition-fragmentation mechanism is consistent with our current $SO(X^3\Sigma^-)$ energy disposal studies for reaction (4).

2. EXPERIMENTAL

The experimental apparatus and procedure for the $SO(X^3\Sigma^-)$ rovibrational state distribution measurements are described elsewhere,[28-31] and will only be briefly summarized here. Ground state $O(^3P)$ atoms were generated by photolysis of NO_2 either by the 351 nm output from a XeF excimer laser or the 355 nm output from a frequency tripled Nd-YAG laser (Continuum, model PL7000) with typical laser fluences of 10-70 mJ/cm^2. $SO(X^3\Sigma^-)$ product from reactions (1) to (4) are monitored by measuring the laser-induced fluorescence (LIF) signal on the $B^3\Sigma^- - X^3\Sigma^-$ transition in the wavelength region of 237 -312 nm. The probe laser is pumped by a XeCl excimer laser operating at 308 nm. The $SO(X^3\Sigma^-)$ excitation wavelengths (237 - 312 nm) are generated by frequency doubling the dye laser fundamental outputs using five different laser dyes (coumarin 480, coumarin 503, coumarin 540A, rhodamine B and rhodamine 6G). Both the photolysis and the probe lasers are collinearly counter-propagated along the extension arm axis of the reaction cell with a time delay of 2-5 μs between the laser pulses using a pulse delay generator. LIF signals were collected

at 90° relative to the laser beam axis by a high gain photomultiplier tube (PMT) through several long-pass filters to minimize the scattered light from the photolysis and probe lasers. The output of the PMT is processed and averaged by a gated integrator, digitized, and stored in a computer for further analysis. For a fixed time delay between the photolysis and the probe lasers, the dye laser was scanned while collecting the total fluorescence to obtain the nascent LIF spectrum.

The gases used, OCS (97.5%) and NO_2 (98%), were obtained commercially (Matheson) and used without further purification, whereas CS_2 (Allied chemicals), C_2H_4S (98°/0, Aldrich), and Cl_2CS(97%, Aldrich) were subjected to several freeze-pump-thaw cycles prior to use. Typically, 20-50 mTorr of NO_2 and 100-200 mTorr of the sulfur containing reactants were used in the current study. The reaction cell was evacuated to 10^{-6} Torr prior to the experiments by a turbomolecular pump, but the gases were flowed by a mechanical pump during the experiments and the pressures were monitored at the exit by a capacitance manometer.

3. Results and Discussion

LIF signals, assignable to the $SO(B^3\Sigma^- - X^3\Sigma^-)$ transition,[32,33] have been obtained under two collision conditions for all the above reactions, where the total pressure in the reaction cell is 100mTorr and the probe delay time is 2 μs. Since vibrational relaxation occurs with more collisions, typically once every 1000 collisions or so, the measured vibrational state distributions for $SO(X^3\Sigma^-)$ from reactions (1) - (4) can be considered as nascent. A typical spectrum for $SO(B^3\Sigma^-, v' = 1 - X^3\Sigma^-, v'' = 6)$ from reaction (1) under two gas kinetic collisions is shown in Fig 1. The $SO(X^3\Sigma^-)$ vibrational state distributions from reaction (1) - (4) were obtained by integrating the area of the LIF spectra for the vibronic transitions terminating on the v' = 1 level of the B state (1, v" = O -9 (depending on the maximum detectable v")), and normalizing for their respective Franck - Condon (FC) factors.[34] The energy disposal into the $SO(X^3\Sigma^-)$ can be obtained from the sum of energy partitioned into each vibrational level: $E_{vib} = \sum c_v (E_v - E_0)$, where c_v is the fractional population in each v" state given in Fig (2), and E_0 is the zero point energy of $SO(X^3\Sigma^-)$.

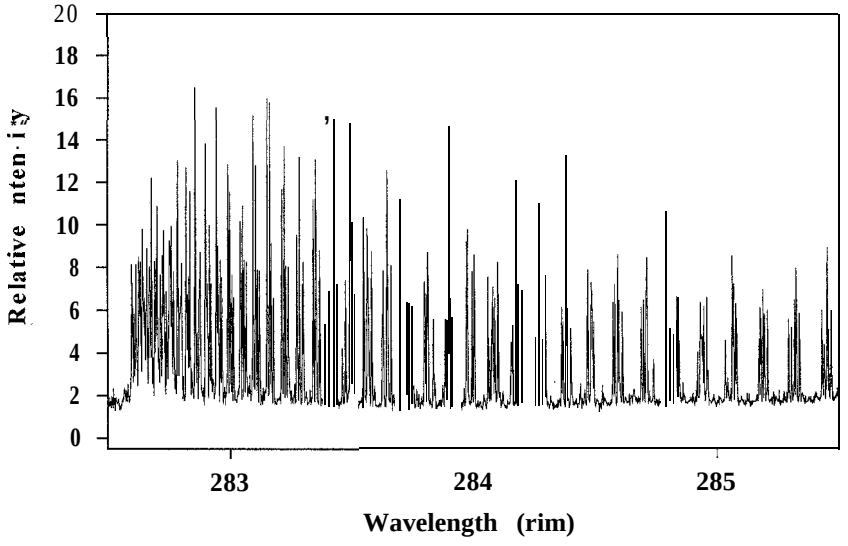


Figure 1. LIF excitation spectrum of SO ($B^3\Sigma^-, v' = 1 - X^3$ X-, v" = 6) transition measured from reaction (1) after two gas kinetic collisions.

The procedure for measuring the $SO(X^3\Sigma^-)$ rotational state distribution is given elsewhere[28-31] and will be discussed only briefly here. The rotational distribution in a single vibrational level can be given by the following Boltzmann expression,

$$P(N'') = (2J''+1)\exp[-B_v N''(N''+1) / k_B T_R] \quad (5)$$

where B_v is the rotational constant of the given vibrational state, k_B is the Boltzmann constant and T_R is the rotational temperature. If a plot of $\ln[P(N'')/2J''+1)]$ versus $BN''(N''+1)$ can be fit to a straight line, then the slope gives the rotational temperature T_R from which the average rotational energy for a given vibrational level is calculated from the expression $k_B T_R$. The following sections will discuss the total rovibrational energy disposed to $SO(X^3\Sigma^-)$ product and the possible mechanisms for reactions (1) - (4).

3.1. $O(^3P) + OCS \rightarrow SO(X^3\Sigma^-) + CO(X^1\Sigma^+)$

The observed vibrational state distribution of $SO(X^3\Sigma^-)$ from this reaction is shown in Fig 2a. and was found to be inverted in v" = 5 with detectable population up to v " = 9. The total vibrational energy of near nascent $SO(X^3\Sigma^-)$ was calculated to be 14.4 kcal mol^{-1}, corresponding to ≈ 26% of the total available energy which includes the center of mass (cm.) kinetic energy of $O(^3P)$ from 351 nm photodissociation of NO_2. Previous results proposed an abstraction mechanism for this reaction with the sulfur end of OCS being the reactive site. Our measured inverted vibrational distribution along with a significant energy disposed into $SO(X^3\Sigma^-)$ vibrations supports an abstraction mechanism or a very short-lived OCSO collisional complex. From our $SO(X^3\Sigma^-)$ rotational energy which was found to be 8.1 % of the available energy, and in conjunction with previous work, greater than 90°/0 of the reaction exoergicity has been measured in the products' degrees of freedom for this reaction.

3.2. $O(^3P) + CS_2 \rightarrow SO(X^3\Sigma^-) + CS(X^1\Sigma^+)$

The measured $SO(X^3\Sigma^-)$ vibrational state distribution from reaction (2) is shown in Fig 2b. It is clear that the distribution is non-Boltzmann with an inversion in v" = 1. From our measured vibrational distribution for $SO(X^3\Sigma^-)$, we find 22% of the available energy disposed into SO vibration which is consistent with previous measurements. Also an inversion in v" = 1 for $SO(X^3\Sigma^-)$ supports the abstraction mechanism for this reaction. An analysis of the $CS(X^1\Sigma^+)$ rovibrational energy distribution along with the $SO(X^3\Sigma^-)$ rotational energy for this reaction is currently underway.

3.3. $O(^3P) + C_2H_4S \rightarrow SO(X^3\Sigma^-) + C_2H_4(^1A_g)$

The experimentally determined $SO(X^3\Sigma^-)$ vibrational state distribution under two collision condition is shown in Fig 2c. As can be seen, the vibrational state distribution is inverted with a maximum in v" = 1, with a detectable population up to v" = 8 and the total vibrational energy for $SO(X^3\Sigma^-)$ was found to be 13% of the available energy. From the measured $SO(X^3\Sigma^-)$ rotational state distribution we find 3°/0 of the available energy in SO rotations. Based on the above results, an abstraction mechanism that occurs by the attack of $O(^3P)$ on the sulfur end of the C_2H_4S leading to $SO(X^3\Sigma^-)$ and $C_2H_4(^1A_g)$ as the primary products, is proposed for reaction (3).

3.4. $O(^3P) + Cl_2CS \rightarrow SO(X^3\Sigma^-) + CCl_2(^1A_1)$

The $SO(X^3\Sigma^-)$ vibrational state distribution under two collision condition is shown in Fig 2d. It is clear that the vibrational distribution for $SO(X^3\Sigma^-)$ in this reaction can be characterized as Boltzmann with a vibrational temperature of 1150 ± 55 K. As a result of a Boltzmann type vibrational distribution for the $SO(X^3\Sigma^-)$ product, we expect the reaction to proceed via a long-lived intermediate and all modes of this

intermediate to participate equally in product formation. We favor the triplet dichloro-sulfine (Cl_2CSO) intermediate, proposed by Slagle et *al.*, as the primary process for this reaction.[35] This triplet Cl_2CSO

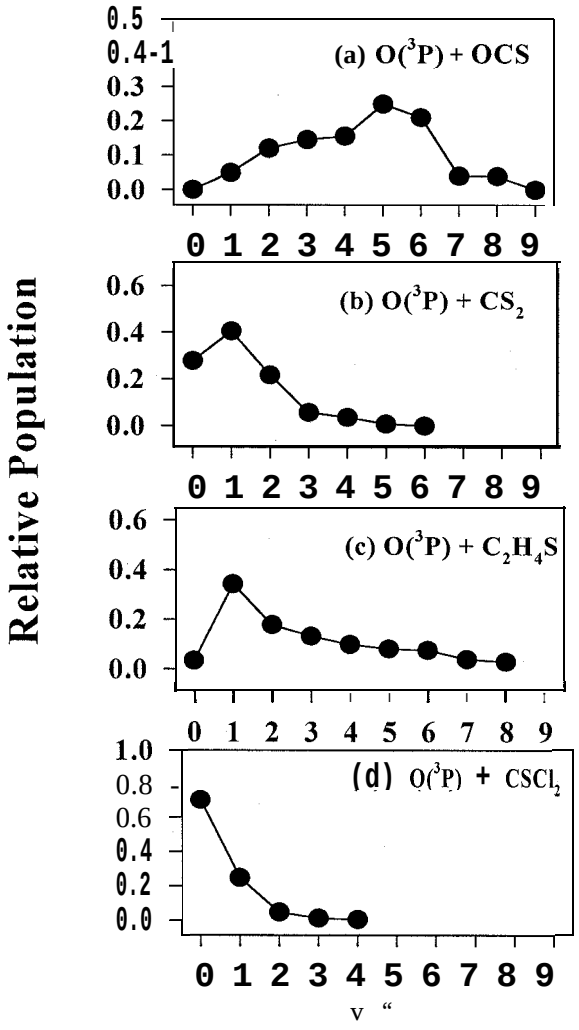


Figure 2. Observed vibrational state distributions of the $SO(X^3\Sigma^-)$ product following reaction (1) - (4).

intermediate undergoes a C - S bond clevage to give $SO(X^3\Sigma^-)$ and $CCl_2(^1A_1)$ products. The nascent vibrational and rotational energies were calculated to be 1.3 and 1 kcal mol^{-1} respectively, which corresponds to 5.1 % and 4% of the available energy.

4. Summary and Conclusions

$SO(X^3\Sigma^-)$ product rovibrational energy distribution has been measured from the reactions of $O(^3P)$ with OCS, CS_2, C_2H_4S and Cl_2CS by using LIF spectroscopy. The energy distributions has provided detailed mechanistic information for these reactions. Non-Boltzmann vibrational distribution for the $SO(X^3\Sigma^-)$ product in the case of $O(^3P)$ reactions with OCS, CS_2 and C_2H_4S suggest an abstraction mechanism, complementing the previous measurements made on these reactions. In the case of Cl_2CS, the reaction

proceeds via a bound dichloro-sulfine (Cl_2CSO) intermediate that statistically partitions the energy among all the vibrational modes, resulting in a Boltzmann type vibrational distribution for the $SO(X^3\Sigma^-)$ product.

5. REFERENCES

1. M. Nicolet, *Can. J. Chem. 52, 1381, 1974.*
2. G.S. Tyndall and A.R. Ravishankara, *Int. J. Chem. Kinet. 23, 483, 1991* and references therein.
3. W.B. DeMore, D.M. Golden, R.F. Hampson, C.J. Howard, M.J. Kurylo, M.J. Molina, A.R. Ravishankara and S.P. Sander, *JPL Publication 94-26,* December 15, 1994 and references therein.
4. P. Warneck, *Chemistry of the Natural Atmosphere",* Academic Press, New York, 1988.
5. R.P. Wayne, *Chemistry of Atmospheres,* Oxford Univ. Press, Oxford, 1991.
6. J.S. Robertshaw and I.W.M. Smith, *Inter. J. Chem. Kinet. 12,729, 1980.*
7. B. Kruger and H.G. Wagner, Z. *Physik. Chem. 126,* 1, **1981.**
8. R. Atkinson, D.L. Baulch, R.A. Cox, R.F. Hampson Jr., J.A. Kerr, and J. Tree, *J. Phys. Chem. Ref. Data 18, 881, 1989.*
9. R.G. Shortridge and M.C. Lin, *Chem. Phys. Lett. 3.5, 146, 1975.*
10. D.S.Y. Hsu, W.M. Shaub, T.L. Burks and M.C. Lin, *Chem. Phys.* 44, 143, 1979.
11. S.L. Nickolaisen, D.W. Veney and H.E. Cartland, *J. Chem. Phys. 100,4925, 1994.*
12. *X.* Chen, F. Wu, and B.R. Weiner, *Chem. Phys. Lett. 247,313, 1995.*
13. B.M.R. Jones, R.A. Cox, and S.A. Penkett, J. *Atmos. Chem. 1,65, 1983.*
14. S.J. Arnold, and H. Rojeska, *Appl. Opt. 12, 169, 1973.*
15. I.W.M. Smith, *Discussions Faraday Sot. 44, 194, 1967.*
16. *G.* Hancock, and I.W.M. Smith, *Trans. Faraday Sot. 67,2586, 1971.*
17. R. D. Suart, P.H. Dawson, and G.H. Kimbell, *J. Appl. Phys.* 43, 1022, 1972.
18. I.R. Slagle, J.R. Gilbert, and D. Gutman, *J. Chem. Phys. 61,704, 1974.*
19. R.E. Graham, and D. Gutman, *J. Phys. Chem. 81,207, 1977.*
20. J.W. Hudgens, J.T. Gleaves, and J.D. McDonald, *J. Chem. Phys. 64,2528, 1976.*
21. P.D. Naik, U.B. Pavanaja, A.V. Sapre, K.V. S. Ramarao, and J.P. Mittal, *Chem. Phys. Lett. 186,565, 1991.*
22. J. Geddes, P.N. Clough, and P.L. Moore, *J. Chem. Phys.* 61,2145, 1974.
23. P.A. Gerry, C.V. Nowikow, and R. Grice, *Mol. Phys. 37,329, 1979.*
24. J.J. Rochford, L.J. Powell, and R. Grice, *J. Phys. Chem. 99, 15369, 1995.*
25. J.H. Lee, R.B. Timmons and L.J. Stief, *J. Chem. Phys. 64,300,.1976.*
26. K. Ravichandran, Y. Gong, F. Wu and B.R. Weiner, *Chem. Phys. Lett. 252,348, 1996.*
27. I.R. Slagle and D. Gutman, *Int. J. Chem. Kinet. 11,453, 1979.*
28. *X.* Chen, F. Asmar, H. Wang and B.R. Weiner, J. *Phys. Chem.* 95,6415, 1991.
29. H. Wang, X. Chen, and B.R. Weiner, *J. Phys. Chem. 97, 12260, 1993.*
30. *X.* Chen, H. Wang, B.R. Weiner, M. Hawley and H.H. Nelson, J. *Phys. Chem.* 97, 12269, 1993.
31. H. Wang, X. Chen and B.R. Weiner, *Chem. Phys. Lett. 216,537, 1993.*
32. R. Colin, *J. Chem. Sot., Faraday Trans. 2, 11, 1139,1982.*
33. R. Colin, *Can. J. Phys. 47,979, 1969.*
34. W.H. Smith, H.S. Liszt, *J. Quan. Spectros. Radiat. Trans. 11,45, 1975.*
35. K. Ravichandran, I. Ayala, Y. Ishikawa and B.R. Weiner, J. *Phys. Chem. 101, 8587,1997.*

98 URC169

Quantity, Revisited: An Object-Oriented Reusable Class

Monica Gayle Funston[1], Walter Gerstle[2], and Malcolm Panthaki[3]
Department of Civil Engineering, University of New Mexico, Albuquerque, NM 87131
Phone: (505)277-3458; Fax: (505) 277-1988; Email: gerstle@unm.edu

Abstract

Quantity, a prototype implementation of an object-oriented class, was developed for two reasons: to help engineers and scientists manipulate the many types of quantities encountered during routine analysis, and to create a reusable software component to for large domain-specific applications. From being used as a stand-alone application to being incorporated into an existing computational mechanics toolkit, *Quantity* appears to be a useful and powerful object.

Quantity has been designed to maintain the full engineering meaning of values with respect to units and coordinate systems. A *Value* is a scalar, vector, tensor, or matrix, each of which is composed o *Value Components,* each of which may be an integer, floating point number, fuzzy number, etc., and its associated physical unit. Operations such as coordinate transformation and arithmetic operations are handled by member functions of *Quantity*. *The* prototype has successfully tested such characteristics as maintaining a numeric value, an associated unit, and an annotation. In this paper we further explore the design o *Quantity*, with particular attention to coordinate systems.

Introduction

Engineers make extensive use of mathematical objects such as scalar, vector, tensor, and matrix. Unlike pure mathematicians, however, engineers attach physical significance to these mathematical objects Thus, unlike pure mathematical quantities, physical units must be attached to the components of engineering quantities Also, because the physical meaning of engineering quantities must be made clear, it is often necessary for engineers to explain, or annotate, quantities A great deal of engineering time and many errors arise in manipulating, displaying, and interpreting engineering quantities A compact, introduction to tensor analysis can be found in (Block 1978). The transformation of tensor components from one coordinate system to another is in general an involved yet precisely defined process: exactly the sort of thing that would be good to hide from the engineer; hence another reason for the development of the object *Quantity*.

During the development of a Computational Mechanics Toolkit (CoMeT) at the University of New Mexico (Atencio et al. 1997), a software component was needed to manipulate engineering quantities. A search was conducted that revealed no reusable component with all the requirements to:

. Store data as values with respect to units and coordinate systems;
. Perform arithmetic operations between quantities;
. Perform unit conversions;
. Perform coordinate transformations;
• Allow annotation of quantities;

[1] Graduate Student
[2] Associate Professor; Corresponding Author
[3] Software Consultant

. Graphically render quantities in a flexible way.

We therefore set forth to design an object-oriented reusable system (Wirfs-Brock, Wilkerson, and Weiner 1990) to address these requirements.

This paper follows upon a previous paper (Funston, Valerie, and Gerstle 1997). In that paper, the concept of *Quantity was* introduced, and the design with respect to units was discussed in detail. However, in our ensuing work, we identified deficiencies in the design, which we now attempt to correct. In this paper, we concentrate upon issues associated with coordinate system transformations and their interactions with unit conversions. The design of *Quantity is* changed in important ways from that described in (Funston, Valerie, and Gerstle 1997).

Redesign of *Quantity* and *Value;* Introduction of *Value Component*

Previously, the *Value* part of a *Quantity was* maintained as a single piece of member data in the various types of *Quantity* classes. Differentiation of *Value* type was all at one level; for example derived classes of *Value* included *Integer Scalar, Float Scalar, Fuzzy Scalar, Integer Vector, Float Vector, Fuzzy Vector, Integer Tensor, Float Tensor,* and *Fuzzy Tensor. The* Unit Symbol was maintained as member data by *Quantity. This* design led to confusion because it is possible for *Values* to have components with different units, as for example, in the case of a position vector in a cylindrical coordinate system.

It has become apparent that we require a class called a *Value Component.* A *Value* now contains as member data a list of *Value Components,* and must collaborate with a *Coordinate System.* A *Value Component can* be of various types: integer, float, fuzzy, etc. Thus, a *Quantity* now maintains a *Value* (of type Scalar, Vector, Tensor, Matrix, etc.), and a *Coordinate System.* A *Value in turn* maintains a list of *Value Components,* and collaborates with *Coordinate System.* A *Value Component* maintains not only a scalar of a specific type, but also a Unit Symbol.

Value is categorized according to the rules of transformation with respect to change of coordinate system. Thus, types of *Value* include *Tensor* (of specified Order and Dimensionality), *Position, Matrix,* etc. A *Matrix* does not transform under change of coordinate system.

A collaboration chart between *Quantity, Value, Value Component, Coordinate System, Coordinate System Dictionary,* and *Unit Dictionary is* shown in Fig. 1.

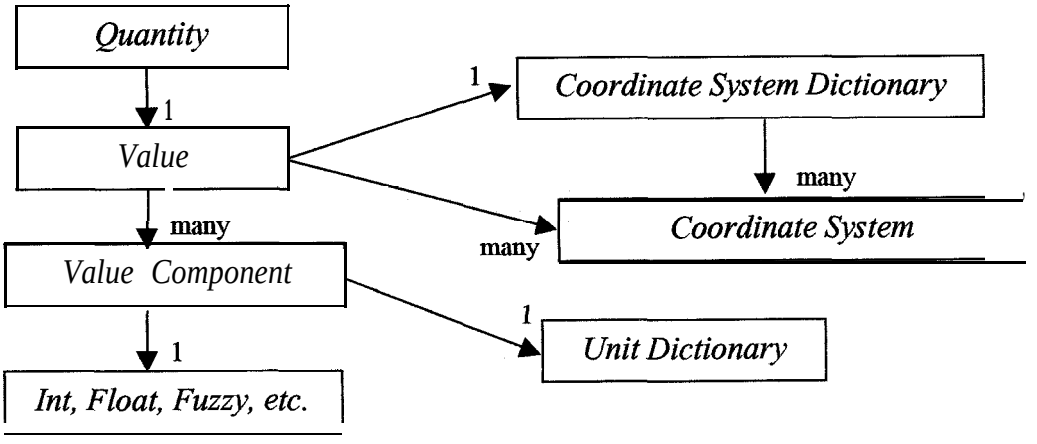


Fig. 1- Collaboration Chart

With this redesign, no changes to *Unit Dictionary* were necessary. However *Value Component,* rather than *Quantity*, now collaborates with *Unit Dictionary* to assign units and perform unit conversions. *Value,* rather than *Quantity*, now collaborates with *Coordinate System Dictionary* to perform coordinate system transformations of its *Value Components.*

Design of *Coordinate System*

We have developed a very general concept of *Coordinate System.* Component data and functions of *Coordinate System* include:

. coordinate system name (A);
. parent **coordinate system name** (P);
. Function to transform from a set of coordinates x_i^A in (A) to coordinates x_j^P in (P);
. Function to transform from a set of coordinates x_j^P in (P) to coordinates x_i^A in (A);
. Function to develop Jacobean transformation between (A) and (P) at a point;
. Function to develop inverse Jacobean transformation between (A) and (P) at a point;
. A list of child coordinate systems.

With these components a *Value,* whether it is a position vector or an n^{th}-order tensor, can be transformed from one *Coordinate System* to another. The formulas for transforming first, second, and third-order tensors from Cartesian coordinate system A to Cartesian coordinate system B are:

$$v_i^A = \frac{\partial x_i^A}{\partial x_j^B} v_j^B ;$$

$$t_{ij}^A = \frac{\partial x_i^A}{\partial x_r^B} \frac{\partial x_j^A}{\partial x_s^B} t_{rs}^B ;$$

$$t_{ijk}^A = \frac{\partial x_i^A}{\partial x_r^B} \frac{\partial x_j^A}{\partial x_s^B} \frac{\partial x_k^A}{\partial x_t^B} t_{rst}^B$$

The **partial** derivatives are elements of the Jacobean transformation. Thus it is necessary for a *Coordinate System* object A to compute derivatives of its coordinates with respect to coordinates of its parent P (elements of the Jacobean matrix), and vice versa' (elements of the inverse of the Jacobean). As is discussed next, for curvilinear coordinate systems, it is necessary to specify the location where the tensor is to be computed.

For linearly related coordinate systems, the elements of the Jacobean are constant with respect to position, while in curvilinear coordinate systems where the mapping is nonlinear the elements of the Jacobean are functions of position. Thus transformations of tensors in linearly-related coordinate systems are independent of position, while transformations of tensors in curvilinear coordinate systems depend upon the location at which the tensor is located. A detailed discussion of tensor transformations in curvilinear coordinates is given in (Block 1978).

Coordinate transformation from coordinate system (A) to coordinate system (B) occurs in the following manner. First the *Value* induces each *Value Component* to convert its unit to the corresponding Base Unit. This ensures that the coordinate transformation is done using consistent units. Then, through a sequence of transformations, the *Value* is transformed from (A) to (B). Finally, each *Quantity Component* is converted from the Base Units to the original units.

The *Coordinate System Dictionary* is modeled after the *Unit Dictionary with* a Base Coordinate System that is the fundamental parent coordinate system of which other Coordinate Systems are children. The *Coordinate System Dictionary* maintains a list of coordinate systems.

A coordinate system cannot be deleted if it is the parent of other coordinate systems. Also, a coordinate system cannot be deleted if a value is defined with respect to it. Thus, it is necessary to count these types of dependencies. The *Coordinate System Dictionary* does this counting.

Fig. 2 shows the hierarchy chart for *Quantity* and its subclasses. Fig. 3 shows the hierachy chart for *Value* and its subclasses.

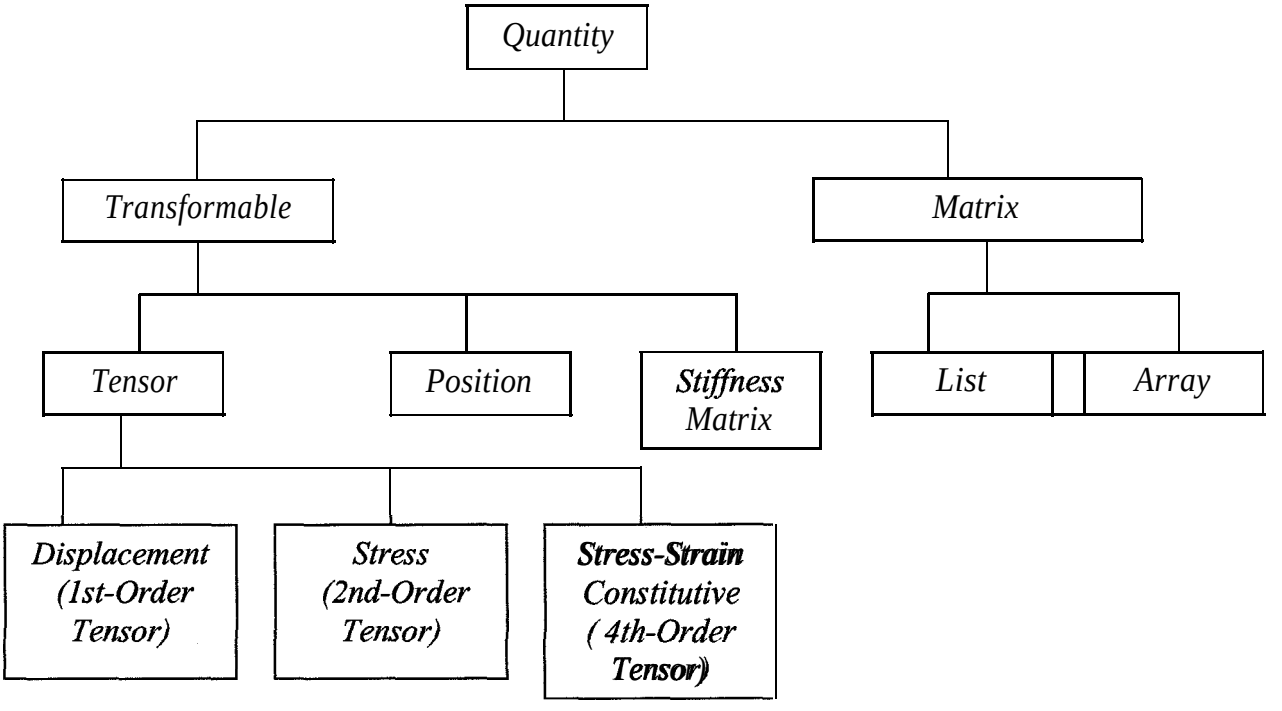


Fig. 2 Hierarchy Chart for *Quantity*.

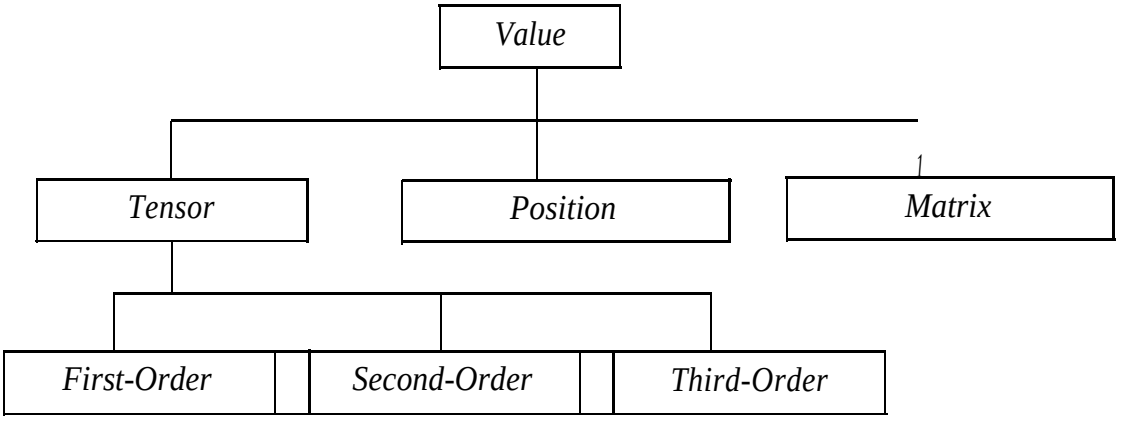


Fig. 3- Hierarchy Chart for Value.

Curvilinear Coordinate Systems

Consider the transformation of a *Quantity* from a Cartesian to a polar coordinate system, as shown in Fig. 4. Assume that we have three *Quantities:* a *Position* a *Force,* and a *Stress. The Position* is represented by (x,y) in the Cartesian and (r,θ) in the polar coordinate system. Because *Position* does not transform as given by the transformation equations for tensors given

above, it is not a true vector. Now, consider a *Force* located at *Position,* represented by (F_x, F_y) in the Cartesian and by (F_r, F_θ) in the polar coordinate system. According to the way a vector transforms, both components of this vector must have units of force (in both coordinate systems). The unit vectors e_i in each coordinate system are aligned with the locus of constant coordinates: constant x and constant y in the Cartesian Coordinate system and constant r and constant θ in the polar coordinate system. Similarly, the *Stress,* which is a tensor of second order, has components corresponding with directions aligned as shown in Fig. 4.

While tensors represented by components in rectangular coordinate systems are independent of position, tensors represented by components in curvilinear coordinate systems are meaningless unless the position at which the tensor is located is known. It is also interesting to note that a tensor located at the origin of the polar coordinate system cannot be defined using components, because the unit vectors, e_i, are indefinite at this point.

If we wish to include curvilinear coordinate systems in our *Quantity,* it will be necessary to indicate not only the coordinate system to which the components are referenced, but also the position within the coordinate system to which the components are referenced. The decision about where to store this information must yet be made: tentatively, it would make sense to store it in *Value.*

$$[X]=\begin{bmatrix} x \\ y \end{bmatrix} \qquad [X]=\begin{bmatrix} r \\ e \end{bmatrix}$$

$$[F]=\begin{bmatrix} F_x \\ F_y \end{bmatrix} \qquad [F]=\begin{bmatrix} F_r \\ F_\theta \end{bmatrix}$$

$$[\sigma]=\begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \qquad [\sigma]=\begin{bmatrix} \sigma_{rr} & \sigma_{r\theta} \\ \sigma_{\theta r} & \sigma_{\theta\theta} \end{bmatrix}$$

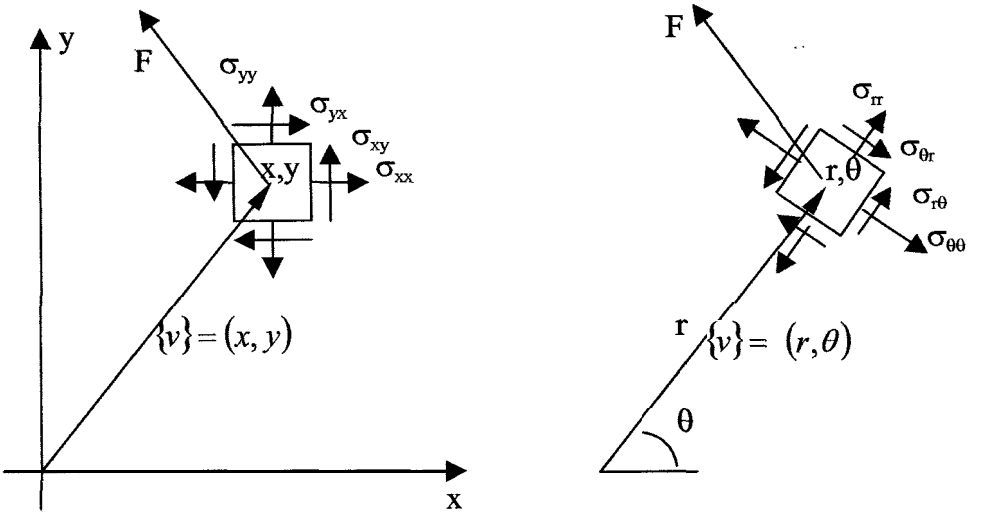


Fig. *4- Cartesian* to Polar Transformation

Conclusions

In this paper we have described the software design of an object called *Quantity*, and associate classes. The intent of Quantity is to perform much of the tedious work that is currently done in a very ad-hoc manner by engineers and scientists.

In the process of designing Quantity, we have had to struggle with fundamental questions. What is a position? What is a vector'? What is a tensor? Is a stiffness matrix a *Quantity*? *The* answers to these questions have not always been obvious. But they have a profound impact upon the "correct" design of *Quantity*.

We view *Quantity as* a high-level engine for interpreting and manipulating values. Massive quantities of data would probably not be stored as *Quantities* because of the high overhead involved. However, we would certainly want to store data in such a manner that it could be resurrected as a *Quantity* when desired.

While the design presented in this paper is by no means complete, we believe it is fundamentally correct. We are currently in the process of implementing *Quantity*. The correctness of the proposed design will ultimately be verified (or the converse) by its usefulness to scientists and engineers.

Acknowledgments

The financial support provided by the Center for Autonomous Control in Engineering (ACE) and the Albuquerque High Performance Computing Center is gratefully acknowledged. Also, the help of graduate student Raikanta Sahu is appreciated.

References

1. Funston, M., Valerie, R., and Gerstle, W. "Creation of Quantity: An Object-Oriented Reusable Class", University of New Mexico, NASA Autonomous Control Engineering, URC-TC Conference 1997.
2. Wirfs-Brock, R., Wilkerson, B. and Weiner, L. "Design of Object-Oriented Software", Prentice Hall, Englewood Cliffs, NJ, 1990.
3. Atencio, L., Gerstle, W., Panthaki, M., and Sahu, R., "Object-Oriented Design and Development of a Computational Mechanics Toolkit", Proceedings of NASA University Research Consortium Technical Conference '97, Albuquerque, NM, Feb. 16-19, 1997.
4. Block, H. D., "Introduction to Tensor Analysis", Charles Merrill Books, Inc., Columbus, Ohio, 1978.

98 URC170

Multipoint Measurements in an Axisymmetric Sudden Expansion

Erroll L. Eaton

Dept. of Mechanical and Aeronautical Engineering
Clarkson University
Potsdam NY

ABSTRACT

A flying wire system was utilized in conjunction with a rake of nine cross-wire probes to obtain simultaneous velocity measurements in an axisymmetric sudden expansion at a Reynolds number of 41,000. From these measurements, the correlation tensor could be calculated. Knowledge of the two-point correlation tensor reveals more in-depth information of the physical attributes of this flow. The two point correlation tensor allowed for calculation of the integrated length scales in both the radial and axial directions. This gives insight into the growth of structures with increasing downstream distance and at different radial locations through out the sudden expansion. The length scales were calculated by integrating the two-point correlation tensor in the radial direction from the centerline to the outer pipe wall and by integrating between several step heights for the axial direction. Calculated correlations at z/h = 8 and 9 at r/R = 0.46 showed a correlation length of 1/3 step height for the radial direction. It was found that length scales in the radial direction became larger with increasing radius with peaks at 0.70 < r/R c 0.81 but then decreased slightly towards the wall. Length scales in the axial direction yielded a recirculating bubble on the order of 3 step heights in the recirculating region. After the recirculating region, the length scales decreased to 1/4 of a step height.

Keywords: Axisymmetric,Sudden Expansion,Correlations,Length Scales,Step Height

1. INTRODUCTION

The axisymmetric sudden expansion is a fundamental separated flow encountered in many engineering situations. This basic configuration is used for sudden dump combustion studies. Using a co-annular setup, swirl is introduced before the expansion in the outer tube with a fuel source, the inner pipe is used for ignition the fluid and the recirculation zone is used as the flame holder. This flow is also useful when studying the effects of separation in ducts and heat exchangers. The separated, reattached and recirculating flow downstream greatly enhance the amount of heat, mass and momentum that are transferred, when compared to fully developed pipe flow at similar Reynolds number. This increase is caused by the large gradients of mean velocity in the separated region, This results in an increase in the turbulent kinetic energy levels in the flow field.

The flow in a suddenly expanded pipe is a complex three dimensional phenomena. It has been categorized into three distinct zones (Johnston 1976): recirculating, reattaching and redeveloping zones. At separation, there exists large velocity gradients and correspondingly high shear coupled with an adverse pressure gradient. The recirculating zone, the most difficult to study, exhibits many attributes. In this region, back flow velocity has been seen to be 10 percent of the axial velocity. This flow region is generated due to the presence of a solid boundary which inhibits entrainment of an 'outer' fluid so that a return flow originating further downstream is developed. This results in the creation of an unsteady eddy structure, Since the mean velocities are small and the turbulent intensities are large, it is difficult to obtain accurate quantitative turbulence measurements. In the reattaching zone, the velocity can be positive or negative with respect to the local mean. This inherently makes it difficult to define an exact point of reattachment. It has been previously defined by Rothe and Johnston(1975) as the region in which the flow is either in forward or backward flow 50 percent of the time. The flow in this region is highly unsteady resulting in a rapid decay of Reynolds stresses. Also large scale structures on the order of the step height are convected through this zone adding to the difficulty in measuring the flow here, The redeveloping zone exhibits pseudo-boundary layer characteristics where the flow will eventually become a fully developed flow if the pipe is long enough.

2. EXPERIMENTAL FACILITY

The flow facility used for this experiment is shown on the table of figures. The airflow was provided by an axial blower driven by a Lancer GPD 502 variable frequency drive. The air travels through a diffuser where it can be conditioned in the settling chamber. The flow chamber consists of a honeycomb section of flow straighteners and fine mesh screens for break up of swirl, The flow then enters an axisymmetric fifth order contraction section, which reduces the flow from a diameter of approximately 44.0" to 8.0" before it enters a linear axisymmetric contraction section that further reduces the diameter to 3.0". In order to ensure a fully developed turbulent flow before entry into the sudden expansion, a 3.0" diameter pipe was selected having a total length of 12ft. An expansion pipe with a diameter of 9.0" was selected that produced a step height of 3.0" and an expansion ratio 3. The total length of the expansion pipe is 9ft and is composed of 2 sections. The test section is 5ft and exit section is 4ft long. This union of pipes allows for easy access to the probes and support . Attached to the exit pipe is a filter made from 0.25in diameter , 2.5" in length sandwiched between fine mesh screens. This filter prevents any large eddies from propagating up through to the test section introduced from the laboratory environment. The wall pressure development was obtained by drilling 19 very fine holes into the test section to for pressure taps. This taps were drilled approximately one step height apart,

A 3/8" slot was milled in the bottom of the test section so that the probes can be propelled through the flow. It was found that this slot effects the flow significantly, if not closed off. To minimize it's effect, a sliding door was designed using two tape measures on either side of the test section to keep the slot closed at all times as the rake flies through the flow. The velocity of the sled containing the rake of cross wires is controlled by a programmable Parker Compumotor stepper indexer/driver. Communication to the Compumotor is achieved by software provided by Parker and interfaced through the RS-232 port of a personal computer. With this software the sled acceleration, velocity and distance can be controlled. Data is collected on the "fly" by the use of a trigger cable connected between the indexer and the A/D converter.

RESULTS

The main thrust of this experiment was to experimentally obtain the two-point velocit y correlation tensor Rij(r,r',z,z'), which wil I be used to calculate the integrated length scales. This tensor has never been measured in the axisymmetric sudden expansion because of its distinct regions discussed earlier. Therefore it was imperative that a "flying wire" system be used. Since the hot-wire is directionally insensitive, it is necessary to impose a velocity on the probes to capture gradients within the recirculating and reattaching zones. In order to estimate the length scales of the fluctuating motion, the congelation between the same fluctuating velocities measured at two different points in space needs to be studied. The length scale of the energy containing eddies is typically referred to as the "integral scale". Physically one can think of the two-point correlation as a measure of the strength of eddies or coherent structures in a flow. Coherent structures are now being considered as the building blocks for the entire flow . Once we identify these structures and determine their relationships to one another, we can determine the role they play in producing a universal characteristic of turbulence. The coherent structure play an important role in mass, heat and momentum transport as well as noise generation.. The size and growth of these coherent structures have been calculated using the integral scale. Distribution of these length scales have been examined separately in the radial and axial direction.

3.1 Correlations

Correlations were first calculated for the radial direction, expressed mathematically as Rij(r,r',z) where z is a fixed axial location. This tensor will provide insight into how well axial fluctuating velocities in the radial direction are correlated at fixed step height locations downstream. The magnitude of this tensor is shown for three distance regions discussed earlier at 9 radial positions. Ruu(r,r'3,) and Ruu(r,r'4) are located in the recirculating region. These plots show strong distinct peaks at the 5 radial location which is the shear layer that is developing from the separated flow. This peak dramatically decreases as we approach the wall where high fluctuations are dissipated. Examining these two plots even closer we can see that the base of the plots are getting wider. This is indicative of larger scales being encountered in the flow field. Also and inherently, the magnitude is increasing as we move downstream. The magnitude increase is due impart by the recirculating bubble that act as retarding agent on the flow as we move away from the inlet pipe. Ruu(r,r'8) and Ruu(r,r'9) are correlations in the reattaching zone. Here we see correlations over larger region, with a increase and then a decrease in magnitude from the recirculating zone. Moving into the reattachment zone, Ruu(r,r, I1) and Ruu(r,r',1 2), we see the magnitude of the tensor dropping off but correlations lengths are over a larger region, approximately 1/3 step height. Correlations Rvv(r,r' z) show the same trends as the Ruu but at considerably less magnitudes,

Axial correlations Ruu(r,z,z') are also presented for the distinct regions. Axial correlations are more pronounced because of the tine resolution due to the flying rake assembly. In the axial direction, 2650 pts were obtained over 42 inches, Therefore, con-elation plots are very fine and contain information at every 10th of a millimeter in the axial direction. Plots are shown for radial locations 1,3,5,7 which correspond to the core flow, outer core, shear layer and near wall positions. Each plot shows the correlation plots for a full step height, Each plot contains 18,000 pts. These plots show first, an increase in magnitude but then a decrease because of the boundary of the wall. For plots in the redeveloping region, correlation lengths are shown to exits over approximately 3 step heights. Overall magnitudes are larger in this region than in the recirculating region. In the redeveloping zone there is an overall decrease in magnitude of the correlation functions. This is due impart to the turbulent diffusions of the large coherent structures. Since coherent structures need high shear to exist, as we move away from the separated region, there is not enough shear to maintain these high fluctuations. Therefore eddies die off as they move downstream of the entry pipe.

3.2 Length Scales

As defined earlier, the length scales are an integral scale that are calculated from integrating the correlation function over the region desired, Since the correlations eventually go to zero at some point, then integration should reveal one distinct value, a length scale.

Length scales were first calculated for the radial direction at distinct step height locations. The spatial variation of Lii are shown to increase steadily up until r/R= 0.81, It is here that length scales start to dissipate due to the presence of the wall, Length scales are shown to grow to about the size of the step height before dissipating, In the axial direction length scales exhibit similar growth patterns. Progression in the axial direction show that length scales grow to be as large as 3 step heights, Physically, this indicates that structures are stretched in the axial direction by as much as 27.5 cm, The plot indicated that a maximum occurs at about 4 step heights downstream, This would indicate that we may have possibly captured the recirculating bubble in this flow.

After about 7 step heights, the length scales, as discussed earlier, decrease due to turbulent diffusion, Eventually there will be a converging of length scales to a distinct size due to redeveloping flow,

SUMMARY AND FUTURE WORK

This analysis has revealed some interesting preliminary results of the growth of length scales radially and axially. Correlations lengths were shown to be on the order of 1/3 of a step height for the radial direction and about 1 step height for the axial direction. Correspondly, the length scales were determined to be on the order of one step height for the radial direction and 3 step heights for the axial direction, both occurring in the recirculating zone. Further investigation into this analysis will reveal a more close relationship as to how these length scales grow in both direction's simultaneously.

APPENDIX:

LENGTH SCALES

$$L_{ij}^3 = \frac{}{R_{ii}(r, r', \theta, \theta', z, z')} \int_{z1}^{z2} \int_0^{2\pi} \int_0^R R_{ij}(r, r', \theta, \text{f?'}, \%, z') r' dr' d\theta' dz'$$

Radial Dependence :

$$L_{ij}^2 = \frac{1}{R_{ii}(r, r', .2)} \int_0^R R_{ij}(r, r', z) r' dr'$$

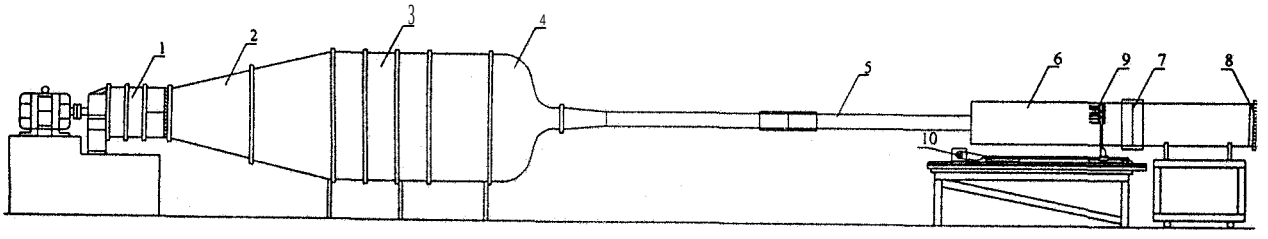
Correlation Tensor

$$R_{ij} = u_i(r, \overline{z_0) u_j(r'}, z_0)$$

Streamwise Dependence :

$$L_{ij} \cdot \frac{1}{R_{ii}(r, z, -Z')} \int_{z_1}^{z_2} R_{ij}(r, z, z') dz'$$

Integrated length scales are calculated over the. full domain of the correlation function because we are dealing .with an inhomogenous flow.



Experimental flow facility,

Figure 2.1

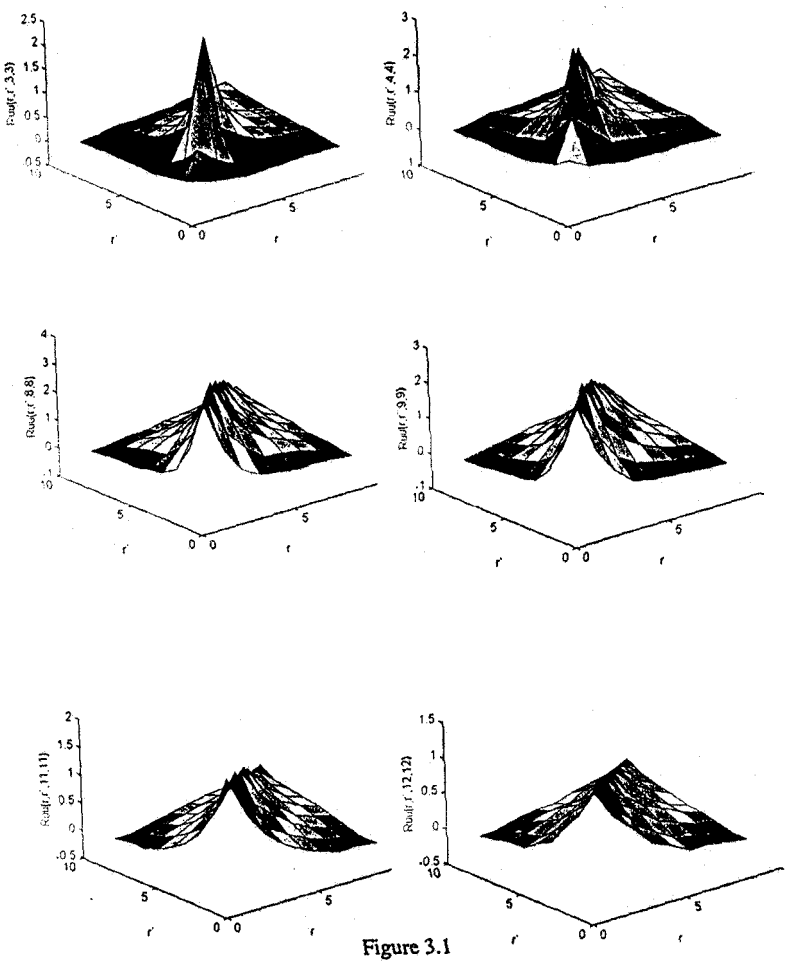


Figure 3.1

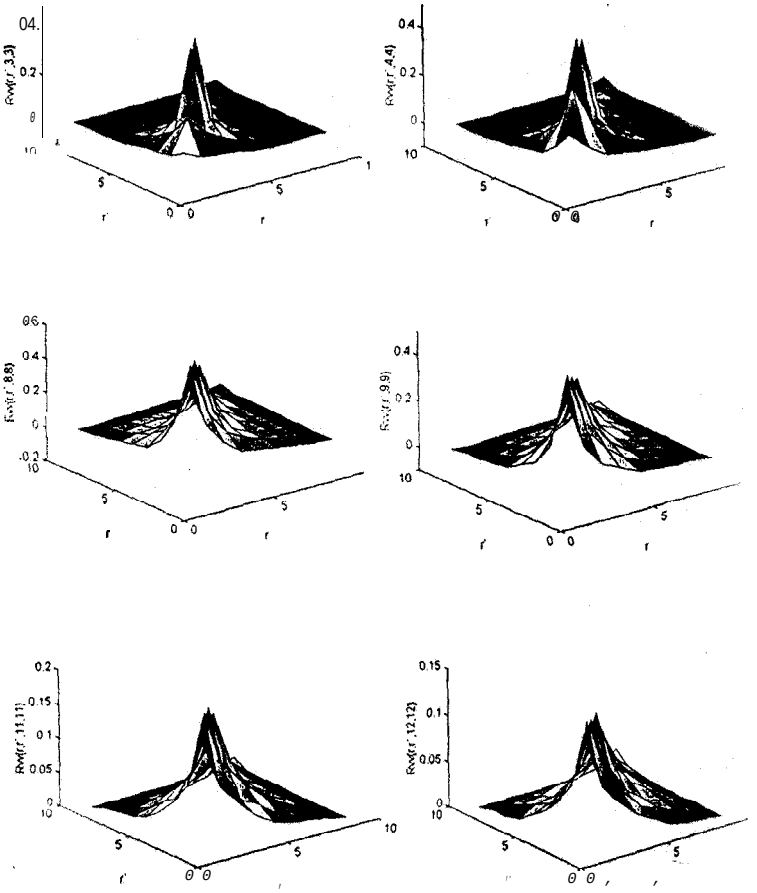
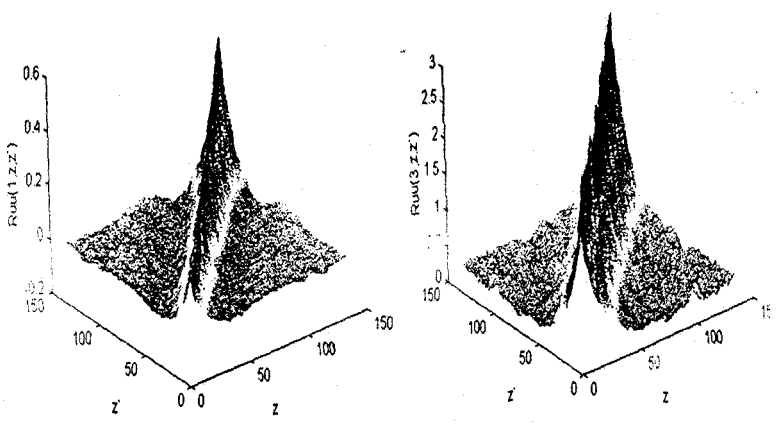
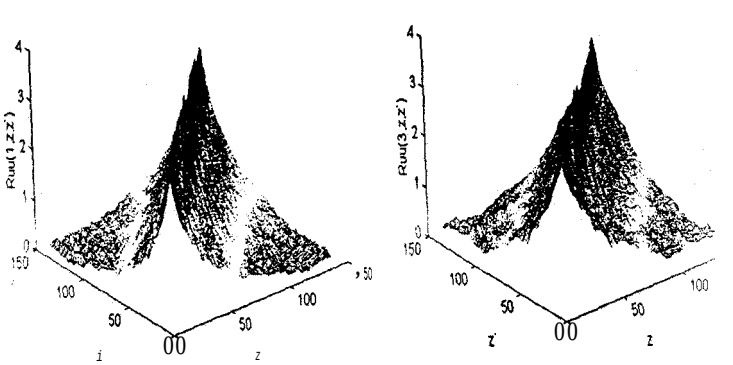


Figure 3.2



STEP HEIGHT 3-4



STEP HEIGHT 8-9

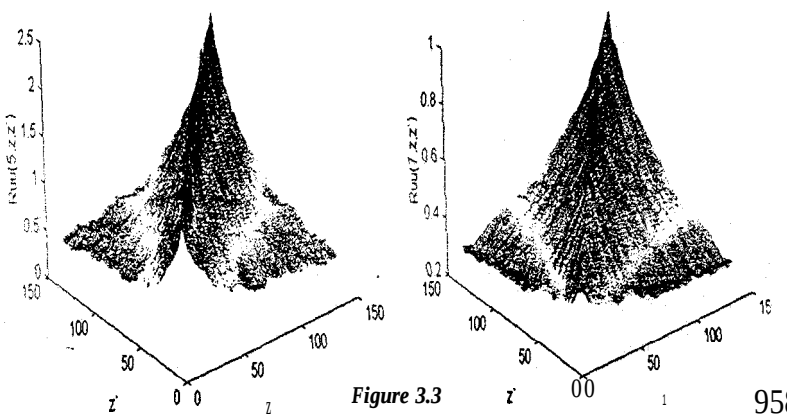


Figure 3.3

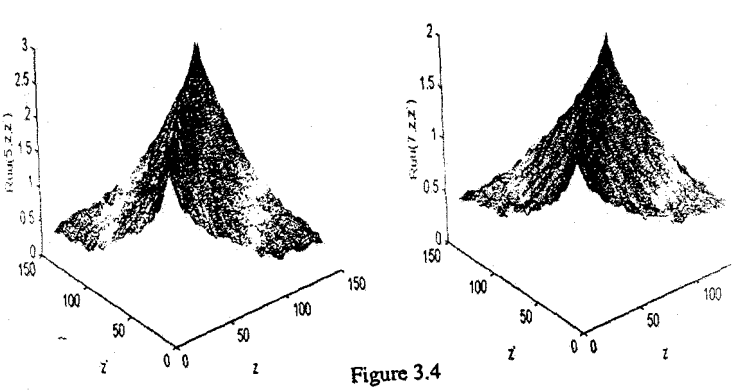
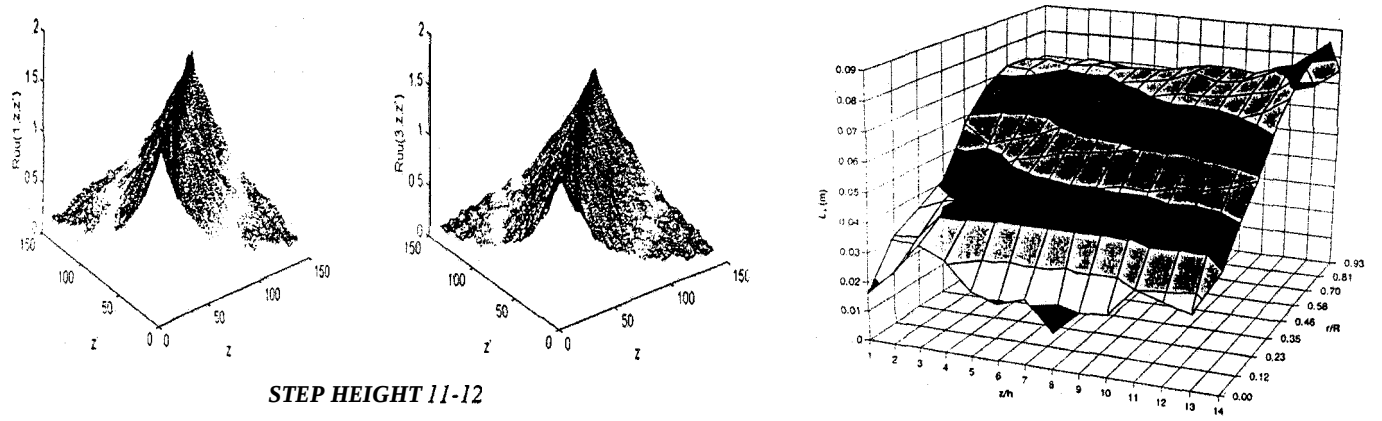


Figure 3.4



STEP HEIGHT 11-12

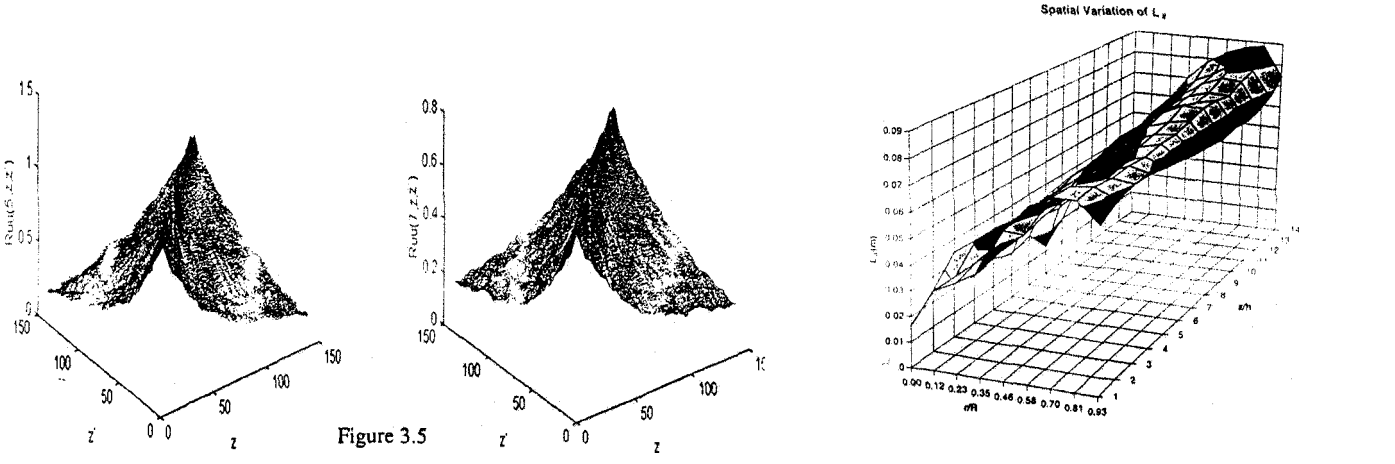


Figure 3.5

Length Scale Growth

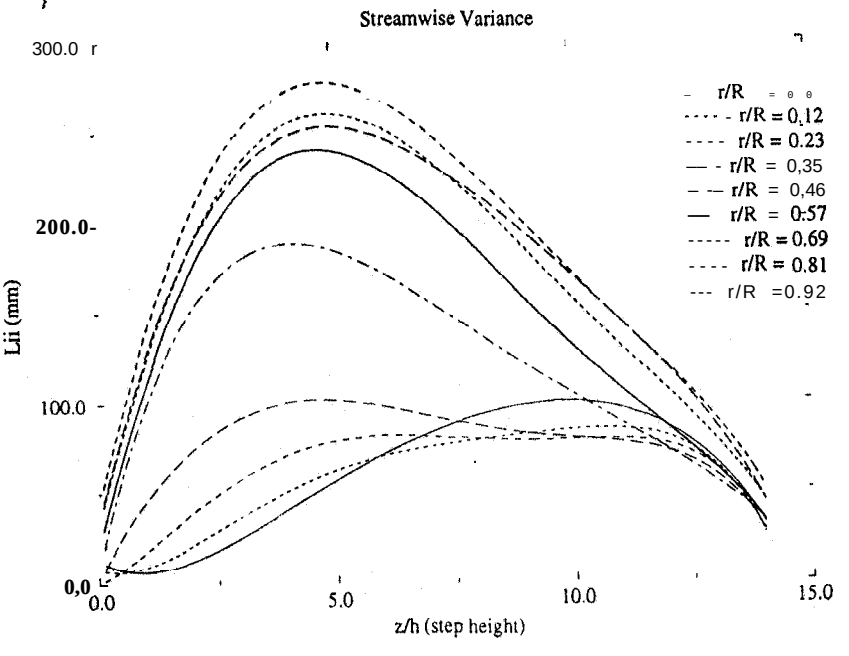


Figure 3.12

Facilitating Student Involvement in NASA Research:

The NASA Space Grant Aeronautics Example

Submitted to the

NASA University Research Centers Technical Conference 1998

Dr. Brent D. Bowen, Director

Michaela M. Schaaf, Coordinator

NASA Nebraska Space Grant Consortium

University of Nebraska at Omaha

Allwine Hall 422,6001 Dodge Street

Omaha, NE 68182-0406

Phone: (402) 554-3772 Fax: (402) 554-3781

nasa@unomaha.edu

Abstract

Many consider NASA programs to be exclusively space-oriented. However, NASA's roots originated in the aeronautical sciences. Recent developments within NASA elevated the declining role of aeronautics back to a position of priority. On a parallel pattern, aeronautics was a priority in the legislation which authorized the National Space Grant College and Fellowship Program. This paper outlines the development of the aeronautics aspect of the National Space Grant College and Fellowship Program, and the resulting student opportunities in research. Results from two aeronautics surveys provide a baseline and direction for further development. A key result of this work is the increase in student research opportunities which now exist in more states and at the national level.

Introduction

The foundation for the National Aeronautics and Space Administration (NASA) stems back to its predecessor, the National Advisory Committee for Aeronautics (NACA). All flying vehicles must rely on the basic fundamentals of flight and aeronautics. Even vehicles that reach outer space must be able to fly through the atmosphere of the earth and return safely home. Thus, the first 'A' in NASA, representing aeronautics, remains an important priority of the agency.

Upon its creation in 1958, NASA assumed the functions, responsibilities, duties, and powers of NACA. All NACA property, personnel, funds and records were also transferred to NASA. The importance of aeronautics to NASA's mission is clearly illustrated in the first sentence of the legislation which created the agency. "AN ACT To provide for research into problems of flight within and outside the earth's atmosphere, . . . " (Public Law #85-568, 1958).

Space Grant History & Development

The NASA Space Grant College and Fellowship Program, established in 1987, consists of 52 actively participating state consortia. The program seeks to enhance space and aeronautical-related research and education, while maintaining the United States' leadership role in aeronautics and space-related industry. Aeronautics is defined by Webster's dictionary as "a science dealing with the operation of aircraft; the art or science of flight" (1989). While this definition does not eliminate some forms of spaceflight, the definition of aeronautics can be limited to include only those activities which affect the safety, environmental compatibility, and the productivity of the air transportation system in our atmosphere.

The need for the role of aeronautics in the Space Grant was realized upon approval of the

original legislation creating this program. The National Space Grant College and Fellowship Act of 1987 (Public Law 100-147) supports the research of space resources which are defined as any tangible or intangible benefit which can be realized from aeronautical and space activities; or advancements in any field related to space.

Early in the program, David Downing, Director of the Kansas Space Grant Consortium, recognized that there were no discussions of aeronautics topics at Space Grant Meetings. Upon this realization, an aeronautics working group was formed to determine the role of aeronautics in the program. Following this formation, the Kansas Space Grant Consortium conducted a survey (1993) to measure interest in the concept.

The results of the survey indicated that a majority of the states have education programs related to aeronautics. Table 1 outlines some of the different aeronautical programs available at colleges and universities in the responding states. Aeronautical-related research is being conducted in some of the states. Categories of this research are identified in Table 2. This table also illustrates how many responding states conduct this activity. While it is not as evident as education and research, aeronautics-related industry also plays a role in the responding states. Table 3 provides data on the amount of aeronautics-related industry in these states.

The Case for Aeronautics

Long before the term space race was popular, the United States already had a 40-year old agency studying aspects of flight (McCurdy, 1993). However, as the space race took over the headlines, aeronautics took a backseat to space flight. Those within NASA realized the vital role aeronautics plays in the national security and economy of the country. Therefore, aeronautical research within NASA continued, but it failed to make the headlines.

Table 1

Aeronautical-related Education Programs in Responding States

Type of Education Program	% of States
Pilot Training	18.75
A&P Training	18.75
BS in Aeronautical Engineering	50.00
BT in Aeronautical Engineering	12.50
Graduate Program in related field	68.75
Other	31.25

Table 2

Aeronautics-related Research in Responding States

Research Area	% of States
Aeronautical Engineering	43.75
Flying/Handling Qualities/Pilot Human Factors	31.25
Flight Safety	18.75
Atmospheric Research relating to Aviation	12.50
Other	37.50

Table 3

Aeronautics-related Industry

in Responding States

Industry	% States
Airframe	25
Powerplant	18.75
Avionics	37.50
Other Systems	31.25

Many sources have been identified which support and define the role of aeronautics in NASA. The "Aeronautics & Space Transportation Technology Enterprise" is a section of the NASA Strategic Plan (1998) devoted to fostering aeronautical research within the agency. It lays out the plan to challenge the international competition so the U.S. can maintain its leadership role in aviation.

In January of 1995, NASA published a fact sheet titled, "Developing Cutting Edge Technology to Keep America the World's Leader in Aviation." This statement reveals a commitment to aviation research and development by NASA. The predictions made reflect a 100% increase in air travel by the year 2005. Experts predict this increase will generate a demand for $600 billion in new jet airliners. Experts further predicted 140,000 jobs and a supersonic transport aircraft marketplace valued at $200 billion by 2015. As NASA's involvement in aeronautics research and development cover general aviation to jumbo jets, from subsonic to hypersonic speeds, the agency has set a goal to recapture the 3 5°/0 of market share lost in the past 25 years.

A key part of NASA's Strategic Plan (1998) is based on the premise that NASA will maintain a leading role in the development of aeronautics technology in conjunction with both industry and other federal agencies. NASA will also actively support the upgrade of the national air transportation system. The Space Grant Program, with a presence in each state, should provide the educational base for aeronautics through research and outreach activities.

Consortia Which Have Joined the Aeronautics Working Group: The Challenge

Aeronautics in the Space Grant Program plays a key role in many space grant consortia projects. Projects such as Aviation Career Education (ACE) Academies are offered in several states, courses in aviation and/or aerospace, mini-grants for aviation studies, fellowship and scholarship research on flight and effects of flight, a jointly sponsored General Aviation Design Competition by NASA and the Federal Aviation Administration (FAA), conferences on the future of aviation, FAA Teacher Resource Centers, and many others represent the desire of Space Grant Consortium administrators, faculty, and students to incorporate aeronautics in the Space Grant College and Fellowship Program. While some states have aeronautics-related programs and/or research, the role of aeronautics has the potential to include more programs and a larger level of activity. Currently, there is no formal means of collaboration with other states on these projects. Deriving the information on current activities from each state and sharing this information will effect the collaboration of Space Grants on aeronautical-related projects within the Space Grant.

At previous Aeronautics discussions at Director Meetings, National Conferences, and Western Regional meetings, the following mission statement resulted:

- "To establish an organization which will aid in the coordination and communication of

activities and the exchange of information among state consortia having an interest in aviation and aeronautics."

- "To alert Space Grant personnel that there are significant interests and activities related to aviation and aeronautics within many state consortia."

To facilitate the mission, an action plan was developed by the initial working group. Some of these action items include:

- Creation of a World Wide Web site as an aviation clearinghouse. This task is being implemented by the University of Nebraska at Omaha Aviation Institute. The site is located at http: //cid..unomaha.edu/~unoai and is updated continuously.
- Foster the development of aeronautics research in the Space Grant Program by sponsoring specific opportunities for publications and presentations of research results.
- Co-sponsor the first air transportation specific scholarly journal. Abstracts of this publication, the Journal of Air Transportation World Wide, can be viewed at http: //cid.unomaha.edu/~unoai/journal.html
- Initiate collaborative research exchange visits.

Aeronautics Survey Update & Additional Action Items

An updated aeronautics survey was administered to each Space Grant in the summer of 1996 to collect input on proposed projects of the Aeronautics Working Group. The survey was designed by the Nebraska Space Grant with input from other states. A 77% response rate was achieved after each Space Grant received the initial mailing of the survey, and those that had not yet responded received two additional reminders. The survey results indicate there are many states that want to exchange aeronautics best practices and collaborate on aeronautics projects.

Of those responding, 48% indicated that there are not enough aeronautical activities within their Space Grant. Through the Aeronautics Working Group these activities are, and will continue to be, made possible.

One suggestion by the working group was the inclusion of aeronautics into the next poster session. Forty-five percent of survey respondents are willing to participate in this activity, and forty percent are undecided. This activity could also include a poster display in the lobby at NASA Headquarters to further promote Space Grant within the organization.

The Aeronautics Working Group will facilitate aeronautics contacts for each Space Grant at the NASA Centers. This is necessary due to the low response of consortia currently collaborating with NASA Centers on aeronautical projects. Sixty-three percent of responding consortia do not currently collaborate with NASA Centers on these projects. This is an item which can be promoted through the Aeronautics Working Group. Dryden Flight Center has been identified as the Center to begin facilitating contacts with for the 1996-1997 time frame.

Each state already includes, 77%, or is willing, 23%, to include institutions with aeronautical-related programs as academic affiliates. To facilitate those states which are willing to include these potential affiliates, a list of institutions with aeronautical-related programs will be provided by the Aeronautics Working Group. Additionally, the Aeronautics Working Group will also assist in identifying aeronautical-related industry contacts for Space Grant consortia. Seventy-two percent of responding states replied they have aeronautical-related industry within their state. Of these, 76% are willing to include them as industry affiliates, while 21 '%0 already include them, and 3% are unwilling to include them. The Aeronautics Working Group can also work with the 28°/0 of the responding states which did not identify aeronautical-related industry

in their state to identify companies and facilitate contacts.

Student Opportunities

The increased awareness of aeronautics in the Space Grant Program has similarly increased student research opportunities. More research scholarships and fellowships are being awarded to students in the aeronautics field as the number of aeronautics research projects rises in many states. The presence of aeronautics in Space Grant led to a national meeting and research collaborations at Dryden Flight Research Center in October 1997. Through faculty collaborations with NASA researchers, more students will be afforded opportunities to get involved in NASA-related research in their home states.

Programs at the national level are also being developed specifically for students. Last year marked the first year of the NASA Dryden Academy. This ten-week research experience at Dryden Flight Research Center afforded five undergraduates the opportunity to work with NASA scientists on research projects at the Center. Scholarships were provided by the Space Grant states to allow these students this once in a lifetime experience.

Nebraska established a program targeted specifically at involving Native American students in aeronautics research. This initiative allows students from the two Native American community colleges and the four tribal school districts to learn more about aeronautics and pursue research in this area. Other states in the Midwest are joining Nebraska in this endeavor.

Conclusion

Overall, the updated survey indicates the trend is turning towards aeronautics. More states are interested in aeronautics today than when the first survey was administered in 1993. This can be attributed to the implementation of the action items of the Aeronautics Working

Group. The group has exposed each state to the importance of aeronautics as a strategic enterprise of NASA. The goals are being accomplished with the simple inclusion of aeronautics topics on the agenda of each Space Grant meeting by explicitly mentioning the strategic initiative in aviation and aeronautics at all NASA Space Grant activities, publications, and presentation opportunities.

This trend is benefiting students as well. More opportunities are now available to students through individual states, and through the national program. The Nebraska effort to involve Native American students, faculty, and administrators in aeronautics outreach and research is setting an example for other states in the region. One of the goals of the national program is to encourage under-represented students to pursue careers and training in math, science, technology, and geography. Through the efforts of the Aeronautics Working Group, the goal is being realized.

Overall, the Aeronautics Working Group wants to ensure that the first “A” in the National Aeronautics and Space Administration is recognized as a vital component of the National Space Grant College and Fellowship Program.

Aeronautics in NASA and NACA. (1993). Washington, D. C.: Langley Research Center.

Glennan, T. K. (1993). The birth of NASA: The diary of T. Kenneth Glennan. Washington, D. C.: National Aeronautics and Space Administration, NASA History Office.

Kansas Space Grant Consortium. (1993). NASA Space Grant Aeronautics Survey. Unpublished manuscript. University of Kansas.

McCurdy, H. E. (1993). Inside NASA: High technology and organizational change in the U.S. space program. Baltimore: John Hopkins University Press.

NASA. (1988). NASA historical data book. Washington, D. C.: Scientific and Technical Information Division, National Aeronautics and Space Administration.

NASA. (1998). NASA strategic Plan. Washington, D. C.: National Aeronautics and Space Administration.

NASA. (January 1995). Developing cutting edge technology to keep America the world's leader in aviation. [Fact Sheet]. Washington, D. C.: National Aeronautics and Space Administration.

Nebraska Space Grant Consortium. (1996). NASA Space Grant Aeronautics Survey. Unpublished manuscript. University of Nebraska at Omaha Aviation Institute.

Public Law #85-568, 72 Stat., 426. (1958). National Aeronautics and Space Act of 1958.

Public Law 100-147. (1987). National Space Grant College and Fellowship Act.

File: White_URC_meeting.wpd Revised: December 15, 1997

Dr. Brent Bowen is Director and Professor, Aviation Institute, Department of Public Administration, University of Nebraska at Omaha. He has been appointed as a Graduate Faculty Fellow of the University of Nebraska System-wide Graduate College. Bowen attained his Doctorate in Higher Education and Aviation from Oklahoma State University and a Master of Business Administration degree from Oklahoma City University. His Federal Aviation Administration certifications include Airline Transport Pilot, Certified Flight Instructor, Advanced-Instrument Ground Instructor, Aviation Safety Counselor and Aerospace Education Counselor. Dr. Bowen's research on the development of the national Airline Quality Rating is regularly featured on ABC's Good Morning America, The Cable News Network, USA Today, The Today Show, The Associated Press, the network evening news shows, and in numerous other national and international media, as well as refereed academic publications. Dr. Bowen has in excess of 200 publications, papers, and program appearances to his credit. His research interests focus on aviation applications of public productivity enhancement and marketing in the areas of service quality evaluation, forecasting, and student recruitment in collegiate aviation programs. He is also well published in areas related to effective teaching. Dr. Bowen is an active industry consultant, pilot, and former fixed base operator and air carrier operator. Bowen has been an invited expert witness before the U.S. House of Representatives Committee on Government Operations and has served on multiple occasions as an invited speaker and panelist at the National Academy of Sciences/Transportation Research Board.

His professional affiliations include the University Aviation Association, Council on Aviation Accreditation, World Aerospace Education Association, International Air Transportation Research Group, Aerospace Education Association, Alpha Eta Rho International Aviation Fraternity, and the Nebraska Academy of Sciences.

Additionally, Dr. Bowen has authored/co-authored numerous successful funding proposals totaling in awards exceeding $9 million. He also serves as program director and principal investigator of the National Aeronautics and Space Administration funded Nebraska Space Grant Consortium and was recently awarded a three-year research grant from NASA's EPSCoR Program exceeding $3 million.

Mrs. Michaela Schaaf is the Coordinator of Research and Special Programs for the Aviation Institute. She holds a Bachelor's degree in Aviation Studies and Master of Science degree, Aviation Concentration, from the University of Nebraska at Omaha. Mrs. Schaaf is currently working towards a Doctorate in Public Administration with an Area of Specialization in Aviation Administration from the University of Nebraska at Omaha. In addition to administering NASA-funded research programs in excess of $5 million, Mrs. Schaaf instructs Writing& Research in Aviation, Diversity in Aviation, and other courses for the Aviation Institute. She serves as the Institute's Honors Coordinator and chairs the Committee on Student Recruiting and Retention. Her Federal Aviation Administration certificates include Private Pilot and Basic Ground Instructor. Mrs. Schaaf is a member of the University Aviation Association, American Society for Public Administration, Omicron Delta Kappa, Alpha Eta Rho, and Civil Air Patrol. Her research interests are in the areas of curriculum development, program evaluation, and space policy.

A

B

C

CONTINUED

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Y

Z

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	22 Feb 98	Conference Proceedings, February 22-25, 1998

4. TITLE AND SUBTITLE

NASA University Research Centers: Technical Advances in Aeronautics, Space Sciencesand Technology, Earth Systems Sciences, Global Hydrology, and Education . Volumes II and III

5. FUNDING NUMBERS

NCC 8-140

6. AUTHOR(S)

T.L. Coleman, B. White, S. Goodman, P.Sakimoto, L. Randolph, and D. Rickman, editors

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Hydrology, Soil Climatology and Remote Sensing
Alabama A&M University
P.O. Box 1208
Normal, AL 35762

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Minority University Research and Education Divison
Office ofEqual Opportunity Programs
Code EU
NASA Headquarters
Washington, DC 20546

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

N/A

11. SUPPLEMENTARY NOTES

Volumes II and III together comprise the proceedings of the NASA URC Technical Conference (URC-TC '98), February 22-25, 1998, Huntsville, Alabama

12a. DISTRIBUTION AVAILABILITY STATEMENT

Publicly available

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

USE FIRST AND SECOND PARAGRAPHS FROM THE FOREWORD IN VOLUME II

14. SUBJECT TERMS

Minority Universities, Historically Black Colleges and Universities, Hispanic Serving Institutions, Earth science, space science, education, aeronautics, space medicine, space life sciences, aerospace technology, hydrology

15. NUMBER OF PAGES

975

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED