

NASA (17 16)
ANALYTIC IN-CAT 99

NASA Contractor Report 188410

PRIMARY 60300



SUBSIDIARIES

60301

P-302

National Aeronautics and Space Administration
(NASA)/American Society for Engineering Education
(ASEE) Summer Faculty Fellowship Program - 1994

Volume 1

Richard Bannerot and Donn G. Sickorez, Editors

(NASA-CR-188410) NATIONAL
AERONAUTICS AND SPACE
ADMINISTRATION (NASA)/AMERICAN
SOCIETY FOR ENGINEERING EDUCATION
(ASEE) SUMMER FACULTY FELLOWSHIP
PROGRAM, 1994, VOLUME 1 (NASA.
Johnson Space Center) 202 p

N95-32418
--THRU--
N95-32434
Unclass

G3/99 0060300

Grant NGT 44-001-800

July 1995

NASA Contractor Report 188410

National Aeronautics and Space Administration
(NASA)/American Society for Engineering Education
(ASEE) Summer Faculty Fellowship Program - 1994

Volume 1

Richard Bannerot, Editor
University of Houston
Houston, Texas

Donn G. Sickorez, Editor
Lyndon B. Johnson Space Center
Houston, Texas

Grant NGT 44-001-800

July 1995

Preface

The 1994 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by Texas A&M University and JSC. The program at JSC, as well as the programs at other NASA centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began nationally in 1964 and at JSC in 1965, are to

1. Further the professional knowledge of qualified engineering and science faculty members.
2. Stimulate an exchange of ideas between participants and NASA.
3. Enrich and refresh the research and teaching activities of participants' institutions.
4. Contribute to the research objectives of the NASA centers.

Each faculty fellow spent at least 10 weeks at JSC engaged in a research project in collaboration with a NASA JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1994. Volume 1 contains reports 1 through 18 and Volume 2 contains reports 19 through 33.

Contents

Volume 1

Acharya, R.S.	Characterization of Microgravity Effects on Bone Structure and Strength Using Fractal Analysis.....	1-1	21
Akl, F.A.	Application of Impact Dampers in Vibration Control of Flexible Structures	2-1	2
	Not used.....	3-1	3
Barhorst, A.A.	Certification of a Hybrid Parameter Model of the Fully Flexible Shuttle Remote Manipulator System.....	4-1	
Barry, S.R.	The Role of the Vestibular System in Manual Target Localization	5-1	4
Berry, F.C.	Modeling of DC Spacecraft Power Systems.....	6-1	5
Blanford, G.E.	Image Analysis Used to Count and Measure Etched Tracks from Ionizing Radiation	7-1	6
Chang, T-H.	Unified Approach for Incompressible Flows.....	8-1	7
Chen, Y-L. A.	Characterizing the Fragmentation Ring on Witness Plates	9-1	8
Duncan, A.B.	Baseline Experimental Investigation of an Electrohydrodynamically Assisted Heat Pipe.....	10-1	9
Figueroa, F.	Bicycle Ergometer Instrumentation to Determine Muscle and Bone Forces During Exercise.....	11-1	10
Gantenbein, R.E.	Validation of Space/Ground Antenna Control Algorithms Using a Computer-Aided Design Tool	12-1	11 4974
Hall, E.B.	The Composite Classification Problem in Optical Information Processing.....	13-1	12 4975
Hunsucker, J.L.	The Background and Theory of Integrated Risk Management....	14-1	13 4976
Ingalls, S.A.	Implementation of a Single-Stage-to-Orbit (SSTO) Model for Stability and Control Analysis.....	15-1	14 4977
Johnson, G.G.	Scanners, Optical Character Readers, Cyrillic Alphabet and Russian Translations.....	16-1	15
	Not used.....	17-1	16
Meade, A.J., Jr.	Numerical Solution of Differential Equations by Artificial Neural Networks.....	18-1	17

Contents
(concluded)

Volume 2

Navard, S.E.	Modeling Early Failures of Electronic Components	19-1
Overmyer, S.P.	Display and Human-Computer Interface (HCI) Design Prototyping for the Payload Video Console (PVC).....	20-1
Perry, E.	Construction and Testing of a Laser-Excited System for Extraction of Oxygen of Silicates and Oxides for Isotopic Analysis.....	21-1
Reffell, M.	White Board Project	22-1
Richmond, E.R.	Software Engineering Analysis of the Dynamic Environment Communications Analysis Testbed (DECAT) Model.....	23-1
Mohammad, A.R.	Spectral Measurements in the Plenum of an Arcjet Wind Tunnel	24-1
Schwartz, D.R.	The Definition and Preliminary Design of a Tool for Design Knowledge Capture.....	25-1
Snow, L.D.	Physiological Alterations and Pharmacokinetic Changes During Space Flight.....	26-1
Tan, A.	Analysis of Orbital Decay of the ODERACS Spheres	27-1
Thompson, M.W.	Interplanetary Communications Systems Designs in the 37.0-40.5 GHz Frequency Bands	28-1
Todd, B.A.	Finite Element Modeling of the Lower Extremities	29-1
Waugaman, D.	Support of Pulse Detonation Engine Testing (Project A) and NASA-NIST Shock Tube Rupture Disk for Transient Pressure Transducer Calibration (Project B).....	30-1
Wilcox, L.M.	Requirements and Design of the DECAT GUI Syntax Checker.....	31-1
Williams, R.D.	Media Selection for Virtual Environment Technology.....	32-1
Dorough, D.	Inertia Properties of the Lower Extremities	33-1

**Characterization of Microgravity Effects on Bone Structure and
Strength using Fractal Analysis**

51-52
4959

Raj S. Acharya, Ph.D.
Director, Biomedical Imaging Group
State University of New York at Buffalo
Buffalo, NY 14260
Phone: (716) 645-2318
email: acharya@eng.buffalo.edu
Date submitted: Oct 12, 94

Linda Shackelford, M.D.
Technical Lead, Bone and Muscle Lab
NASA Johnson Space Center
Houston, Texas 77058
Phone: (713) 483-7100
Fax: (713) 483 6227

Abstract

The effect of micro-gravity on the musculoskeletal system has been well studied. Significant changes in bone and muscle have been shown after long term space flight. Similar changes have been demonstrated due to bed rest. Bone demineralization is particularly profound in weight bearing bones. Much of the current techniques to monitor bone condition use bone mass measurements. However bone mass measurements are not reliable to distinguish Osteoporotic and Normal subjects. It has been shown that the overlap between normals and osteoporosis is found for all of the bone mass measurement technologies: single and dual photon absorptiometry, quantitative computed tomography and direct measurement of bone area/volume on biopsy as well as radiogrammetry. A similar discordance is noted in the fact that it has not been regularly possible to find the expected correlation between severity of osteoporosis and degree of bone loss.

Structural parameters such as trabecular connectivity have been proposed as features for assessing bone conditions. In this report, we use fractal analysis to characterize bone structure. We show that the fractal dimension computed with MRI images and X-Ray images of the patella are the same. Preliminary experimental results show that the fractal dimension computed from MRI images of vertebra of human subjects before bedrest is higher than during bedrest.

1 Introduction

The effect of micro-gravity on the musculoskeletal system has been well studied. Significant changes in bone and muscle have been shown after long term space flight. Similar changes have been demonstrated due to bed rest. Bone demineralization is particularly profound in weight bearing bones. Much of the current techniques to monitor bone condition use bone mass measurements. However bone mass measurements are not reliable to distinguish Osteoporotic and Normal subjects [29]. It has been shown that the overlap between normals and osteoporosis is found for all of the bone mass measurement technologies: single and dual photon absorptiometry, quantitative computed tomography and direct measurement of bone area/volume on biopsy as well as radiogrammetry. A similar discordance is noted in the fact that it has not been regularly possible to find the expected correlation between severity of osteoporosis and degree of bone loss [30].

Structural parameters such as trabecular connectivity have been proposed as features for assessing bone conditions [29]. It has been shown that in vertebral crush fracture patients, elements such as vertical trabeculae are retained more or less intact, while elements such as horizontal bracing trabeculae are resorbed entirely [31] [32]. This results in disconnection of large number of trabecular elements. However, in non-fracture patients connections between elements were preserved. Long vertical trabeculae are subject to buckling under loading. When they lose their lateral connections to adjacent trabeculae, the degree of buckling may exceed the inherent strength of the bone. Structure can be thus be seen as an important feature in assessing bone condition.

2 Significance

Protecting humans against extreme environmental conditions requires a thorough understanding of the pathophysiological changes resulting from the exposure to those extreme conditions. The knowledge of the degree of medical risk associated with the exposure is of paramount importance in the design of effective prophylactic and therapeutic measures for space exploration. Major health hazards due to musculoskeletal systems include the signs and symptoms of hypercalciuria, lengthy recovery of lost bone tissue after flight, possibility of irreversible trabecular bone loss, the possible effect of calcification in the soft tissues, and the possible increase in fracture potential. Our research to relate the local trabecular structural information to microgravity conditions is an important initial step in understanding the effect of microgravity and countermeasures on bone condition and strength. The proposed research is also closely linked with Osteoporosis and will benefit the general population.

3 Hypothesis/Rationale

- The rationale for this research is based on the premise that microgravity conditions change bone structure as well as bone mass.
- Bone structure can be characterized by fractal geometry.

- Fractal characterization of bone structural changes due to microgravity conditions is not only optimal but also pragmatic.

4 Specific Tasks Achieved

The overall goal is the characterization of bone structural changes due to microgravity with the aid of fractals.

- We have related the fractal dimension computed from MRI images to those obtained with X-Rays of the same specimen.
- We have computed the fractal dimension from MRI images of subjects before, during and after bed rest.
- We have computed the fractal dimension of faxitron images of Osteoporotic and Normal Bone slices of human subjects.

During the summer fellowship, we have implemented and tested the fractal algorithms on sample MRI bone images.

The long term goal we are pursuing is to integrate fractal and finite element analysis. We are interested in studying the effect of various external parameters on local structural information provided by our method. This makes it possible to quantify the effect of countermeasures on local structural information of the bones.

5 Summary of Results Achieved

- Fractal Dimension of the human subject tested is higher before bedrest than during bedrest. This shows that the Fractal Dimension decreases due to bedrest.
- Fractal Dimension of Osteoporotic subjects is less than the Fractal Dimension of Normal Subjects in the Faxitron Image Study.
- Fractal Dimension measured with the MRI image and X-Ray image of the Patella is the same.

6 Fractal Model

With natural objects, familiar metrics from classical geometry such as *length*, *area* and *volume* depend on the scale at which we decide to look at the object (*e.g.* the size of the “yard stick”). As an example, one can show that the surface area of a sand grain is entirely dependent on the scale at which one chooses to look at it. The smaller the scale, higher the surface area since more nooks and crannies become visible.

Fractal geometry [12] characterizes this ability of an n dimensional object to fill the $n + 1$ dimensional space where the relationship of a measure M with topological dimension n and the scale ϵ is expressed as

$$M(\epsilon) \propto \frac{1}{\epsilon^r}, \quad 0 \leq r \leq 1 \quad (1)$$

where the quantity $r + n$ (denoted by D) is called the *fractal dimension* or *Hausdorff-Besicovich dimension* and characterizes the degree of erratic behavior. For an ideal object, the measure M is independent of the scale ϵ and hence $n = D$. Thus a fractal object can be defined as a set for which the fractal dimension is greater than the topological dimension. There are number of ways in which fractal geometry can be applied to the analysis of texture in images. Pentland et.al [19] used a method related to the coocurance matrix technique of texture classification based on fractal dimension. They find the standard deviations of the difference of gray levels separated by a given vector and plot it against the vector lengths as a log-log graph. In another technique [18] two dimensional gray level image is represented as a three-dimensional surface whose height at each point represents the gray level at that point and the surface area is measured at different scales. It has also been shown that [9] a n -dimensional fractal object can be characterized by the fractional Brownian motion of n variables and that the relationship between the power spectral density and r are independent of the projection [7]. This makes the fractal dimension computed from the projections of n -dimensional fractal object represent that of the original object.

It has been shown that the power spectra of a fractal object exhibits an inverse-power relation to frequency. For one-dimensional signals, this can be expressed as,

$$S(\omega) \propto \frac{1}{|\omega|^r} \quad (2)$$

and the parameter r is related to the fractal dimension by the equation,

$$D = \frac{5 - r}{2} \quad (3)$$

Thus one can exploit the above relationships to estimate the fractal dimension in frequency domain by finding the slope of the plot of log power spectra vs. log frequency. For multi-dimensional case, the corresponding relationship has been shown to be

$$S(\omega_1, \omega_2, \dots, \omega_n) \propto \frac{1}{(\omega_1^2, \omega_2^2, \dots, \omega_n^2)^{\frac{r}{2}}} \quad (4)$$

where

$$D = \frac{2n + 3 - r}{2} \quad (5)$$

In the next section, we reiterate some basic morphological operations and present an algorithm to obtain the fractal dimension of a three dimensional surface. This approach gives us the added capability of dealing with shape using different structuring elements. Since the underlying theme is based on fractals, all of the above properties hold for morphological fractals as well.

Our approach is based on the above method where the digitized radiograph is represented as a surface whose height represents the gray level at each point. The surface area at different scales is estimated using a series of dilations of this surface by a given structuring element whose size determines the scale. The derivations in the next section shows how one can obtain the surface area from the volume of the dilates and some simplifications which allows one to use dilations by a fixed size structuring element.

7 MORPHOLOGICAL FRACTALS

Mathematical Morphology as developed by Matheron and Serra [25] is basically a Set Theory and uses set transformations for Image Analysis. It extracts the impact of a particular shape on images via the concept of Structuring Element (SE). The SE encodes the primitive shape information. In a discrete approach, the shape is described as a set of vectors with respect to a particular point, the Center, which does not necessarily belong to the SE. During Morphological transformation, the Center scans the whole image and matching shape information is used to define the transformation. The transformed image is thus a function of the SE distribution in the original image.

In particular, *Dilation* of a set X with a SE Y is given by the expression

$$X \oplus Y = \{x : Y^x \cup X \neq \emptyset\} \quad (6)$$

where Y^x indicates the translation of set Y with x . The operations gives a set whose surface is the path traced by the center of the SE Y when it traverses along the surface of X . Using the above operation, surface area of a compact set X with respect to a compact convex SE Y which is symmetrical with respect to the origin is given by [25] :

$$S(X, Y) = \lim_{\rho \rightarrow 0} \frac{V(\partial X \oplus \rho Y)}{2\rho} \quad (7)$$

where ∂X is the boundary of set X and \oplus denotes the dilation of the boundary of set X by the SE Y scaled by a factor ρ . $V(X)$ gives the volume of set X . It has been observed that even though for “regular” classes of sets, the surface area $S(X, Y)$ is finite, for many “natural” objects, this can be infinite.

From the above expressions, it can be seen that dilating by ρY hides all structures smaller than ρY and therefore is equivalent to looking at the surface at scale ρ . For experimental purposes, we can calculate the surface area of a set X at scale ρ

$$S(\partial X, Y, \rho) = \frac{V(\partial X \oplus \rho Y)}{2\rho} \quad (8)$$

If the object is regular, the surface area will not change with ρ_i . For a fractal object, $S(\partial X, Y, \rho)$ increases exponentially with decreasing ρ as seen from equation (1). By taking the logarithm, we now have

$$\begin{aligned}\log(S(\partial X, Y, \rho)) &= \log(K) - r \log(\rho) \\ D &= 2 + r\end{aligned}\tag{9}$$

Where K is the proportionality constant. We can now estimate D by plotting $\log[S(\partial X, Y, \rho_i)]$ Vs. $\log[\rho_i]$ for a given set of scale factors ρ_i , $i = 1, 2, \dots, N$ and calculating the gradient of the line that fits the plot.

The series of dilation of X by $\rho_i Y$ required for the above computation can be further reduced to dilation by the unit element Y by observing that if $X_Y^\rho = X \oplus \rho Y$, then

$$X_Y^{\rho+1} = X \oplus (\rho + 1)Y = (X \oplus \rho Y) \oplus Y = X_Y^\rho \oplus Y\tag{10}$$

Apart from having projection angle and scale invariance, with the morphology method, we also have the freedom of selecting a structuring element suited for the problem at hand. This, coupled with the fact that the method only involves dilation makes the implementation straightforward compared with the other methods where it is required to find the covering of the boundary of X (∂X) by spheres in 3D and disks in 2D.

The surface area $S(X, Y, \rho_i)$ can be iteratively calculated as follows: Let the image X be defined as the set of triplets $\{< f(x, y); x = 1, N; y = 1, M\}$ and the the structuring element Y be given as a set of triplets $\{< x_i, y_i, z_i >, i = 1, P\}$. The ρ^{th} dilate $f_\rho(x, y)$ is calculated as

$$f_\rho(x, y) = \max\{f_{\rho-1}(x + x_i, y + y_i) + z_i, i = 1, 2, \dots, P\}\tag{11}$$

The initial condition $f_0(x, y)$ is set to $f(x, y)$.

The surface area at each step can be calculated by using equation(8) where

$$S(X, Y, \rho) = \frac{V(X, Y, \rho)}{2\rho} = \frac{\sum_{x \in 1, N; y \in 1, M} (f_\rho(x, y) - f(x, y))}{2\rho}\tag{12}$$

8 Alternating Sequential Filters

In this section we will use Morphological Pyramids to compute the Fractal Dimension. Generally, dilation and erosion are applied in pairs to make the transformation independent of the origin of the SE. The opening operation of a set F by a SE B is defined as,

$$F \circ B = (F \ominus B) \oplus B$$

The dual operation closing is defined as

$$F \bullet B = (F \oplus B) \ominus B$$

A composite opening-closing mapping can be defined as,

$$M_B(F) = (F \circ B) \bullet B$$

An iterative application of such an operator is defined as a Alternating Sequential Filter (ASF) [15],[26].

$$ASF(F) = M_{B_N} M_{B_{N-1}} \dots M_{B_1}(F)$$

where N is an integer and B_N, B_{N-1}, \dots, B_1 are SEs with different sizes satisfying the constraint $B_N \supseteq B_{N-1} \supseteq \dots B_1$.

During a morphological transformation, the structuring element scans the whole image and modifies each point depending on the structural similarity of the SE with the image at that point. During erosion, any structure in the foreground smaller than the SE is removed from the image. Similarly, dilation removes such structures in the background. The composite operation of opening and closing provides the same result with the added advantage of independence from the origin of the SE. The result of such a transformation can be interpreted as a transformation where the details of the original image smaller than the SE are removed. Thus the repeated application of such transformations using SE with increasing size will result in a sequence of images with decreasing details. This is equivalent to a representation of the image at decreasing resolution. This multiresolution representation is used for the surface area measurement.

Though opening and closing are sufficient for such a purpose, use of ASF is preferred. ASF is more robust and introduces less distortion than individual application of opening or closing [15]. A smaller structuring element should be used before a larger one.

The direct application of ASF needs iterative morphological transformation using increasing structuring elements. As the SE size increases, the computation involves increases. In [15], it has been shown using morphological sampling theorem [4] that ASF pyramid can be obtained equivalently by decreasing the size of the image instead of increasing the size of structuring element. At any level of the pyramid, the next level representation is obtained by subsampling the image and then transforming the sampled image using a constant sized element. As the image size is decreased and the SE size remains unchanged, this process is more efficient.

Once the morphological pyramid is obtained, the surface area is computed using a *piece-wise planer* approximation. The image is viewed as a sampled version of continuous two-dimensional surface. In the approximation, the surface is considered as the union of non-overlapping triangular areas. The triangles are formed in two steps. First, the support of the surface is divided into a number of squares, each having the side equal to one grid spacing. An $M \times N$ image is composed of MN such unit squares. Each square is represented by the pixels at four corners $(p, q), (p + 1, q), (p, q + 1)$ and $(p + 1, q + 1)$. In the next step, each such square is divided into two triangles, one of which is represented by the pixels at $(p, q), (p + 1, q)$ and $(p, q + 1)$ and the other by $(p + 1, q), (p + 1, q + 1)$ and $(p, q + 1)$. The area of any such triangle is computed using the pixel gray level values of its three corners. The surface area of the image is approximated by the sum of the area of all the triangles. The image size decreases as the resolution decreases. To represent the correct surface area at all resolutions, the area computation takes the resolution step into account by normalizing the grid space (for example, the grid spacing in the 1/4-th resolution represents 4 gridspacing of the original resolution).

The slope of the log-log plot of the surface area against the resolution is computed using least-square estimation.

9 Preliminary Experimental Results

9.1 Relating the Fractal Dimension of X-Ray and MRI images of the patella.

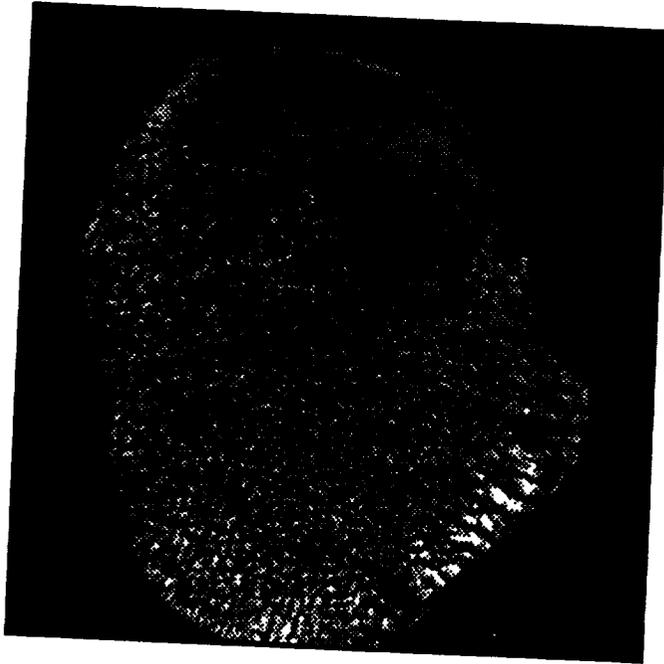
An isolated human patella was scanned in MRI scanner. Additionally, X-Ray images of the same patella was obtained. Fractal dimension was computed in four different regions of the MRI image with a window size of 50. Fractal Dimension was also computed in the same regions on the X-Ray image. Figure 1 shows the MRI image and the X-Ray image of the patella. The computed fractal dimension in each of the four different areas are shown. The tabulated results in figure 1 show that there is no significant difference between the fractal dimension computations between the MRI image and the X-Ray image.

9.2 Faxitron Bone Images of Osteoporotic and Normal Subjects

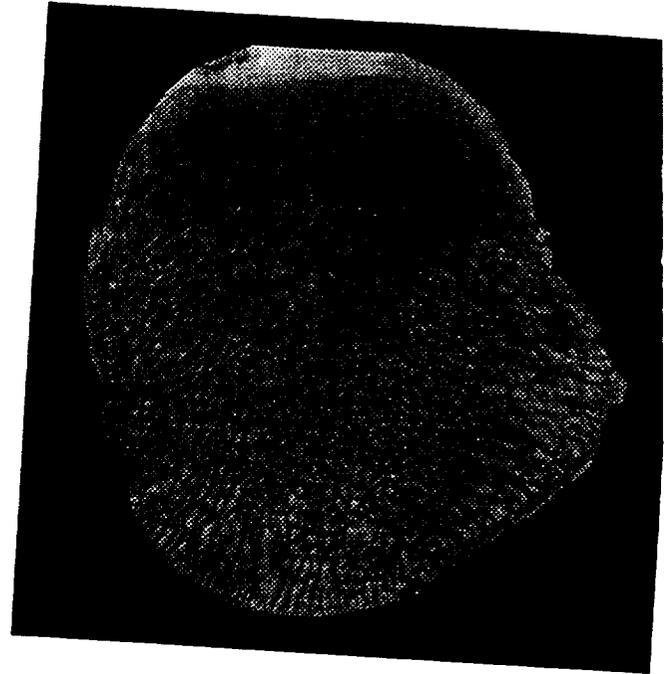
Faxitron bone images of the normal and human subjects were obtained. Fractal dimension was computed with a window size of 110 by 110 on six slices of normal and 8 slices of osteoporotic bone images. The fractal dimension computed with the flat and pyramid SEs are shown. The fractal dimension computed on the osteoporotic subjects was smaller than the fractal dimension computed with the normal subjects.

9.3 Bedrest Studies

Fractal dimension was computed on a series of MRI images of the vertebra of human subjects scanned before, during and after bedrest. The computed result on figure 3 shows that the fractal dimension of the subject before bedrest was higher than the fractal dimension during bedrest.



MRI image of patella

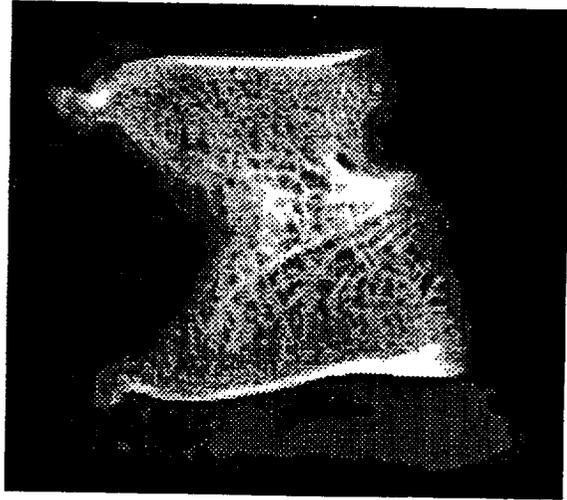
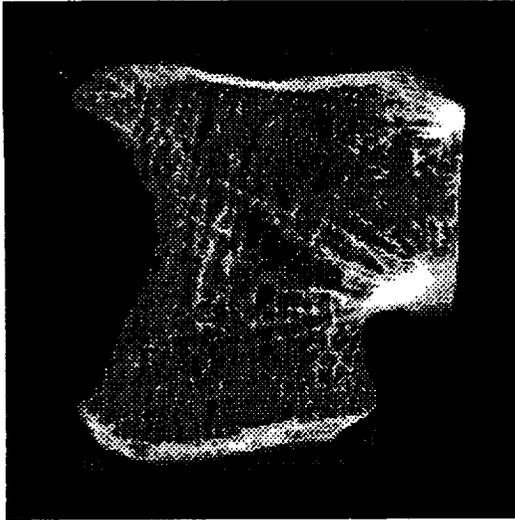


X-Ray image of patella

s/patella_pieces2/MRI-patella.1.	50	D=2.799259
s/patella_pieces2/MRI-patella.2.	50	D=2.797808
s/patella_pieces2/MRI-patella.3.	50	D=2.736876
s/patella_pieces2/MRI-patella.4.	50	D=2.694368

s/patella_pieces2/XRAY_patella.1.	50	D=2.804438
s/patella_pieces2/XRAY_patella.2.	50	D=2.750343
s/patella_pieces2/XRAY_patella.3.	50	D=2.735890
s/patella_pieces2/XRAY_patella.4.	50	D=2.622544

Fig 1: Fractal Dimension of X-Ray and MRI image



Please note that the fractal number can vary between 2 and 3.
The differences obtained between the groups is quite significant.

/* **** Results from 152_L4 stack using morphology **** */
/* **** Each region-of-interest is of dim. 110x110 **** */

		Flat	Pyramid	
es/152_L4_1.1.	110	D=2.609848	D=2.543952	Normal
es/152_L4_2.1.	110	D=2.670599	D=2.595224	
es/152_L4_3.1.	110	D=2.661320	D=2.598183	
es/152_L4_4.1.	110	D=2.667757	D=2.597461	
es/152_L4_5.1.	110	D=2.667157	D=2.593838	
es/152_L4_6.1.	110	D=2.674416	D=2.610721	

/* **** Results from 16_L4 stack using morphology **** */
/* **** Each region-of-interest is of dim. 110x110 **** */

s/16_L4_2.1.	110	D=2.478800	D=2.417768	Osteoporotic
s/16_L4_3.1.	110	D=2.523201	D=2.451674	
s/16_L4_4.1.	110	D=2.518456	D=2.447915	
s/16_L4_5.1.	110	D=2.514385	D=2.446387	
s/16_L4_6.1.	110	D=2.572345	D=2.481285	
s/16_L4_7.1.	110	D=2.572345	D=2.481302	
s/16_L4_8.1.	110	D=2.450142	D=2.402326	
s/16_L4_9.1.	110	D=2.482301	D=2.425839	

Fig 2: Fractal Dimension of Normal & Osteoporotic images.

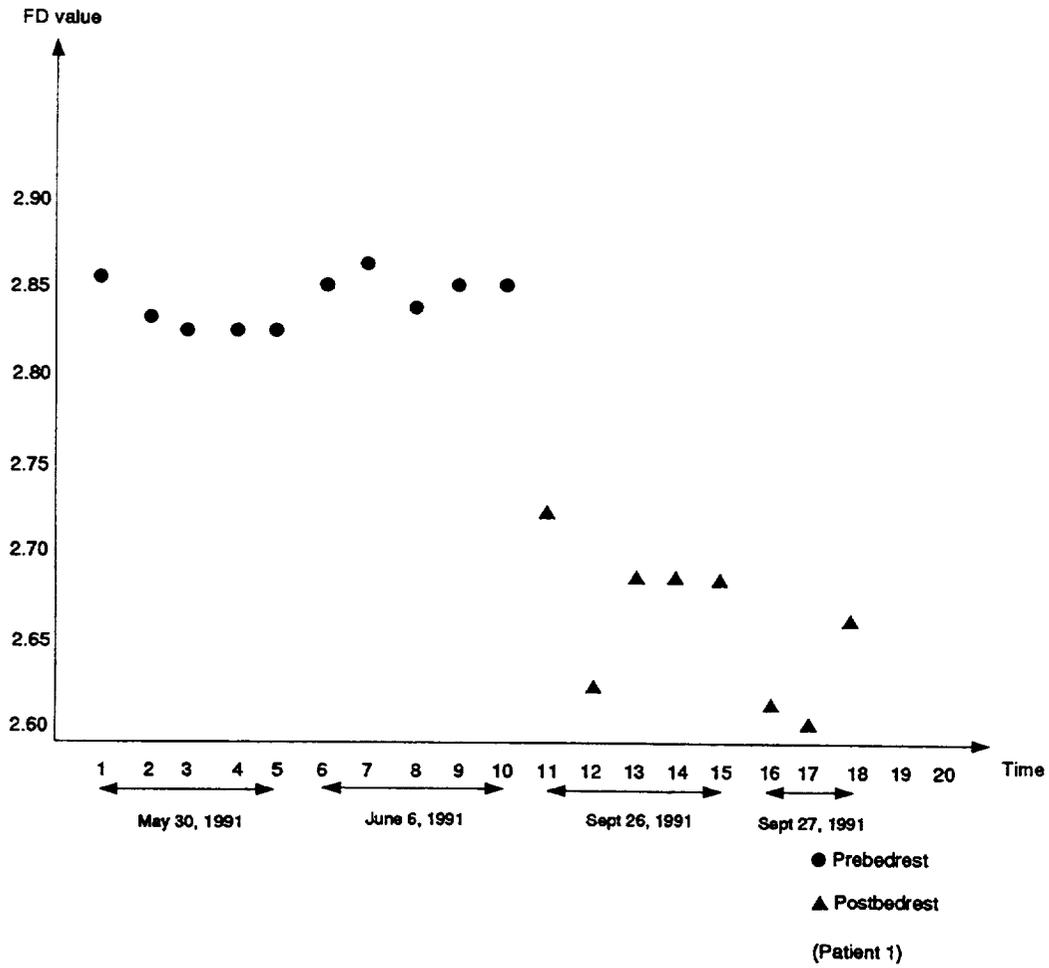


Fig 3: Bedrest Study

10 Conclusion

We use fractal dimension to characterize bone structure. We see that the fractal dimension computed with MRI images and X-Ray images of the patella are the same. We find that the fractal dimension of the Osteoporotic subjects to be less than the fractal dimension of Normal subjects. We also show that the fractal dimension of the subjects during bedrest is less than the fractal dimension before bedrest.

References

- [1] C. B. Caldwell and et al. Characterization of Mammographic Parenchymal pattern by Fractal Dimension. *Phys. Med. Biol*, 35, 1990.
- [2] R. M. Haralick. Statistical and structural approaches to texture. *Proc. IEEE* 67, pages 786–809, May 1979.
- [3] R. M. Haralick and et al. Image Analysis Using Mathematical Morphology. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 9, July 1987.
- [4] R. M. Haralick and et al. The Digital Morphological Sampling Theorem. *IEEE Trans. Acous. Speech and Signal Proc.*, 37, December 1989.
- [5] E. Hausmann, K. Allen, L. Christersson, and R. Genco. Effect of x-ray beam vertical angulation on radiographic alveolar crest measurement. *J periodontal Res.*, 24:8–19, 1989.
- [6] J. Keller and et al. Texture Description and Segmentation Through Fractal Geometry. *Comput. Vision, Graph. Image. Proc.*, 45, 1989.
- [7] W. S. Kuklinski, K. Chandra, U. E. Ruttimann, and R. Webber. Application of fractal texture analysis to segmentation of dental radiographs. In *Medical Imaging III: Image Processing*, volume 1092. SPIE, 1989.
- [8] S. Y. Lu and K. S. Fu. A syntatic approach to texture analysis. *Computer Graphics and Image Processing*, 7:303–330, 1978.
- [9] T. Lundahl, W. J. Ohley, S. M. Kay, and R. Siffert. Fractional brownian motion: A maximum likelihood estimator and it's application to image texture. *IEEE Transactions in Medical Imaging*, MI-5:152–161, Sept. 1986.
- [10] A. Lynch and et al. Analysis of texture in Macroradiographs of osteoarthritic knees using fractal signatures. *Phys. Med. Biol.*, 36, 1991.
- [11] S. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 11, July 1989.

- [12] B. B. Mandelbrot. *Fractals: Form, Chance and Dimension*. Freeman, San Fransisco, 1977.
- [13] B. B. Mandelbrot. *Fractal Geometry of Nature*. Freeman Press, 1982.
- [14] D. Marr. Analyzing natural images: A computational theory of texture vision. *Cold Spring Habor Symposia on Quantitative Biology*, XL:647–662, 1976.
- [15] A Morales and R. Acharya. Morphological Pyramids with Alternating Sequential Filters, *IEEE Trans on Image Processing* (In Press).
- [16] T. Pavlidis. A review of algorithms for shape analysis. *Computer Graphics and Image Processing*, 7:243–258, 1978.
- [17] S Pelag and et al. Multiple Resolution Texture Analysis. *IEEE Trans. Pattern Anal. Mach. Intelligence*.
- [18] S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(4):518–523, July 1984.
- [19] A. Pentland. Fractal-based description of natural scenes. In *IEEE Computer Society Conference, Computer Vision and Pattern Recognition*, pages 210–209, Washington, DC, June 1983.
- [20] A. Rosenfeld. A note on automatic detection of texture gradients. *IEEE Transactions on Computers*, 10:988–991, 1975.
- [21] U. Ruttimann and et al. Fractal Dimension from Radiographs of Peridental Alveolar Bone: A possible Diagnostic Indicator of Osteoporosis.
- [22] U. E. Ruttimann and J. A Ship. The use of fractal geometry to quantitate bone structure from radiographs. *J. Dent. Res*, page 69, 1990. Abstract no. 1431.
- [23] J Samarabandu and et al. Analysis of BOne X-Rays Using Morphological Fractals. *IEEE Trans. Med. Imag.*, 12:466–70, September 1993.
- [24] N. Sarkar and B. B. Chaudhuri. An Efficient Approach to Estimate Fractal Dimension of Texture Images. *Pattern Recog.*, 25, 1992.
- [25] J. Serra. *Image Analysis and Mathematical Morphology, vol 1*. Academic Press, 1982.
- [26] S. Sternberg. Grey Scale Morphology. *Comput. Vision, Graph. Image. Proc.*, 35, 1986.
- [27] H. Voorhees and T. Poggio. Detecting textons and texture boundaries in natural images. In *1st International Conference on Computer Vision*, pages 250–258, London, June 8–11 1987.

- [28] R. L. Webber, J. B. Hazelrig, R. J. Patel, H. R Van Den Berg, and J. E. Lemmons. Evaluation of site-specific differences in trabecular bone using fractal geometry. *J. Dent. Res.*, page 70, 1991. Abstract no. 2095.
- [29] R. Heaney, "Osteoporotic Fracture Space: An Hypothesis," *Bone and Mineral*, 6(1989), 1-13.
- [30] J. Podenphant et al, "Bone mass, bone structure and vertebral fractures in osteoporotic patients," *Bone*, 1987, 8:127-130.
- [31] M. Kleerekoper et al, "Cancellous bone architecture and bone strength," *Osteoporosis 1987*, 294-300.
- [32] J.E. Aaron et al, "The microanatomy of of trabecular bone loss in normal aging men and women," *Clin Orthop Res*, 1987, 215:260-271.
- [33] J. Samarabandu, R. Acharya, E. Hausman, C. Allen, "Analysis of Bone Images using Morphological Fractals", *IEEE Trans on Medical Imaging*, 12, Sep 1993, pp466-470.

Application of Impact Dampers in Vibration Control of Flexible Structures

Final Report
NASA/ASEE Summer Faculty Fellowship Program--1994
Johnson Space Center

Prepared by: Fred A. Akl, Ph.D. P.E.
Aamir S. Butt, M.S.

Academic Rank: Professor
Graduate Student

College and Department: Louisiana Tech University
Department of Civil Engineering
Ruston, LA 71272

NASA/JSC

Directorate: Engineering

Division: Structures and Mechanics

Branch: Loads and Structural Dynamics

JSC Colleague: David A. Hamilton

Date Submitted: August 5, 1994

Contract Number: NGT-44-005-803

ACKNOWLEDGMENT

The writers wish to thank the staff at the of the Vibration and Acoustic Test Facility (VATF) for their valuable assistance in all phases of this investigation. Our deep appreciation is extended to the NASA and Lockheed Engineering Services Corporation employees. The amount of work completed in this ten weeks period would not have been possible without their assistance.

REFERENCES

1. Chalmers, R. and S. Semercigil, "Impact Damping the Second Mode of a Cantilever Beam," *Journal of Sound and Vibration*, Vol. 146, pp. 157-161. 1991.
2. Masri, S. F., "Forced Vibration of Class of Nonlinear Dissipative Beams," *Journal of the Engineering Mechanics Division, ASCE* Vol. 99, No. EM4, Proc. Paper 9959, pp. 669-683. August 1993.
4. D.B. Halpin, D.B., "Modal Special Studies Packet - Preliminary Tests of the Impact Damper Test Article," *Vibration and Acoustics Test Facility, NASA JSC*. June 1994.
5. *Modal Testing - Theory and Applications*. D.J. Ewins, Ed. 1983. Brüel & Kjaer Instrumentation, Inc.
6. Roy, R. K., R.D. Roche, and J.E. Foster, "The application of Impact Dampers to Continuous Systems", *Trans. of the ASME, Journal of Engineering for Industry*, November, pp. 1317-1324. 1975.
7. Yousef, S. and F. Akl, "Forced Vibration Analysis of a Slender Stack Equipped with Sliding and Pendulum Impact Dampers," *Developments in Theoretical and Applied Mechanics*, Vol. 14, SECTAM XIV, University of Mississippi, pp. 433, April 1988.

TABLE 4.- MULTIPLE REGRESSION ANALYSIS OF THE RESULTS FOR THE FIRST MODE

Multiple Regression Y₁:DAMPING (%) 3 X variables

DF:	R:	R-squared:	Adj. R-squared:	Std. Error:
19	.938	.881	.858	.07

Analysis of Variance Table

Source	DF:	Sum Squares:	Mean Square:	F-test:
REGRESSION	3	.581	.194	39.399
RESIDUAL	16	.079	.005	p = .0001
TOTAL	19	.66		

Residual Information Table

SS[e(i)-e(i-1)]: e ≥ 0:		e < 0:		DW test:
.147	11	9		1.863

1

Multiple Regression Y₁:DAMPING (%) 3 X variables

Beta Coefficient Table

Parameter:	Value:	Std. Err.:	Std. Value:	t-Value:	Probability:
INTERCEPT	.315				
Im(H)/FRQ	-.017	.007	-.337	2.5	.0237
MASS*MS^2	7.362	.974	.813	7.555	.0001
GAP (mm)	.02	.005	.495	4.277	.0006

2

Multiple Regression Y₁:DAMPING (%) 3 X variables

Confidence Intervals and Partial F Table

Parameter:	95% Lower:	95% Upper:	90% Lower:	90% Upper:	Partial F:
INTERCEPT					
Im(H)/FRQ	-.032	-.003	-.029	-.005	6.248
MASS*MS^2	5.296	9.428	5.66	9.063	57.072
GAP (mm)	.01	.03	.012	.028	18.289

3

TABLE 3.- DAMPING ESTIMATES FOR THE THIRST MODE

Test File	Damping %	Mass (kg)	Gap (m)	Im(FRF) (m/s ₂ /N)	Freq. (Hz)	Φ_d
TA_11110F3	0.356	0.017	0	188.9	22.12	1.436
TB_11111F3	0.692	0.017	0.013	100.4	22.2	1.436
TB_11112F3	0.384	0.017	0.008	134.66	22.2	1.436
TA_11113F3	1.27	0.017	0.003	35.82	22.2	1.436
TB_11121F3	0.693	0.013	0.013	99.22	22.16	1.436
TB_11122F3	0.473	0.013	0.008	164.28	22.16	1.436
TA_11123F3	1.535	0.013	0.003	41.54	22.16	1.436
TA_11131F3	0.461	0.008	0.013	145.48	22.38	1.436
TB_11132F3	0.638	0.008	0.008	86.48	22.38	1.436
TA_11133F3	0.82	0.008	0.003	83.17	22.38	1.436
TD_11210F3	0.318	0.017	0	101.61	24.32	0.837
TD_11211F3	0.408	0.017	0.013	84.71	24.38	0.837
TD_11212F3	0.446	0.017	0.008	70.64	24.38	0.837
TD_11213F3	0.504	0.017	0.003	73.66	24.38	0.837
TC_11221F3	0.357	0.013	0.013	95.23	24.38	0.837
TD_11222F3	0.392	0.013	0.008	81.41	24.38	0.837
TD_11223F3	1.086	0.013	0.003	35.35	24.38	0.837
TD_11231F3	0.371	0.008	0.013	91.95	24.38	0.837
TD_11232F3	0.341	0.008	0.008	98.01	24.38	0.837
TD_11233F3	0.559	0.008	0.003	62.49	24.38	0.837

TABLE 2.- DAMPING ESTIMATES FOR THE SECOND MODE

Test File	Damping %	Mass (kg)	Gap (m)	Im(FRF) (m/s ₂ /N)	Freq. (Hz)	Φ_d
TA_11110F2	0.37	0.017	0	170.86	13.05	1.393
TA_11111F2	0.566	0.017	0.013	113.2	13.07	1.393
TA_11112F2	0.728	0.017	0.008	92.94	13.07	1.393
TA_11113F2	0.62	0.017	0.003	99.33	13.07	1.393
TA_11121F2	1.97	0.013	0.013	32.77	13.07	1.393
TA_11122F2	1.19	0.013	0.008	53.49	13.07	1.393
TA_11123F2	0.765	0.013	0.003	87.67	13.07	1.393
TA_11131F2	0.593	0.008	0.013	112.06	13.19	1.393
TA_11132F2	1.08	0.008	0.008	60.38	13.19	1.393
TA_11133F2	0.621	0.008	0.003	104.19	13.19	1.393
TA_11510F2	0.447	0.017	0	127.11	13.61	0.988
TA_11511F2	0.684	0.017	0.013	82.82	13.62	0.988
TA_11512F2	0.652	0.017	0.008	85.96	13.62	0.988
TA_11513F2	0.828	0.017	0.003	70.39	13.62	0.988
TA_11521F2	0.602	0.013	0.013	93.04	13.7	0.988
TA_11522F2	0.63	0.013	0.008	90.05	13.7	0.988
TA_11523F2	0.608	0.013	0.003	92.73	13.7	0.988
TA_11531F2	0.592	0.008	0.013	94.83	13.67	0.988
TA_11532F2	0.601	0.008	0.008	92.8	13.67	0.988
TA_11533F2	0.683	0.008	0.003	83.09	13.67	0.988

TABLE 1.- DAMPING ESTIMATES FOR THE FIRST MODE

Test File	Damping %	Mass (kg)	Gap (m)	Im(FRF) (m/s ₂ /N)	Freq.	Φ_d
TN_21000	0.336	0.017	0	72.34	4.44	1.857
TM_21111	0.976	0.017	0.013	25.66	4.44	1.857
TM_21112	0.797	0.017	0.008	33.11	4.44	1.857
TM_21113	0.503	0.017	0.003	54.83	4.44	1.857
TM_21121	0.855	0.013	0.013	29.56	4.467	1.857
TM_21122	0.59	0.013	0.008	49.86	4.467	1.857
TN_21123	0.461	0.013	0.003	63.16	4.467	1.857
TN_21131	0.617	0.008	0.013	34.68	4.5	1.857
TN_21132	0.532	0.008	0.008	54.8	4.5	1.857
TN_21133	0.391	0.008	0.003	70.81	4.5	1.857
TA_21510F1	0.31	0.017	0	41.06	5.535	0.869
TA_21511F1	0.482	0.017	0.013	30.06	5.54	0.869
TA_21512F1	0.359	0.017	0.008	37.41	5.54	0.869
TA_21513F1	0.344	0.017	0.003	38.29	5.54	0.869
TA_21521F1	0.53	0.013	0.013	24.84	5.55	0.869
TA_21522F1	0.456	0.013	0.008	30.89	5.55	0.869
TA_21523F1	0.356	0.013	0.003	39.7	5.55	0.869
TA_21531F1	0.452	0.008	0.013	30.23	5.565	0.869
TA_21532F1	0.383	0.008	0.008	34.61	5.565	0.869
TA_21533F1	0.311	0.008	0.003	43.48	5.565	0.869

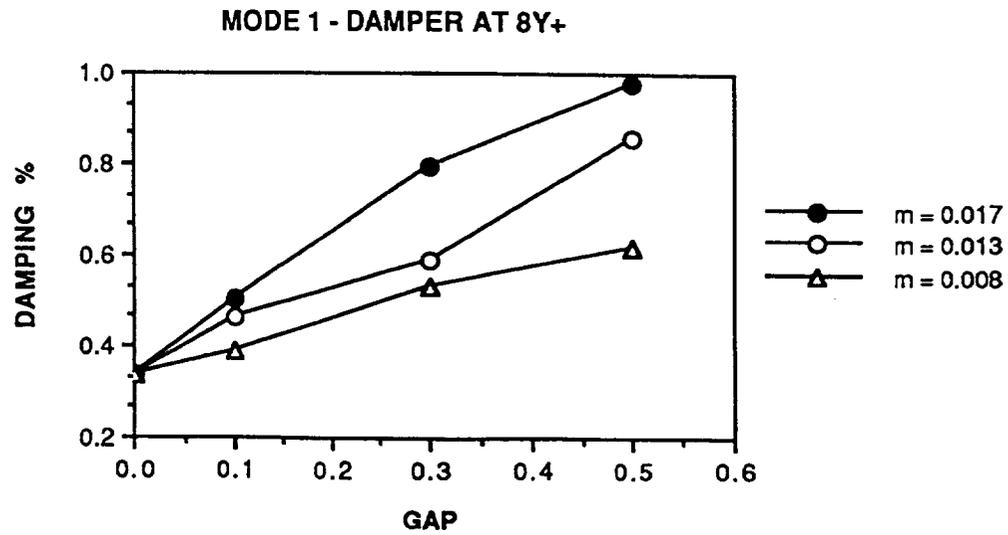
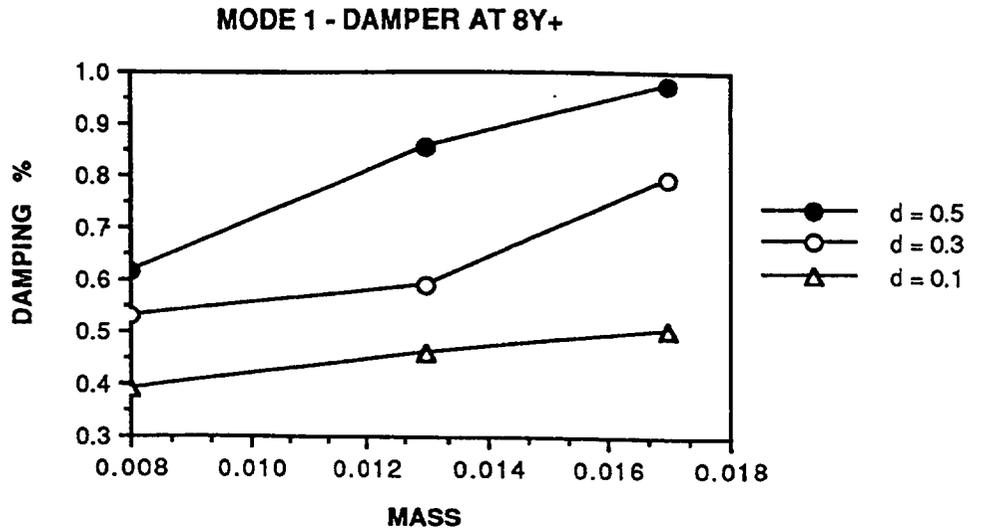


Figure 4.- Variation of Damping with Mass and Gap

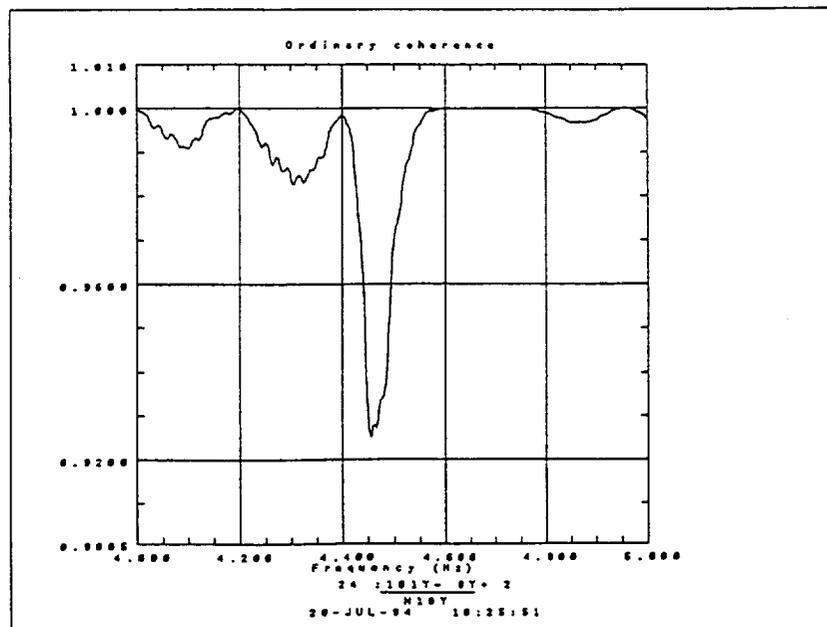
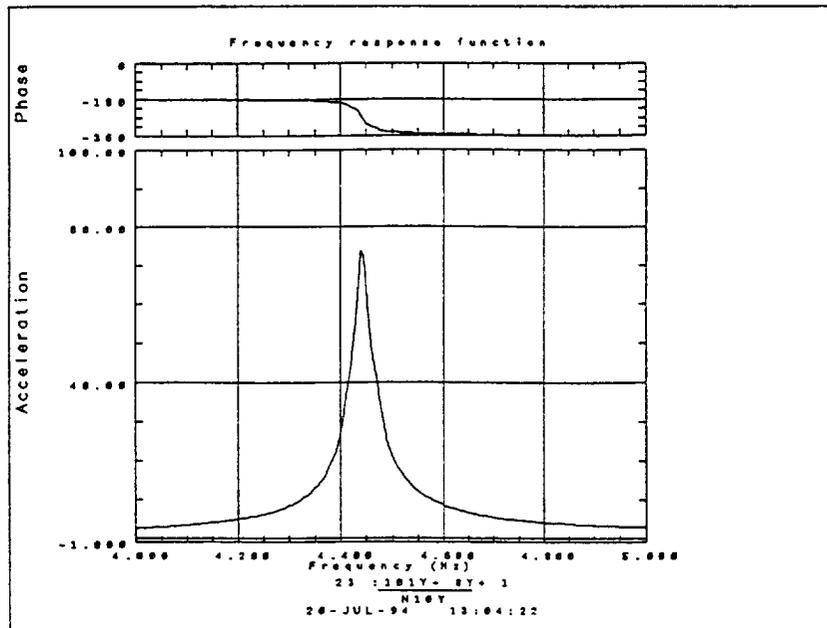
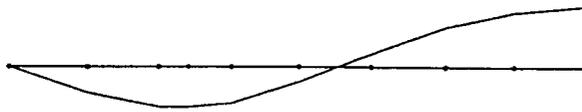


Figure 3.- Frequency Response Function and Coherence Plots for the First Mode



Mode 1 ($f_1 = 5.53$ Hz) - Spring at 81.5 in



Mode 2 ($f_2 = 13.04$ Hz) - Spring at 121.25 in



Mode 3 ($f_3 = 22.12$ Hz) - Spring at 121.25 in

Figure 2.- Modes Shapes of the Test Structure

to broaden the scope and the range of the parameters governing the behavior of flexible structures. Numerical simulation of the behavior of the problem would be needed to predict the behavior of the flexible structure under operating conditions in space.

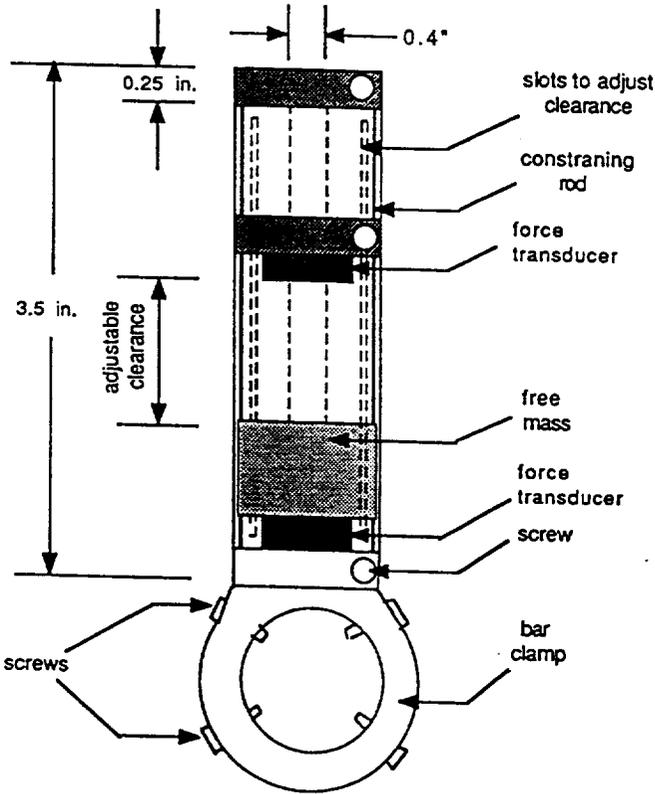


Figure 1.- Details of the Impact Damper

as follows.

$$\int_l v_i(x) dm + \int F_s dt + F_d (\Delta T) = \int_l v_f(x) dm$$

Where:

$v_i(x), v_f(x)$	the initial and final velocities, respectively.
l	the length of the test structure.
dm	an infinitesimal mass of the test structure.
F_s	the force supplied by the shaker.
$F_d(\Delta T)$	the impulse exerted by the damper.

The increase in the damping ratio of the test structure may be attributed to the presence of the third term in the L.H.S in the impulse and momentum equation above. The momentum of the beam is expected to be in the same direction before and after the impact. Therefore if the impulse is in the opposite direction to the momentum of the test article, it is then expected that the resulting momentum at the end of the half-cycle duration will be less due to the contribution of the impulse from the impact damper.

Figures 4 shows a plot of the variation of the damping ratio as a function of the moving mass of the damper and the gap. It is somewhat difficult to draw general conclusions about the behavior of the test structure from these figures.

A multiple linear regression analysis was performed using the *StatView* software. We observed, from the numerous tests we conducted, that in order for the damper to be active, it is necessary to drive the test structure at a certain level of excitation necessary to overcome friction between the damper and the guide rods on which it travels. We further hypothesize that the effectiveness of the damper depends on:

1. The velocity of the impact mass {X1 = Gap}
2. The mass multiplied by the square of the amplitude of the mode shape at the location of the damper, assuming a unit modal mass for the test structure {X2 = $m \cdot \Phi_d^2$ }, where Φ_d is the ordinate of the mode shape at the location of the damper.
3. The velocity of the test structure at the location of the damper {X3 = $\text{Im}(\text{FRF})/\text{Freq}$ }, where $\text{Im}(\text{FRF})$ is the imaginary part of the inertance frequency response function.

A multiple regression analysis based on the above hypothesis is presented in Table 4. From the statistical point view, the results indicate that the hypothesis is valid. For the first mode, the statistical analysis indicates that 88% of the variation in the damping ratio can be attributed to the independent variables {X1, X2, X3}.

CONCLUSIONS

Impact dampers can be effective in increasing the damping ratio of lightly damped flexible structures. The increase in damping is attributed to contribution of the impulse of the damper to the impulse and momentum equation of the test structure. Statistical analysis of the data obtained offers a degree of confidence in the multiple regression analysis performed in this study. It is believed that additional analysis of the data obtained can offer some insight in the behavior of the test structure. Additional experimental work on earth and in space is necessary

2. Using the multiple regression method to correlate the data.

INSTRUMENTATION

A 50-lb electromagnetic shaker was placed at a distance of 40.5 in from the center of the metal bracket. Eight accelerometers were placed approximately equally spaced along the length of the test structure (Halpin 1994). In addition a ninth accelerometer was used as a reference accelerometer at the point of excitation. A force transducer was attached to the test article at the location of the drive point to measure the input force. In addition, two force transducers were used to measure the impulsive force of the impact damper. Each steel billet was also instrumented with four accelerometers. Data acquisition was accomplished with the ZONIC 7000 48-channel data acquisition system.

TEST RESULTS AND ANALYSIS

Initial tests conducted on the test structure with impact occurring between the bare metal of the test article and the mass was recorded on magnetic tape and later digitized at rate of 10^6 samples per cycle. The results indicated that impact occurs at two impact per cycle at frequencies of 5, 10 and 20 Hz. It also showed that the impact force has the general form of an impulse with a duration of about 0.15 ms. A cushioning tape layer of about 0.1 in was used at both ends of the moving mass to widen the duration of the impact. This was necessary, first to eliminate the ringing effect of the noisy impact, and second to allow the capture of the impulse within the sampling limitations of the ZONIC 7000 system of 12,800 samples per second. The introduction of the cushioning tapes at both ends of the travelling mass resulted in impulse duration of about 1.5 ms, well within the limits of the ZONIC system.

To determine the damping ratios, twenty sine-sweep tests were performed for each of the first three natural frequencies for a total of sixty tests. Tables 1,2 and 3 show the results obtained. Six tests were conducted with the impact damper inactive, ie. the mass is restrained from moving. These tests form the baseline for comparison. It is noted that the intrinsic damping ratios ζ for the test structure are (0.336% and 0.31% for the first mode, 0.37% and 0.447% for the mode second, and 0.356% and 0.318% for third mode of vibration. Figure 2 shows the mode shapes for the test structure used in this investigation.

In performing the data acquisition, the maximum allowed block size of 4096 samples under triggered-continuous condition with 50% overlap was used. The Nyquist frequencies used were equal to 8, 16, 32 Hz for the first, second and third mode, respectively. The linear sweep rate recommended by Ewins (1983) was used to achieve accuracy.

The circle fit method was used to obtain an estimate of the damping ratio for each mode. Figure 2 shows a sample FRF and Coherence Functions. The results indicates that the parameters which was utilized in the acquisition and analysis of the data produced a good quality fit and hence an accurate estimate of the damping ratios. It is observed that the impact damper increases the damping ratio by as much as five folds.

Consider a mechanical system consisting of the test structure acted upon by the force supplied by the shaker and two impulses per cycle from the impact damper. The equation of impulse and momentum in the y-direction can be written covering a duration of one half cycle

TEST SETUP AND CONFIGURATION

The test structure consists of a 122 inch long brass tube, weighing 3.043 lb. It is simply supported at one end using a metal bracket and a universal joint. A linear helical spring is used to support the test structure at a variable distance from the simple support end. The stiffness of the spring was determined in the laboratory to be equal to 15 lb/in. The impact damper consists of a PVC cylindrical tube housing two aluminum discs at the top and bottom (Figure 1). The bottom disc is secured to a ring which fits around the brass tube with 4 screws. A sliding aluminum disc is used to adjust the distance in which the impacting mass is allowed to travel. The reader is advised to refer to Halpin (1994) for additional information related to the test setup and configuration.

The test structure was installed on two 2,000 pound steel billets. The billets were placed on steel supports to raise their levels high enough for shaker installation. The test structure was installed to the long sides of the billets for testing in the horizontal direction (Y+). One billet was used to attach the metal bracket providing the simple support for the test structure. It remained stationary throughout the test. The spring support was attached to the other billet which was moved as needed to vary the span between the two supports.

The test configurations can be summarized by the following conditions:

- a. Two spring support locations.
- b. Three impact damper locations.
- c. Three gaps.

TEST AND ANALYSIS PROCEDURES

To achieve the desired objectives of this investigation, the following test procedures were observed:

1. Check the quality of input forcing functions and driving point responses.
2. Confirm requested frequency range of 3 to 30 Hz is sufficient to obtain the desired flexural modes of test article.
3. Perform three tests at frequencies of 5, 10 and 20 Hz. Record the data on magnetic tape.
4. Digitize the data at a rate of one million per second and investigate the characteristics of the impact force at the top and bottom of the damper.
5. Determine optimum block size for data acquisition to have proper resolution for lowest test article flexural mode.
6. Find optimal input force level.
7. Document linearity and reciprocity of test structure (Halpin 1994).
8. Document repeatability of test structure from test to test.
9. Document modal characteristics of test structure.
10. Perform data quality review.
11. Perform sine sweeps of the test structure and record data.
12. Record time histories of all channels for selected tests an at selected frequencies.

Analysis of collected data was completed in two phases:

1. Using the SDRC-IDEAS software to obtain an estimate of the damping ratio using the circle fit method. The same software was used to obtain the mode shapes and natural frequencies.

INTRODUCTION

Attenuation of excessive vibration is an important design consideration in structural systems exposed to dynamic loads during their service. This can be achieved by actively or passively controlling the dynamic behavior of the structure. Methods of passive vibration control are quite often successful in achieving their objective at a relatively lower cost. They are often preferred over comparable active methods of vibration control if the structure will be used in a hostile operating environment, such as in outer space, where maintenance or monitoring is expensive.

An impact damper belongs to the category of passive vibration control devices. It consists of a free mass constrained to travel between two container walls. It produces substantial amount of damping in the structure it is attached to through momentum transfer.

Since the 1930's, numerous analytical and experimental studies have been conducted on SDOF and MDOF impact damped systems. However the experimental investigation of impact damped continuous systems was not undertaken until the early seventies.

In 1973 Masri (1973) performed experiments on a class of nonlinear dissipative cantilever and simply supported beams subjected to external sinusoidal excitation. He concluded that a heavier mass at an optimal clearance added to the systems damping. In addition the damper was more effective when located away from the node of the mode shape. In 1975 Roy and others (1975) produced similar results for fixed-fixed and simply supported beams subjected to base excitation.

Yousef and Aki (1987) performed a study of the free and forced vibration response of a vertical cantilever steel stack. Sliding and pendulum impact dampers were used in that investigation. A condition of two impacts per cycle was observed at frequencies higher than the fundamental frequency of the system. In addition it was concluded that an optimal clearance was associated with each impactor's mass.

Recently the second mode of a cantilever beam was studied by Chalmers and Semercigil (1992). They performed experiments with and without a rubber lining inside the boundaries of the damper's container. It was concluded that the rubber produced a characteristic behavior which was quite insensitive to the changes in particle clearance. Maximum damping was obtained by locating two dampers each at the locations of the largest amplitude.

OBJECTIVES

The objectives of this study are:

1. to investigate the effect of an impact damper on the dynamic behavior of a flexible structure where the effect of gravity is minimized.
2. to perform a number of tests on the flexible structure to study the effect of the mass, gap and location of the impact damper on the damping in the first three modes of vibration.
3. to analyze the data obtained in order to gain a better understanding of the behavior of the impact damper.

ABSTRACT

Impact dampers belong to the category of passive vibration devices used to attenuate the vibration of discrete and continuous systems. An Impact damper generally consists of a mass which is allowed to travel freely between two defined stops. Under the right conditions, the vibration of the structure to which the impact damper is attached will cause the mass of the impact damper to strike the structure. Previous analytical and experimental research work on the effect of impact dampers in attenuating the vibration of discrete and continuous systems have demonstrated their effectiveness. It has been shown in this study that impact dampers can increase the intrinsic damping of a lightly-damped flexible structure. The test structure consists of a slender flexible beam supported by a pin-type support at one end and supported by a linear helical flexible spring at another location. Sinusoidal excitation spanning the first three natural frequencies was applied in the horizontal plane. The orientation of the excitation and the test structure in the horizontal plane minimizes the effect of gravity on the behavior of the test structure. The excitation was applied using a linear sine sweep technique. The span of the test structure, the mass of the impact damper, the distance of travel, and the location of the impact damper along the span of the test structure were varied. The damping ratio are estimated for sixty test configurations. The results show that the impact damper significantly increases the damping ratio of the test structure. Statistical analysis of the results using the method of multiple linear regression indicates that a reasonable fit has been accomplished. It is concluded that additional experimental analysis of flexible structures in microgravity environment is needed in order to achieve a better understanding of the behavior of impact damper under conditions of microgravity. Numerical solution of the behavior of flexible structures equipped with impact dampers is also needed to predict stresses and deformations under operating conditions of microgravity in space applications.

Certification of A Hybrid Parameter Model of the Fully Flexible Shuttle Remote Manipulator System

Final Report

NASA/ASEE Summer Faculty Fellowship Program-1994

Johnson Space Center

Prepared By:	Alan A. Barhorst
Academic Rank:	Assistant Professor
University & Department	Texas Tech University Department of Mechanical Engineering Lubbock, Texas 79409-1021

NASA/JSC	
Directorate:	Engineering
Division:	Structures and Mechanics
Branch:	Loads and Structural Dynamics
JSC Colleague:	Dave Hamilton
Date Submitted:	August 9, 1994
Contract Number:	NGT-44-005-803

Section 3 not used.

ABSTRACT

The development of high fidelity models of mechanical systems with flexible components is in flux. Many working models of these devices assume the elastic motion is small and can be superimposed on the overall rigid body motion. A drawback associated with this type of modeling technique is that it is required to regenerate the linear modal model of the device if the elastic motion is sufficiently far from the base rigid motion. An advantage to this type of modeling is that it uses NASTRAN modal data which is the NASA standard means of modal information exchange. A disadvantage to the linear modeling is that it fails to accurately represent large motion of the system, unless constant modal updates are performed. In this study, which is a continuation of a project started last year, the drawback of the currently used modal snapshot modeling technique is addressed in a rigorous fashion by novel and easily applied means.

INTRODUCTION

In the spirit of continuous improvement, dynamic models of complex systems continue to improve. Included in this improvement is the current ability to model flexible systems in a modal snapshot/linear fashion. The literature is bulging with ever-improving ways to model the distributed effects [1]. There are a diverse cross-section of techniques. Some are intuitive to a design engineer [2, 3, 4, 5], while others are mathematically elegant but beyond the training of many practicing engineers [6, 7]. The purpose of this study is to further examine the efficacy of the author's attempt at developing a rigorous yet usable method for modeling complicated systems [5].

During the summer of 1993, the author began work on a rigorous quasi-automated means to model large motion [8]. This summer the task was continued and the algorithms have been more fully developed. Actual simulation and animation of simplified Remote Manipulator Systems (RMS) were generated by the quasi-automated method. Increasingly complex RMS models are being developed as a shake down tool for the algorithms.

METHODOLOGY

Present Capabilities

Based on discussions,¹ the author understands that the fidelity of the model for the present Shuttle Remote Manipulator System (RMS) simulation is limited to small amplitude vibrations about any "snap shot" configuration of the system. This limitation manifests itself because of the linear finite element (NASTRAN) model used as the progenitor for the modal basis functions. Therefore, RMS slewing maneuver studies are not within the fidelity of the linear model. There exist techniques which allow an analyst to study the slewing maneuvers of systems like the RMS, but these modeling techniques are computationally expensive and/or hard to understand [1], therefore they are not always implemented by practicing engineers. The author believes the technique discussed below gives analysts a familiar yet powerful modeling tool.

New Capabilities

The main motivating factor for the development of another modeling method was the need to easily derive complete models of complex elastic systems [1, 4, 9, 10, 11, 12, 13]. Although the method discussed herein is still relatively mathematically intense (compared to an equal number of rigid bodies), it has a predisposition (as demonstrated in the work) for symbolic manipulation. Another impetus for this work is that a simpler method may make it possible to bring rigorous flexible system modeling out of the academic domain and into use by product designers. Another catalyst for this effort is that a simple (ultimately an automated) method will make it possible for researchers to rapidly regenerate models based on new continuum assumptions.

¹Orientation meetings with various engineers from the Structures and Mechanics Division of JSC.

The approach is variational in nature. It retains most of the attributes of the analytical approach (i.e. Hamilton's principle), but eliminates most of the pitfalls, such as the need to use Lagrange multipliers for constraints, and excessive algebra. The methodology is vector based and requires the analyst to perform operations comparable to the operations required for implementing Lagrange's equations. However, it is claimed that the net algebra with the method herein will be less than the net algebra associated to Hamilton's principle or Lagrange's equations. Analysts familiar with Kane's [14] form of d'Alembert's principle will find the technique affable. The complete derivation of the method is shown elsewhere [1, 5].

Closed Form Model

Why should an analyst develop closed form models when there exist other tools that seem to adequately model these systems? The author believes that using tools that are traditionally from the structural analysis realm such as NASTRAN models unnecessarily limit the model to the linear motion about some configuration. It is felt that if the approach of writing complete models first (then reducing to linear if desired) is feasible, in a timely manner, then engineers will utilize these more exact models. In order to facilitate the clock, computer aided modeling is desired; *Mathematica* [15] is an excellent tool for this process. Another advantage to working directly with the closed form model is that the "zero times zero" multiplications that arise in straight out matrix models are avoided. Also repetitive multiplications and additions are readily recognized and can be assigned to a memory location for instant recall. This tight code will make running these complicated models more feasible.

***Mathematica* Algorithms**

Mathematica [15] algorithms were developed to mimic the procedure outlined in the previous work [5]. The standard notation for *Mathematica* was adjusted so as to mimic engineering vector notation. Then algorithms were developed that recognize the vector dot and cross products, the triple products, and other identities. Differentiation of vectors in multiple coordinate frames was defined. Standard order for the symbols was defined so symbolic cancelation was facilitated. Functions that aid in the gathering of terms, the distribution of terms, and general manipulation were developed. At this point these algorithms are used via a *Mathematica* notebook running on a NeXT computer. They are not limited to this computer system because the notebooks are portable across multiple computer systems. An example of how one enters symbols for manipulation is shown in the report [8].

RMS MODELS

The first model of the RMS studied this summer was a two link rigid model that was used to shake down the modeling algorithms and fine tune the symbolic manipulation functions. It was also used to iron out means to show animations of the device, which is actually trivial with the aid of the *Mathematica* notebook front-end. The model and animation worked well. The simulation was done in *Mathematica*.

The next model studied, was a simple beam model of a single link planar manipulator. The beam neutral axis served as a rotating frame, and all flexibility was referenced to the rotating frame. The weak formulation of the field equations was utilized. The boundary conditions are rigorously supplied by the underlying modeling method. The modeling algorithms worked well. The equations of motion were output in FORTRAN form and simulated external to *Mathematica*. The output variable were then re-loaded into *Mathematica* for plotting and animation.

The third model studied was a two flexible link model of the RMS. The flexibility was modeled with simple beam models (weak formulation), referenced to rotating neutral axes. The automated modeling went well, simulation was done external to *Mathematica*. The simulation consisted of 16 coupled ordinary first order differential equations in the form:

$$A\dot{u} = f$$

where A is a full matrix and is a nonlinear function of time and the configuration coordinates. The right-hand side f is nonlinear function of time, and the coordinates and speeds of the system. Two cases were studied, with data not consistent with RMS data, in a demonstrative fashion. The system is a flexible double pendulum. Snapshots of the simulation are shown in figure 1, figure 2, and figure 3. The progression is from top to bottom than left to right. Figure 1 shows a case where the elastic motion stays within the realm of the simple beam model. Figure 2 shows the motion of the root beam in the rotating frame. Figure 3 shows a case where the elastic motion is large. The accuracy of the beam model is suspect in this case. It is presented to show the large overall motion simulation capability. Both of these cases were run on Motorola 68040 hardware.

The fourth model was an attempt to bring a NASTRAN modal model of the beams from the model above. This work is still in progress. It is used as a building block to be able to represent the flexible continuum of any body with NASTRAN modal basis functions.

The fifth model is a three link planar system with data from the RMS. This effort is incomplete and will be used to iron out future problems. This will aid in the verification of the full flex model alluded to in the work from last year [8].

SUMMARY

This project was a continuation of an effort to enable computer assisted modeling of systems of flexible bodies. The resulting models are not restricted to linear ranges of slewing maneuvers. An automated means to write the equations of motion is nearing completion. Much progress was made during the last two summers, but much work is still needed.

REFERENCES

- [1] A. A. Barhorst, *On Modeling the Dynamics of Hybrid Parameter Multiple Body Mechanical Systems*. PhD thesis, Texas A & M University, College Station, Texas, 1991.
- [2] D. R. Glandorf, "Equations of motion for the station control simulator, memo 23338," Tech. Rep., Lockheed EMSCO, December 1986.
- [3] T. R. Kane, R. R. Ryan, and A. K. Banerjee, "Dynamics of a cantilever beam attached to a moving base," *Journal of Guidance, Control, and Dynamics*, vol. 10, pp. 139–151, March-April 1987.
- [4] L. Meirovitch, "State equations of motion for flexible bodies in terms of quasi-coordinates," in *Proc. IUTAM/IFAC Symposium on Dynamics of Controlled Mechanical Systems*, (Berlin), pp. 37–48, Springer-Verlag, 1988.
- [5] A. A. Barhorst, "Systematic closed form modeling of hybrid parameter multiple body systems," in *Proceedings of the ASME Winter Annual Meeting*, (New Orleans, LA), November 1993.
- [6] J. C. Simo, "Nonlinear dynamics of flexible structures, a geometrically exact approach," in *Proc. of the Workshop on Multibody Simulation*, pp. 235–296, JPL D-5190, April 1988.
- [7] M. G. Nasser, *Numerical Methods for Multibody Elastic Systems and Contact*. PhD thesis, The University of Houston, Houston, Texas, 1992.
- [8] A. A. Barhorst, "Modeling the shuttle remote manipulator system—another flexible model," Final Report, NASA/ASEE JSC Summer Faculty Fellows Program, August 1993. Contract # NGT-44-001-800.
- [9] K. W. Lips and R. P. Singh, "Obstacles to high fidelity multibody dynamics simulation," in *Proc. of the American Control Conference*, pp. 587–594, June 1988.
- [10] P. W. Likins, "Multibody dynamics: A historical perspective," in *Proc. of the Workshop on Multibody Simulation*, pp. 10–24, JPL D-5190, April 1988.
- [11] K. H. Low and M. Vidyasagar, "A lagrangian formulation of the dynamic model for flexible manipulator systems," *Journal of Dynamic Systems, Measurement and Control*, vol. 110, pp. 175–181, June 1988.
- [12] S. Lee and J. L. Junkins, "Explicit generalizations of lagrange's equations for hybrid coordinate dynamical systems," Tech. Rep. 91-0301, Texas A&M University, Aerospace Engineering Department, March 1991.
- [13] J. C. Simo and L. Vu-Quoc, "On the dynamics in space of rods undergoing large motions—a geometrically exact approach," *Computer Methods in Applied Mechanics and Engineering*, vol. 66, pp. 125–161, 1988.

- [14] T. R. Kane and D. A. Levinson, *Dynamics Theory and Applications*. New York: McGraw-Hill, 1985.
- [15] S. Wolfram, *Mathematica—A System for Doing Mathematics by Computer*. Redwood City, California: Addison-Wesley, 2 ed., 1991.

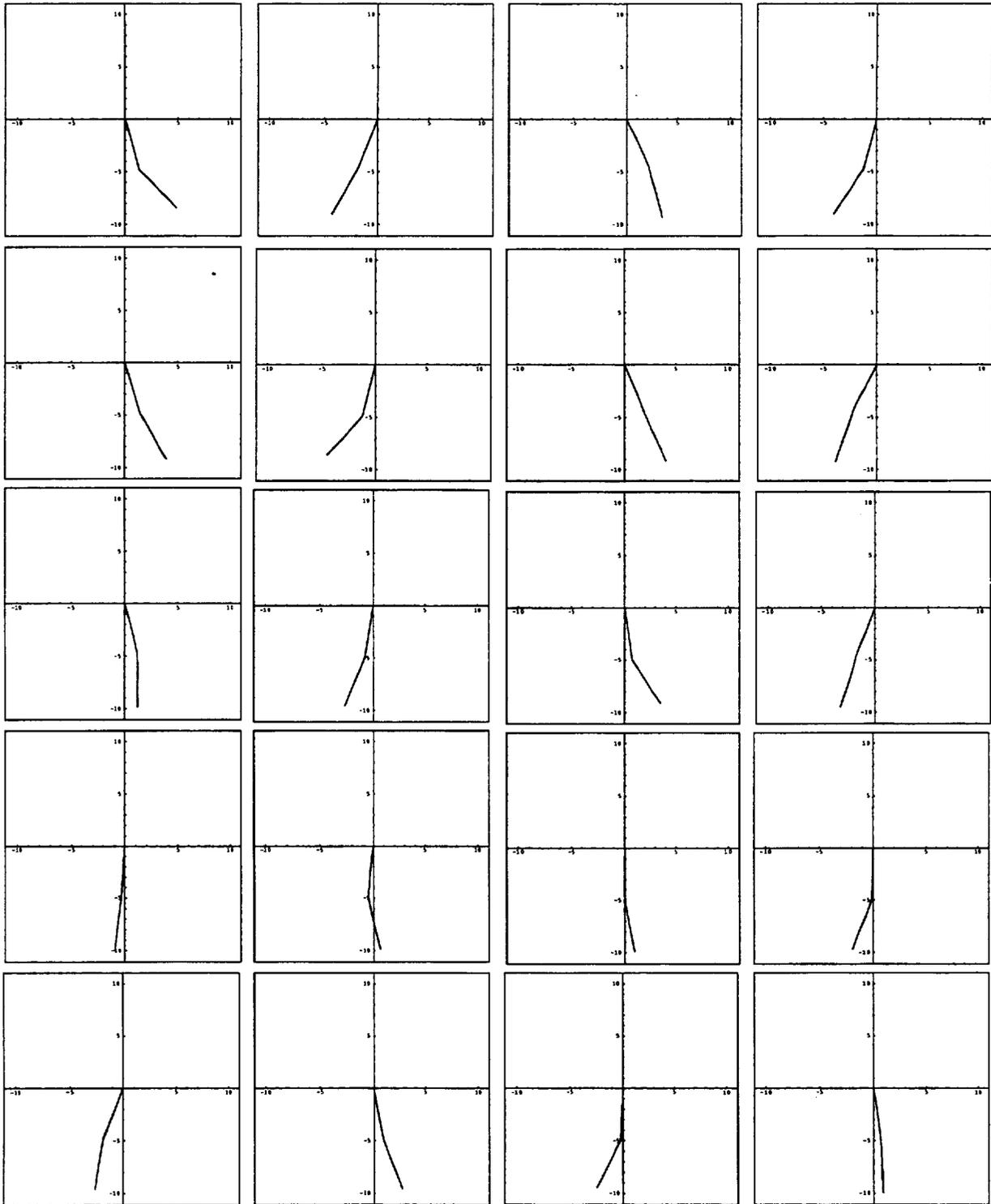


Figure 1: Two Link Flexible Manipulator–Low Elastic Energy

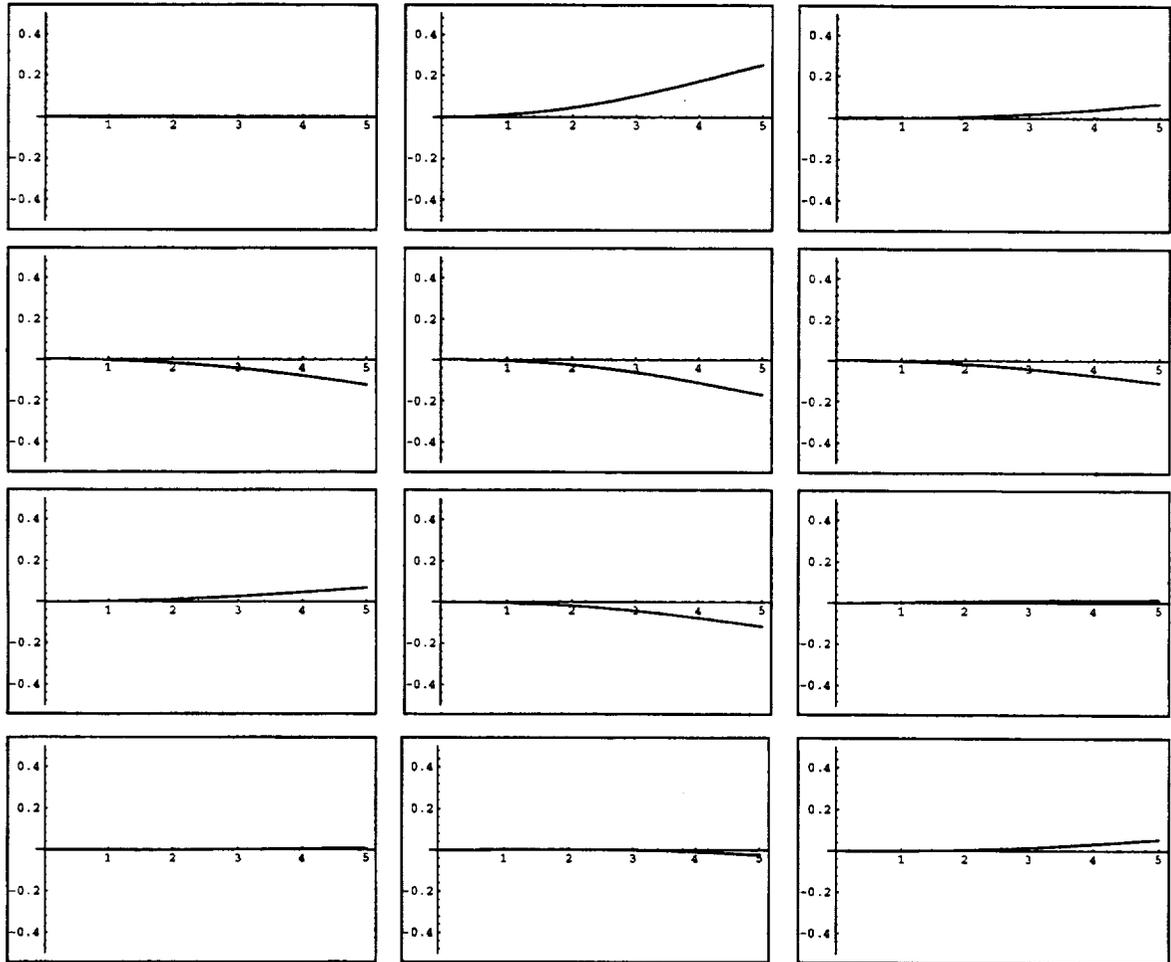


Figure 2: Base Link Flexible Motion—Low Elastic Energy

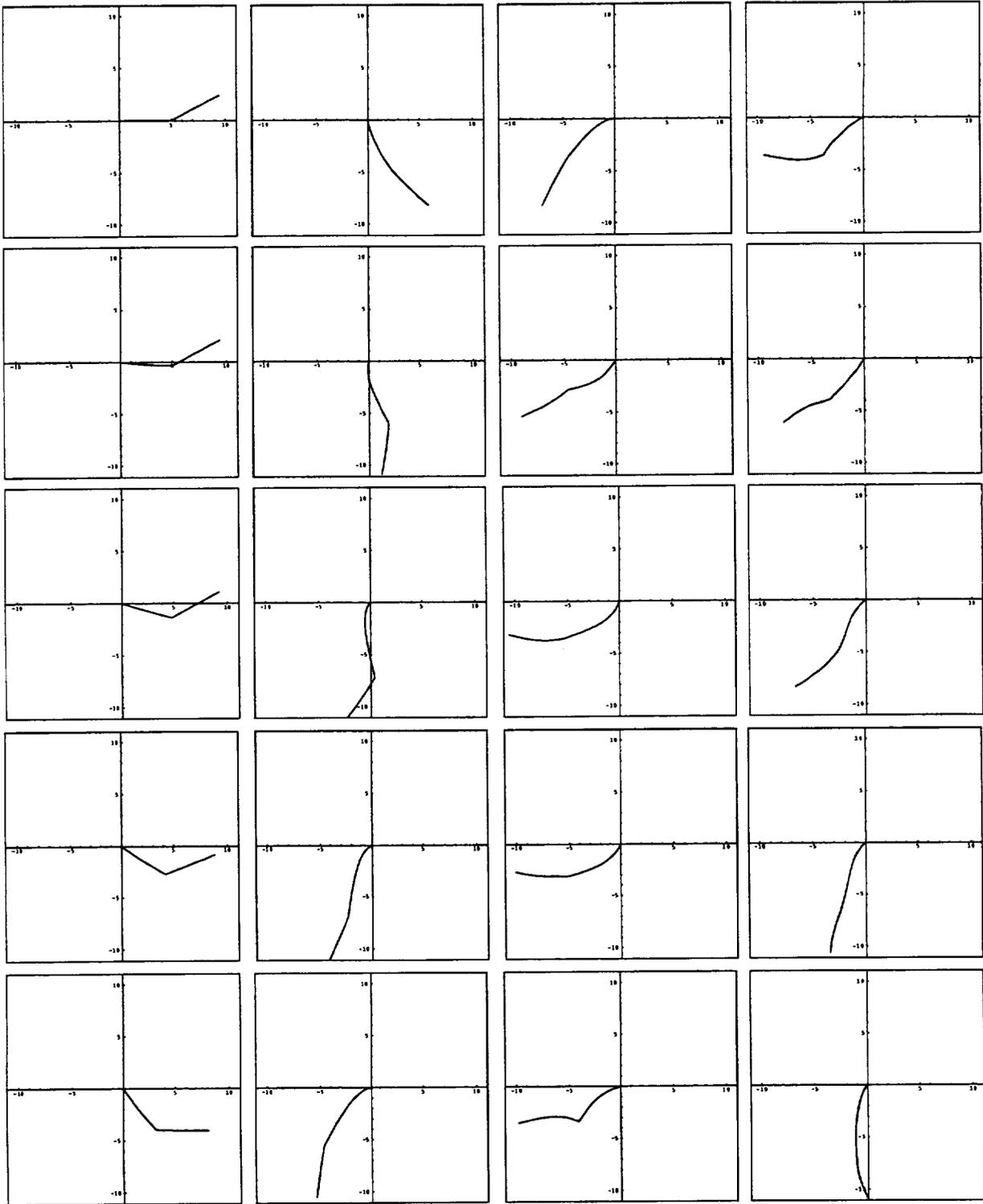


Figure 3: Two Link Flexible Manipulator-High Elastic Energy

54-53
4967

THE ROLE OF THE VESTIBULAR SYSTEM IN MANUAL TARGET LOCALIZATION

Final Report
NASA/ASEE Summer Faculty Fellowship Program -- 1994
Johnson Space Center

Prepared By: Susan R. Barry, Ph.D.
S. Alyssa Mueller

Academic Rank: Assistant Professor,
Student

University & Department: Mount Holyoke College
Dept. of Biological Sciences
South Hadley, MA. 01075

NASA/JSC

Directorate: Space and Life Sciences

Division: Medical Sciences

Branch: Space Biomedical Research Institute

JSC Colleague Jacob Bloomberg, Ph.D.

Date Submitted: September 2, 1994

Contract Number: NGT-44-005-803

Approved By: Jacob Bloomberg

Date Approved: Sept. 2, 94

ABSTRACT

Astronauts experience perceptual and sensory-motor disturbances during spaceflight and immediately after return to the 1-g environment of Earth. During spaceflight, sensory information from the eyes, limbs and vestibular organs is reinterpreted by the central nervous system so that astronauts can produce appropriate body movements in microgravity. Alterations in sensory-motor function may affect eye-head-hand coordination and, thus, the crewmember's ability to manually locate objects in extrapersonal space. Previous reports have demonstrated that crewmembers have difficulty in estimating joint and limb position and in pointing to memorized target positions on orbit and immediately postflight.

One set of internal cues that may assist in the manual localization of objects is information from the vestibular system. This system contributes to our sense of the body's position in space by providing information on head position and movement and the orientation of the body with respect to gravity. Research on the vestibular system has concentrated on its role in oculo-motor control. Little is known about the role that vestibular information plays in manual motor control, such as reaching and pointing movements. Since central interpretation of vestibular information is altered in microgravity, it is important to determine its role in this process.

This summer, we determined the importance of vestibular information in a subject's ability to point accurately toward a target in extrapersonal space. Subjects were passively rotated across the earth-vertical axis and then asked to point back to a previously-seen target. In the first paradigm, the subjects used both visual and vestibular cues for the pointing response, while, in the second paradigm, subjects used only vestibular information. Subjects were able to point with 85% accuracy to a target using vestibular information alone. We infer from this result that vestibular input plays a role in the spatial programming of manual responses.

INTRODUCTION

Astronauts experience perceptual and sensory-motor disturbances during spaceflight and immediately after return to the 1-g environment of Earth (Young et al., 1984). During spaceflight, sensory information from the eyes, limbs and vestibular organs is reinterpreted by the central nervous system so that astronauts can produce appropriate body movements in microgravity. Alterations in sensory-motor function may affect eye-head-hand coordination and, thus, the crewmember's ability to manually locate objects in extrapersonal space. Previous reports have demonstrated that crewmembers have difficulty in estimating joint and limb position and in pointing to memorized target positions on orbit and immediately postflight (Watt et al., 1985).

The ability to point or reach toward an object or perform other manual tasks is essential for safe Shuttle operation and may be compromised particularly during re-entry and landing sequences and during possible emergency egress from the Shuttle. An understanding of eye-head-hand coordination and the changes produced during space flight is necessary to develop effective countermeasures. This summer's project expanded upon last summer's research on the sensory cues used in the manual localization of objects.

Last summer, we determined that people use an egocentric, as opposed to an allocentric, reference frame to point toward a target. In an egocentric reference frame, the object is localized in relation to the position of the subject's body while in an allocentric reference frame, the target is localized in relation to other objects in the external visual world. Thus, people use internal, egocentric cues, such as the direction of gaze and the position of the limbs, to locate objects in extrapersonal space. One set of internal cues that may assist in the manual localization of objects is information from the vestibular system. Since central interpretation of vestibular information is altered in microgravity, it is important to determine its role in this process.

Vestibular receptors include the otolith organs and semicircular canals located in the inner ear. This system contributes to our sense of the body's position in space by providing information on head position and movement and the orientation of the body with respect to gravity. Along with visual, auditory and proprioceptive cues, vestibular input contributes to an internal spatial map within the central nervous system of body position.

Healthy human subjects can use vestibular information to estimate, without the aid of vision, the magnitude of self-angular displacement when briefly rotated about an earth-vertical axis. Bloomberg et al., (1991) demonstrated that subjects, who are rotated in total darkness, can make accurate saccadic eye movements back to a previously seen earth-fixed target. Since the rotation and the saccadic eye movements occurred in total darkness, vestibular cues must have been used to update the internal spatial map, code the target position in space, and generate accurate eye movements. Research on the vestibular system has concentrated on its role in oculo-motor control.

Little is known about the role that vestibular information plays in manual motor control, such as reaching and pointing movements.

This summer, we determined the role of vestibular information in a subject's ability to point accurately toward a target in extrapersonal space.

Subjects were passively rotated about the earth-vertical axis and then asked to point back to a previously-seen target. In the first paradigm, the subjects used both visual and vestibular cues for the pointing responses, while in the second paradigm, subjects used only vestibular information. Subjects were able to point with 85% accuracy to a target using vestibular information alone. We infer from this result that vestibular input plays a role in the spatial programming of manual responses.

METHODS

Subjects:

Eight healthy subjects, five males and three females, ranging in age from 21 to 40 were tested. Five subjects were right-handed while three were left-handed.

Experimental set-up:

In order to measure pointing accuracy, subjects were seated in a rotatable chair located two meters from the center of a screen. The target was illuminated on the center of the screen at a height of 56 inches, average eye-level for a seated subject.

A laser for pointing was mounted onto a plastic finger splint. The tip of the laser rested on a ball joint so that it could be rotated in any direction. The laser was secured onto the index finger of the dominant hand of the subject with Velcro™ straps. For each test, subjects practiced pointing with the finger-tip laser at the illuminated target on the screen and adjusted the position of the laser. Subjects were able to match the laser beam location with the perceived pointing location.

The target was initially displayed on a computer monitor and then projected onto the viewing screen using an overhead projector equipped with a special display panel (Proxima Corporation, San Diego, CA.). The display panel possessed an auxiliary scanning device that was used to record the laser beam spot on the display screen when the subject pointed with the laser at the remembered target position. The display window had a resolution of 640 units horizontal and 480 units vertical. Software was written by Dr. William Huebner to display the target momentarily on the screen and then to record the coordinates of the laser spot.

The subjects were seated in a rotatable chair whose movements were controlled by a servo-motor and computer. The chair was rotated at a rate of 40 degrees per second to displacements of 0 degrees, +/-10 degrees, +/-20 degrees, or +/- 30 degrees. In each experiment, the chair was rotated ten times to each of the seven possible displacements for a total of seventy rotations. The order of the displacements was random but consistent across

subject population. A search coil system was used to verify the position of the chair before and after each rotation.

During the experiment, the subject was placed in a comfortable head restraint to prevent any head movement independent of the movement of the chair. The subject wore earphones to receive verbal cues during the trials, such as when to fixate the target and when to point. The earphones also masked outside auditory signals in the room which might have given the subject cues as to rotational displacement or where to point.

Experimental protocols:

Each subject was tested in two experimental paradigms.

Paradigm one: The subject was asked to fixate his/her gaze at the target on the illuminated screen. The chair was then rotated while the subject maintained his/ her gaze at the target. The projector light was then extinguished, and the subject used the finger-fixed laser to point back at the remembered location of the target on the screen. For each displacement, the subject pointed three times at the remembered target location. Second and third attempts were corrective motions.

Paradigm two: The subject was asked to fixate the target on the illuminated screen. The projector light was then extinguished. The chair was rotated in total darkness, and, after rotation, the subject pointed back to the remembered location of the target. For each displacement, the subject pointed three times at the remembered target location.

For all eight subjects, paradigm one was always tested before paradigm two. At least two days elapsed for each subject between tests.

In both paradigms, the subjects received minimal feedback of how accurately they had pointed. In paradigm one, however, the subject could use both visual and vestibular cues to locate the position of the target during rotation. In paradigm two, the subject could use only vestibular cues during and after rotation to locate the position of the target and to point.

Data analysis:

For each pointing trial, the linear deviation of the laser point from the x-axis of the target was measured. These data were used to calculate angular deviation from the target given a constant distance from the chair's center of rotation to the target (Figure 1). This deviation was defined as the pointing error. The values of pointing error for the first pointing attempt at each displacement in a given experiment were averaged. Mean pointing errors were calculated in an identical fashion for the second and third pointing attempts. Data are displayed as mean pointing error +/- standard error of the mean. Mean values of pointing error were compared using a two-tailed Student's T-test. A p value of less than 0.05 indicated a significant difference between the means.

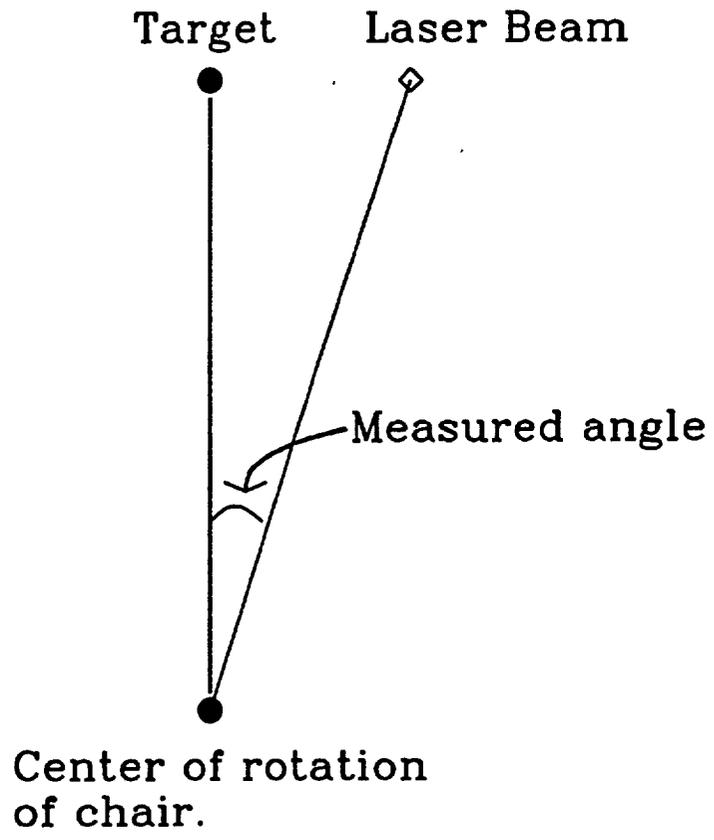


Figure 1.— Angular deviation of laser beam relative to target.

RESULTS

For both paradigms, subjects were rotated both clockwise and counterclockwise for 10, 20 or 30 degree rotations. Thus, following 50% of the rotations, the subjects pointed away from their body to aim at the remembered location of the target. For the other 50% of trials, the subjects pointed across their bodies to aim at the target position. No significant differences were found in pointing accuracy between crossed and uncrossed pointing movements. No differences in performance were noted between left-handed and right-handed subjects.

In paradigm one, subjects were rotated while they maintained gaze on the stationary target. After the rotation was complete, the target light was extinguished, and the subjects pointed back to the remembered location of the target. For the first attempt following a ten degree rotation, subject pointing error averaged 1.3 ± 0.3 degrees; for a twenty degree rotation, pointing error averaged 1.6 ± 0.2 degrees, and for a thirty degree rotation, pointing error averaged 2.4 ± 0.5 degrees (Figure 2). Since subjects pointed three times with each chair rotation, we were able to observe whether any improvement occurred with subsequent pointing attempts. For the ten and twenty degree rotations, a significant improvement was found in pointing accuracy between the first and second attempts for paradigm one.

Pointing error was also normalized against the angle of chair rotation. For a ten degree rotation, subjects erred by 0.13 degrees per degree of rotation. For both the twenty and thirty degree rotations, subject erred by only 0.08 degrees per degree of rotation (Figure 3).

In paradigm two, subjects fixated on the target as in paradigm one. However, the target light was extinguished before rotation, and the subject pointed back to the remembered location of the target once the rotation was complete. Pointing error was greater in this paradigm. For a ten degree rotation, subject pointing error averaged 2.1 ± 0.4 degrees; for a twenty degree rotation, pointing error averaged 3.1 ± 0.6 degrees; for a thirty degree rotation, pointing error averaged 4.7 ± 0.6 degrees (Figure 4). In contrast to paradigm one, no significant improvement in pointing accuracy was observed with subsequent pointing attempts.

Normalized pointing error were 0.21 degrees per degree of rotation for a ten degree rotation and 0.15 degrees per degree of rotation for both twenty and thirty degree rotation (Figure 5).

For both paradigms one and two, subjects tended to undershoot the target. This was true regardless of whether crossed or uncrossed pointing movements were used.

Finally, there was no significant difference between the pointing error for a zero degree rotation between paradigm one and two. In paradigm one, pointing error averaged 0.85 ± 0.28 degrees while in paradigm two, pointing error averaged 0.87 ± 0.21 degrees for the first attempts (Figure 6). In the absence of rotation, the pointing task for paradigm one is identical to that for paradigm two so that the measurements of pointing error should not be and are not different.

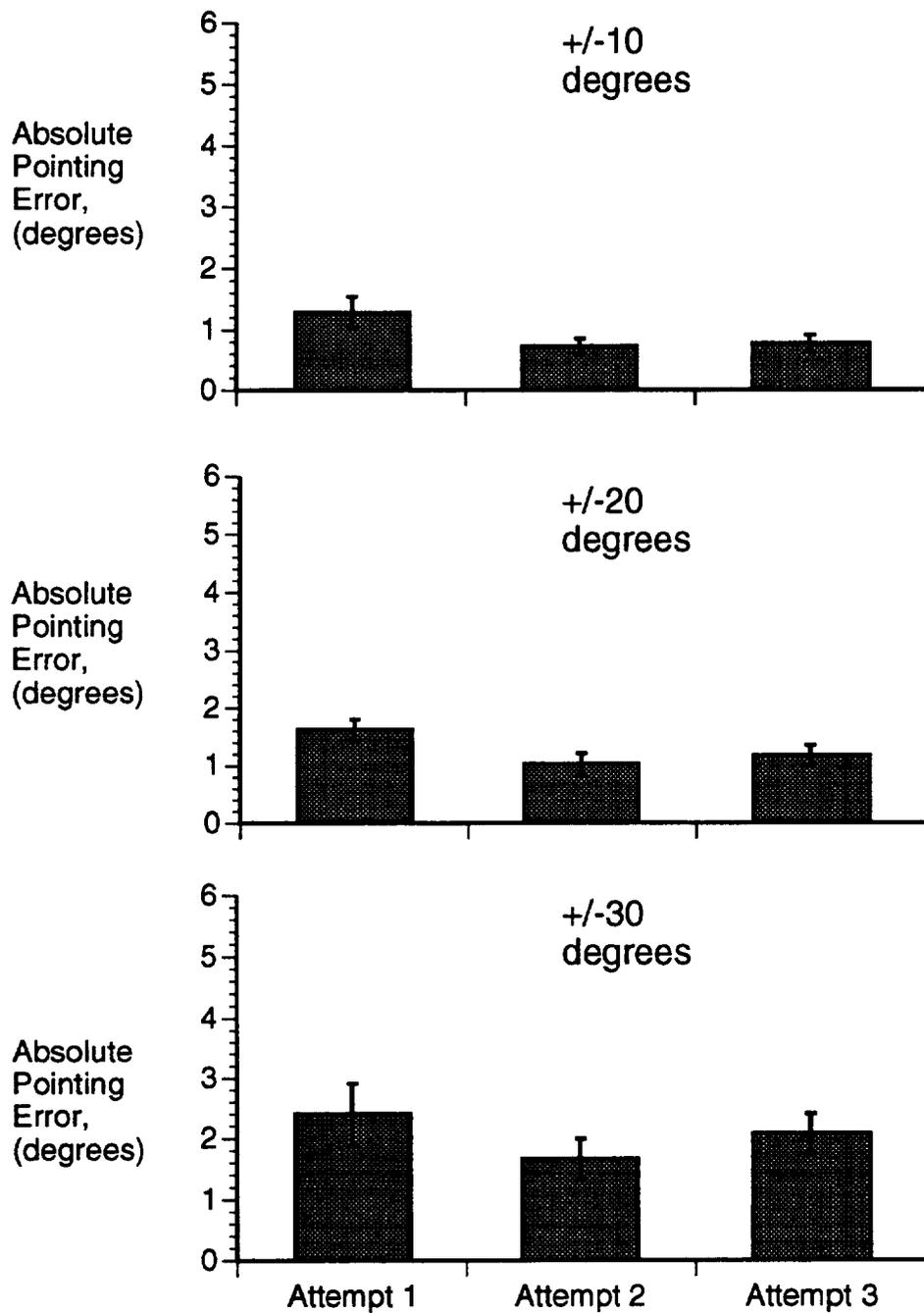


FIGURE 2.- Absolute pointing error for paradigm one (rotation in the light) for 10, 20, 30 degree displacements. Data represent mean +/- SEM for 8 subjects.

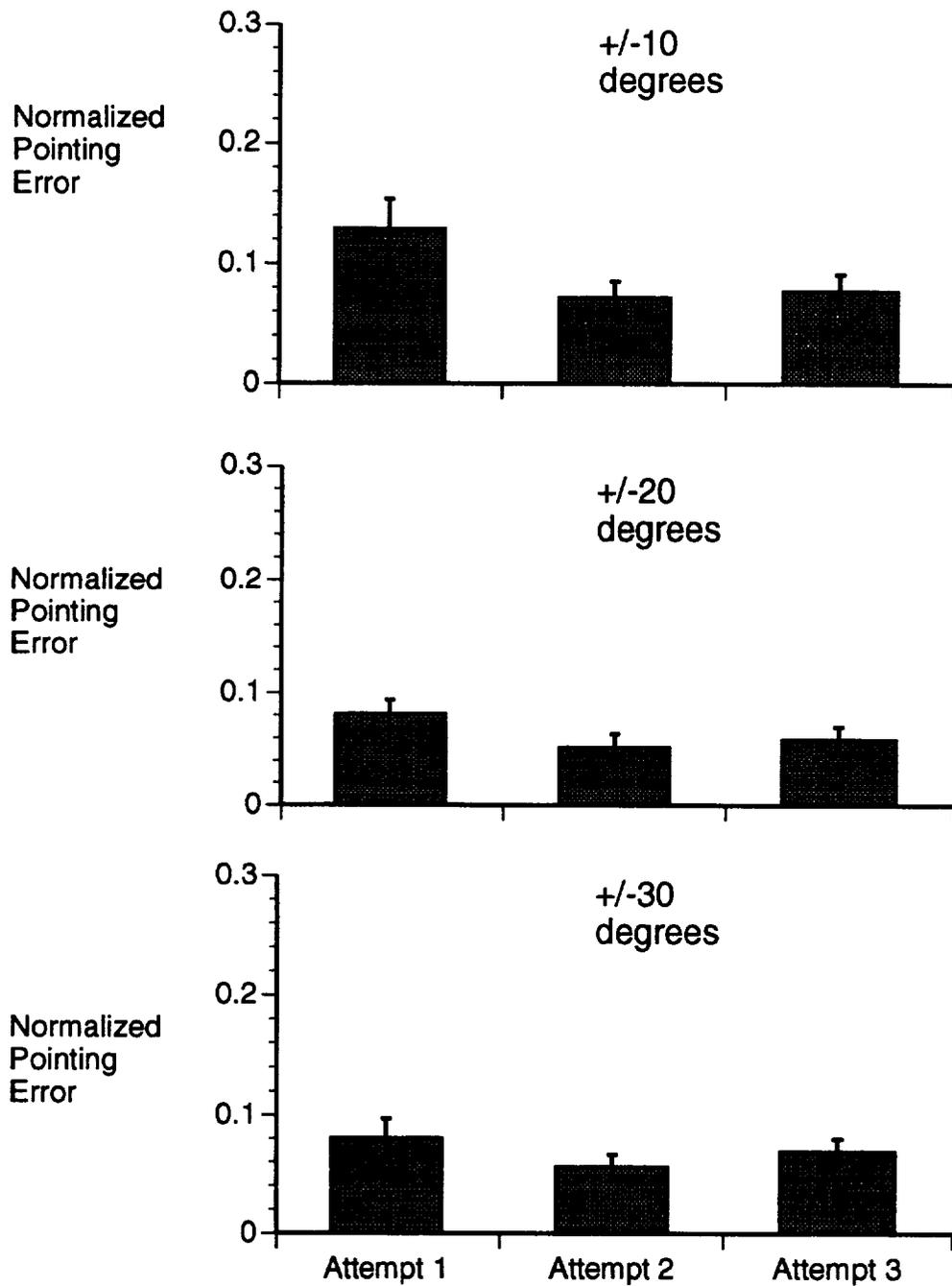


FIGURE 3.- Normalized pointing error for paradigm one (rotation in the light) for 10, 20 and 30 degree displacements. Data represent mean \pm SEM for 8 subjects.

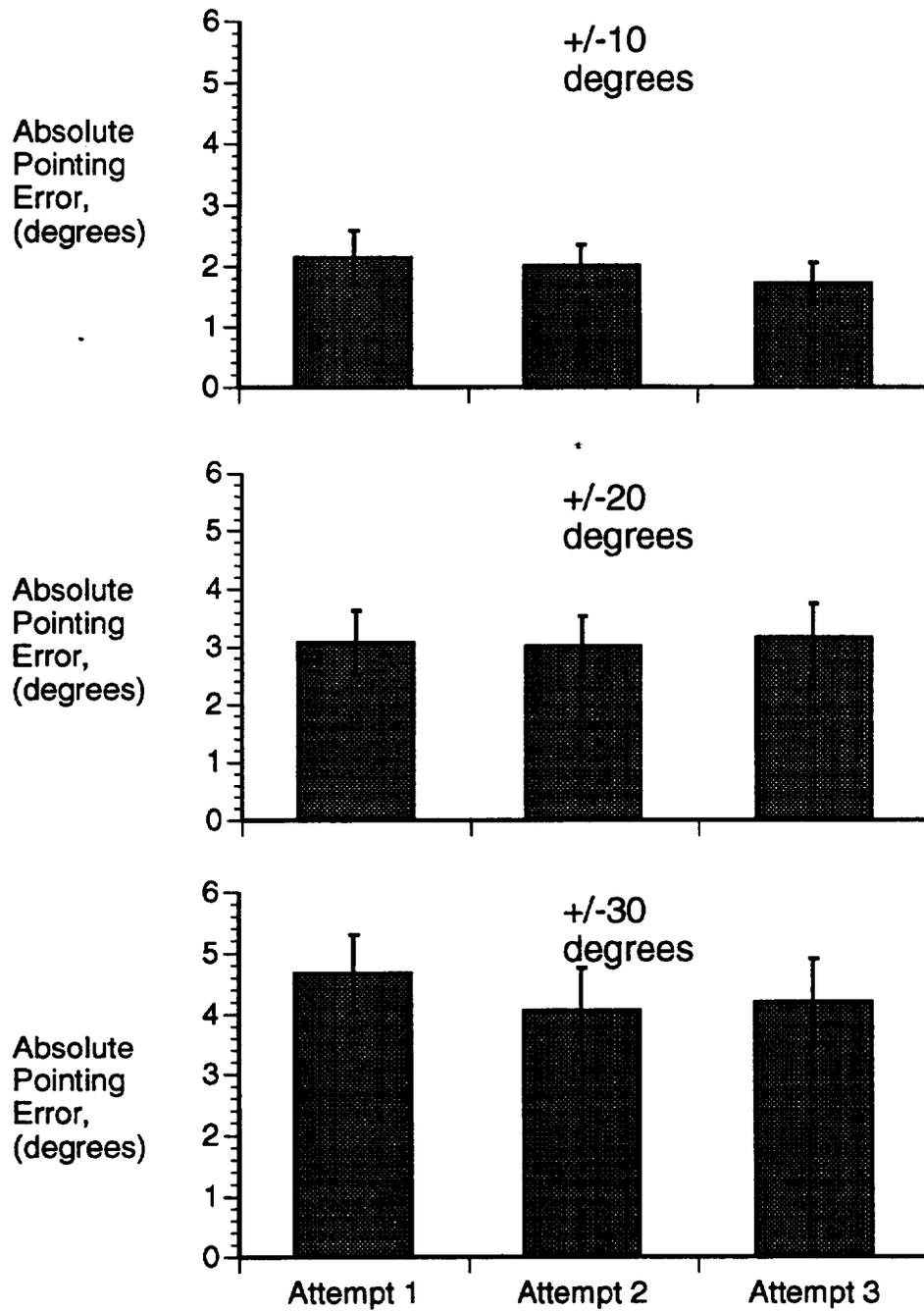


FIGURE 4.- Absolute pointing error for paradigm two (rotation in the dark) for 10, 20 and 30 degree displacements. Data represent mean \pm SEM for 8 subjects.

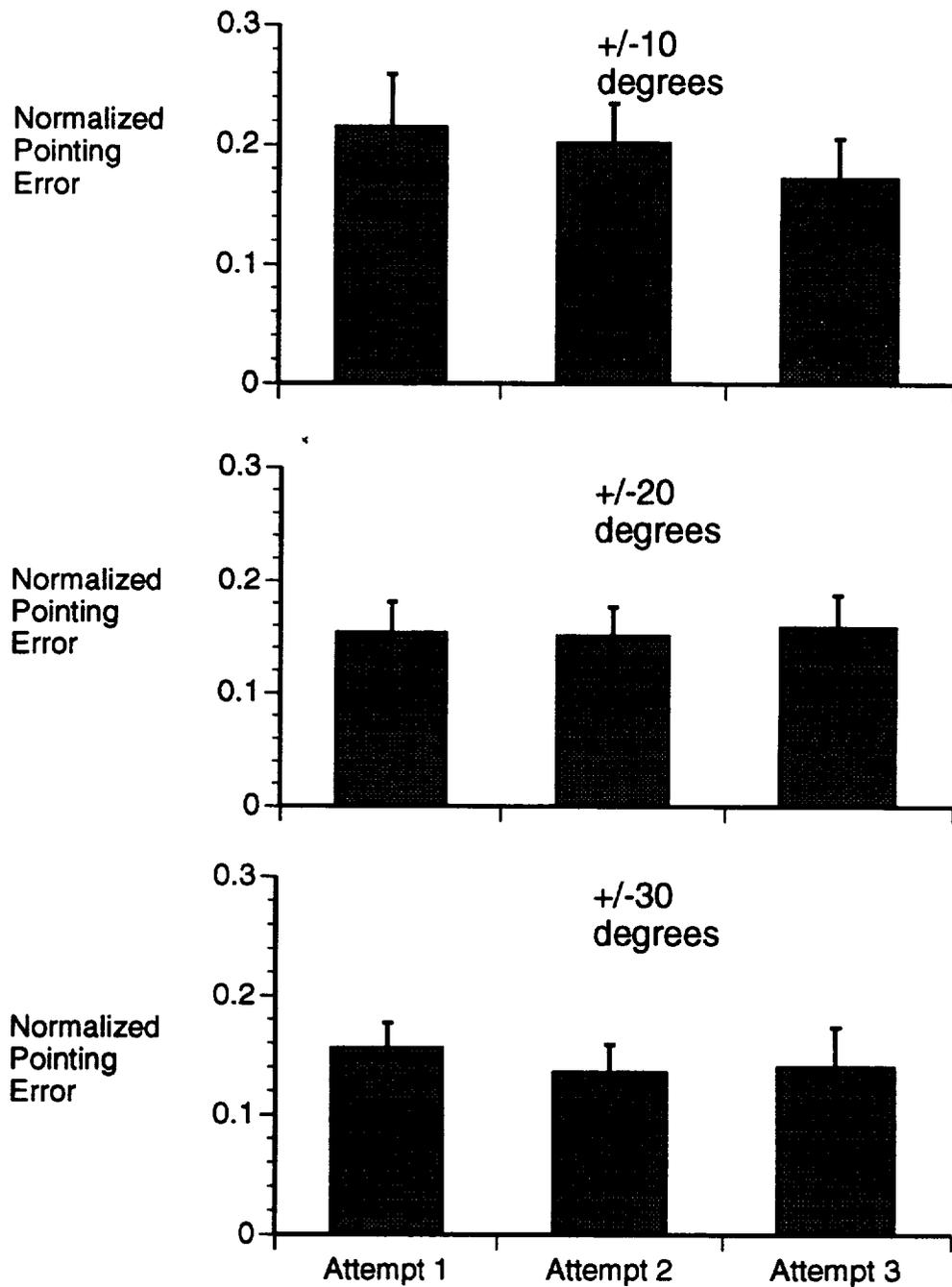


FIGURE 5.- Normalized pointing error for paradigm two (rotation in the dark) for 10, 20 and 30 degree displacements. Data represent mean +/- SEM for 8 subjects.

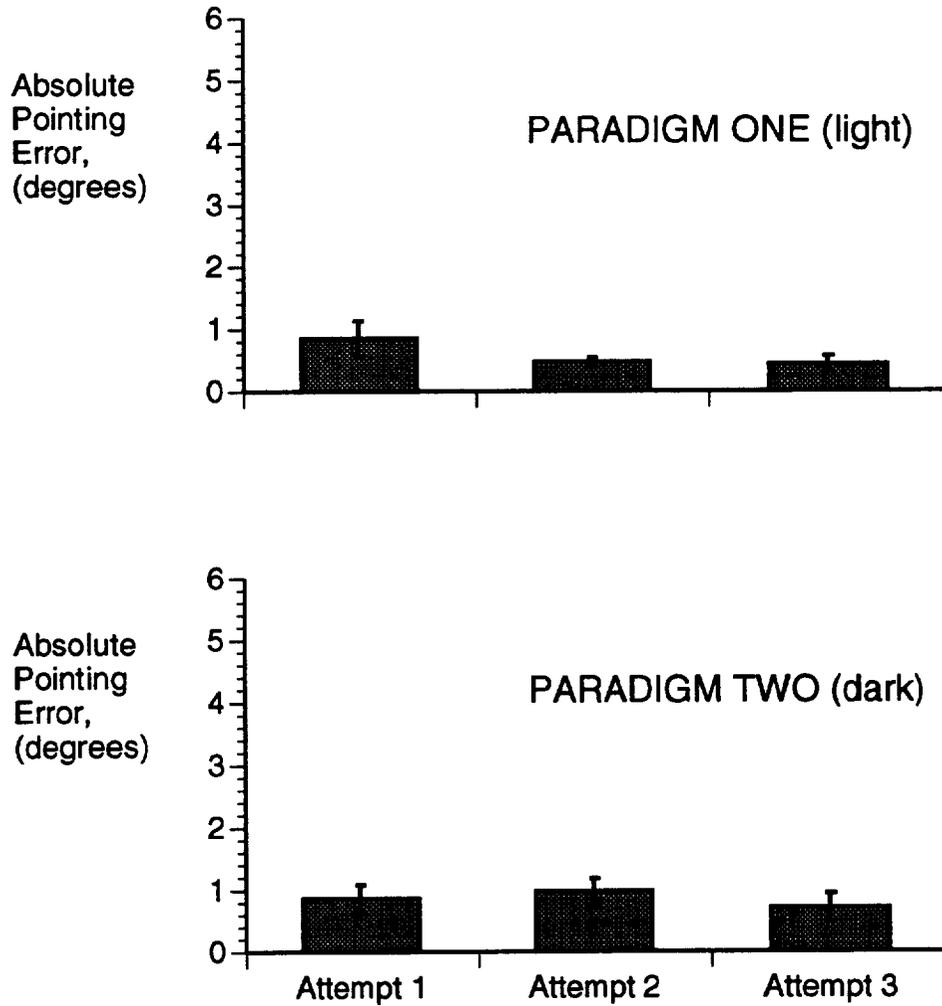


FIGURE 6.- Absolute pointing error for 0 degree displacement for both paradigms. Data represent mean +/- SEM for 8 subjects.

DISCUSSION

Subjects pointed approximately twice as accurately in paradigm one than in paradigm two. In paradigm one, the subjects were able to track the target visually during rotation, while in paradigm two, subjects were rotated in total darkness; no visual feedback of tracking was possible. Thus, visual feedback of target position during rotation improved pointing performance. Nevertheless, subjects were able to point back to the target in paradigm two with a great deal of accuracy. The pointing error for a 20 or 30 degree rotation averaged 0.15 degrees per degree of rotation. Since the subjects were able to point back to targets in paradigm two, a vestibularly derived percept of head rotation must have been the primary if not the only source of information to the motor command networks for pointing movements. This vestibular information has derived primarily from the semicircular canals. To our knowledge, these experiments represent the first demonstration of a role of the vestibular system in manual pointing.

In paradigm one, pointing performance improved between the first and second pointing attempts for 10 and 20 degree rotations. Since the subjects pointed in the dark, they must have compared the position of the laser beam after the first pointing attempt with a stored version of target position within an internal spatial map. They were able to adjust the direction of the laser beam to approximate more accurately the remembered position of the target.

In contrast to their performance in paradigm one, subjects did not improve their pointing performance with second pointing attempts in paradigm two. When the pointing error was normalized against the degree of rotation for 20 and 30 degree rotations, the error averaged 0.15 degrees per degree of rotation.

In paradigm two, subjects were primarily if not entirely dependent on the vestibular system for pointing performance. The normalized error of 0.15 degrees per degree of rotation, therefore, may represent a limit to the accuracy of internal perception of rotation obtained with exclusively vestibular information. Alternatively, vestibular input may provide a more accurate estimate of the degree of chair rotation than the normalized error value. The normalized error may indicate other limits of accuracy. These limits are reached as motor output signals are sent to the pointing arm.

In related, similar studies, subjects were asked to make saccadic eye movements back toward an earth-fixed target after a 20 degree rotation in total darkness. In these experiments, the vestibular-ocular reflex was inhibited so that saccades were necessary to bring the eyes back on target (Bloomberg et al., 1991b). As in paradigm two, vestibular input provided most if not all the sensory information used by these subjects to generate saccades. Saccades brought the eyes to within 1.6 degrees of the target. As in our pointing experiments, subjects tended to make saccades that undershot the target. The normalized eye movement error was 0.08 degrees per degree of rotation, about half of the normalized pointing error reported here. Thus, vestibular input may provide a more accurate measure of rotational

displacement than we can measure from the pointing output. The motor command networks directing eye movements may receive more information from the vestibularly-derived spatial map than the motor command networks for pointing movements. Alternatively, vestibular input to both systems may be identical but the motor command networks for pointing may not produce as accurate an output as those for eye movements.

Our data provide a measure of an individual's ability to point to a remembered target location using only vestibular cues. These data can be contrasted with similar data derived from subjects whose eye-head coordination and internal spatial map have been perturbed, for example, by wearing minimizing lenses. In addition, these data can be compared to data collected from astronauts exposed to the microgravity environment of space.

REFERENCES

- 1.) Bloomberg J, Jones GM, Segal B (1991) Adaptive modification of vestibularly perceived rotation. *Exp. Brain Res.* 84: 47-56.
- 2.) Watt DGD, Money KE, Bondar RL, Thirsk RB, Garneau M, Scully-Power P (1985). Canadian medical experiments on shuttle flight 41-G. *Can Aeronaut. Space J.* 31:215-226.
- 3.) Young LR, Oman CM, Watt DGD, Money KE, Lichtenberg BK (1984). Spatial orientation in weightlessness and readaptation to earth's gravity. *Science.* 225: 205-208.

53-26
4/10/8

N95- 32423

MODELING OF DC SPACECRAFT POWER SYSTEMS

**Final Report
NASA/ASEE Summer Faculty Fellowship Program--1994
Johnson Space Center**

Prepared By:	F.C. Berry
Academic Rank:	Associate Professor
University & Department:	Department of Electrical Engineering Louisiana Tech University Ruston, Louisiana 71272
NASA/JSC	
Directorate:	Engineering
Division:	Propulsion And Power
Branch:	Power
JSC Colleagues:	Tom Jeffcoat
Data Submitted:	August 5, 1994
Contract Number:	NGT-44-005-803

ABSTRACT

Future spacecraft power systems must be capable of supplying power to various loads. This delivery of power may necessitate the use of high-voltage, high-power dc distribution systems to transmit power from the source to the loads. Using state-of-the-art power conditioning electronics such as dc-dc converters, complex series and parallel configurations may be required at the interface between the source and the distribution system and between the loads and the distribution system.

This research will use state-variables to model and simulate a dc spacecraft power system. Each component of the dc power system will be treated as a multiport network, and a state model will be written with the port voltages as the inputs. The state model of a component will be solved independently from the other components using its state transition matrix.

A state-space averaging method is developed first in general for any dc-dc switching converter, and then demonstrated in detail for the particular case of the boost power stage. General equations for both steady-state (dc) and dynamic effects (ac) are obtained, from which important transfer functions are derived and applied to a special case of the boost power stage.

INTRODUCTION

The basic dc-dc level conversion function of switching converters is achieved by repetitive switching between two linear networks consisting of ideally lossless storage elements, inductances and capacitances. In practice, this function may be obtained by use of transistors and diodes that operate as synchronous switches. On the assumption that the circuit operates in the continuous conduction mode in which the instantaneous inductor current does not fall to zero at any point in the cycle, there are only two different states of the circuit. Each state can be represented by a linear circuit model or by a corresponding set of state-space equations.

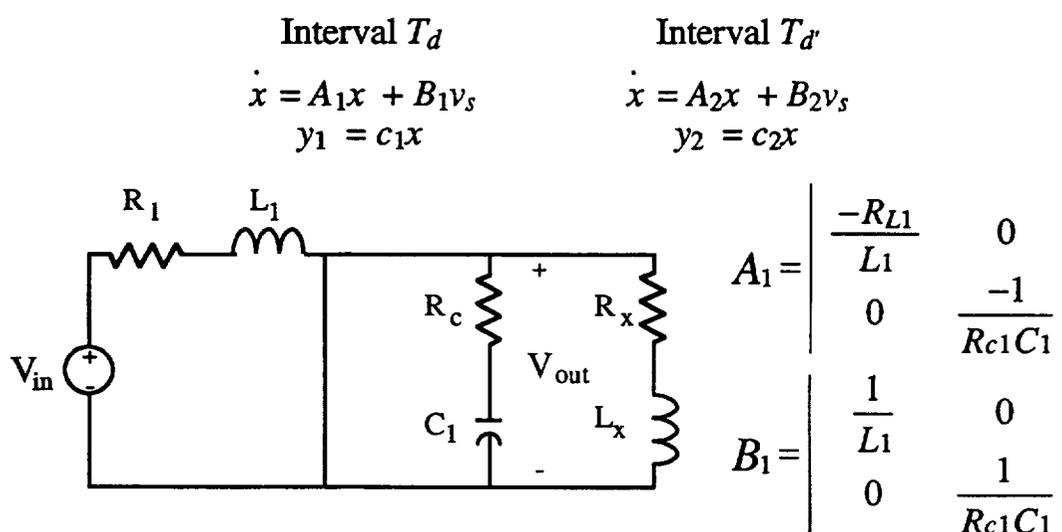


Fig. 1. Network for formulating A_1 and B_1 .

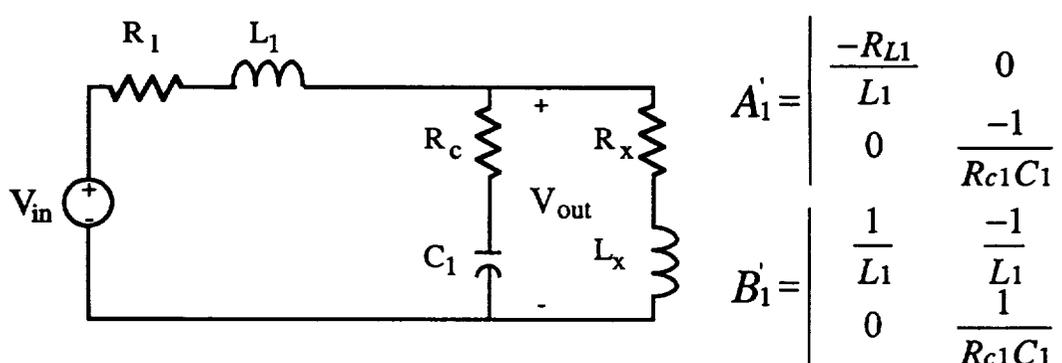


Fig. 2. Network for formulating A_1' and B_1' .

The interval T_d denotes the interval when the switch is in the on state and T_d' denotes the interval when the switch is in the off state. The state model is

formulated for a boost circuit assuming the circuit operates in the continuous conduction mode. The A and B matrices for each of the two linear networks are given along with the network model that was used to obtain the A and B matrices.

SIMULATION

In the analysis of a system by the state-space approach, the system is characterized by a set of first-order differential or difference equations [1]. The dc-dc converter can be described by a set of state-space equations. These state-space equations have the following form:

$$\dot{x} = Ax + Bv$$

where x is a vector of state-variables, v is a vector of inputs, and A and B are matrices. The state-variables, x , are associated with the capacitor voltage and the inductor current. The solution for x is given by

$$x(t) = \Phi(t)x(0) + \int_0^t \Phi(t - \tau)Bv(\tau)d\tau$$

where $\Phi(t)$ is the state transition matrix and $x(0)$ is a vector containing initial values for the state variables. This solution for $x(t)$ is valid if the matrices A and B are constant.

The dc-dc converter is a switching converter and is modeled by two different linear networks (Figures 1 and 2). Therefore, the matrices A and B are not constant over the interval $[0, T]$. The limits of integration must be changed to an interval over which A and B are constant. Also, v will contain entries that are unknown functions of time. All of these conditions will make the convolution integral very difficult to evaluate.

To evaluate the convolution integral it was assumed that the inputs, v , were constant or clamped over the interval of the integration. Therefore a recursive relationship for x can be developed if the appropriate interval, T , is selected for matrices A and B to be constant.

$$x(k+1) = \Phi(T)x(k) + \Theta(T)Bv(k)$$

In this equation, $x(k+1)$ and $x(k)$ are the values of the state vector at $t=(k+1)T$ and $t=kT$, respectively; $v(k)$ is the value of the voltage at $t=kT$. The variable T is referred to as the time step, and it must be selected such that A and B are constant and the inputs, v , does not change appreciably

over the interval $[0, T]$. The variable $\Phi(T)$ is the state transition matrix, which is calculated using the following power series;

$$\Theta(T) = IT + \frac{AT^2}{2!} + \frac{A^2T^3}{3!} + \frac{A^3T^4}{4!} + \dots$$

The variables $\Phi(T)$ and $\Theta(T)$ are related by the following equation;

$$\Phi(T) = A\Theta(T) + I$$

Simulation 1

The $\Phi(T)$ and $\Theta(T)$ are calculated for each time step. A set of these matrices is required for each linear network model of the boost circuit. The solution of $x(k+1)$ is a function of $x(k)$ and $u(k)$, and the $\Phi(T)$ and $\Theta(T)$ for that particular linear network. The calculation of $x(k+1)$ requires one vector addition and two multiplications of a matrix and a vector. This simulation procedure requires two different sets of matrices to be maintained, one set for each linear network. The simulation procedure is summarized in the flowchart of Figure 3.

An example dc power system is simulated using the state variable approach. This system consists of a dc power source connected to a boost network with a resistive-inductive load. The results of this simulation are given in Figure 4.

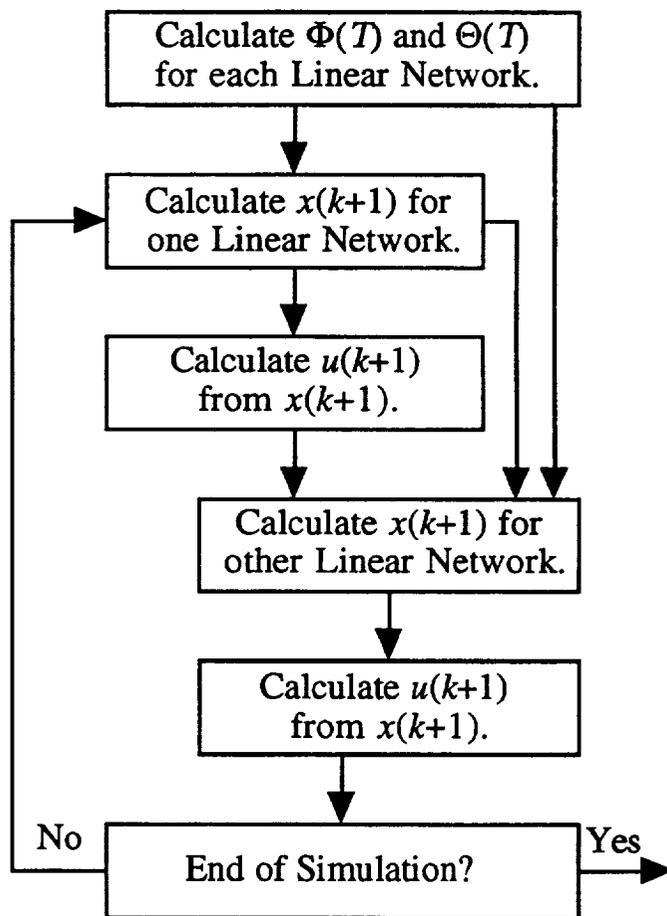


Fig. 3. Flowchart for Simulation 1.

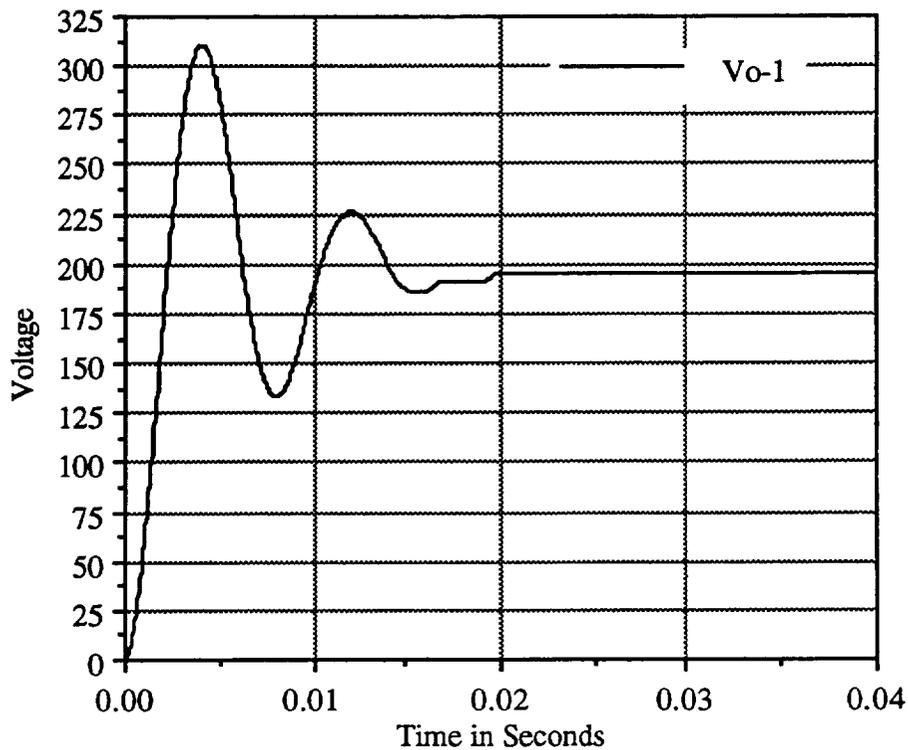


Fig. 4. Output voltage for Simulation 1.

Simulation 2

The objective is to replace the state-space description of the two linear circuits by a single state-space description that represents the approximate behavior of the circuit across the period;

$$T_s = T_d + T_{d'}$$

To represent the approximate behavior of the circuit across the period, T_s , the following averaging method was used. The average was taken by summing the equations for the intervals of T_d and $T_{d'}$ and multiplying each set of equations by d and d' respectively. The following linear continuous systems results;

$$\begin{aligned} \dot{x} &= (dA_1 + d'A_2) x + (db_1 + d'b_2) v_s \\ y &= (dc_1 + d'c_2) x \end{aligned}$$

This model is the basic average model that is the starting model for all other derivations.

The model represented is an averaged model over a single period T_s . If the duty ratio d is assumed to be constant from cycle to cycle, the $d = D$ (the steady-state dc duty ratio). Then the averaged model becomes;

$$\begin{aligned}\dot{x} &= A x + B v_s \\ y &= C x\end{aligned}$$

where

$$A = (DA_1 + D'A_2) \quad B = (DB_1 + D'B_2) \quad c = (DC_1 + D'C_2)$$

The averaged model is a linear system. Therefore, superposition holds can be applied.

The $\Phi(T)$ and $\Theta(T)$ are calculated for each time step, T_s . One set of matrices is required for the averaged linear model of the boost circuit. The solution of $x(k+1)$ is still a function of $x(k)$ and $u(k)$, and $\Phi(T)$ and $\Theta(T)$. The calculation of $x(k+1)$ requires one vector addition and two multiplications of a matrix and a vector. The simulation procedure is summarized in the flowchart of Figure 5.

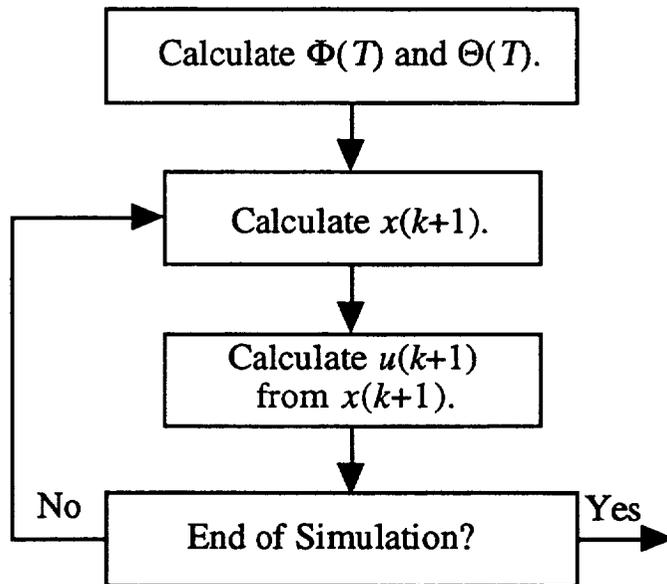


Fig. 5. Flowchart for Simulation 2.

An example dc power system is simulated using the state variable approach. This system consists of a dc power source connected to a boost network with a resistive-inductive load. The results of this simulation are given in Figure 6.

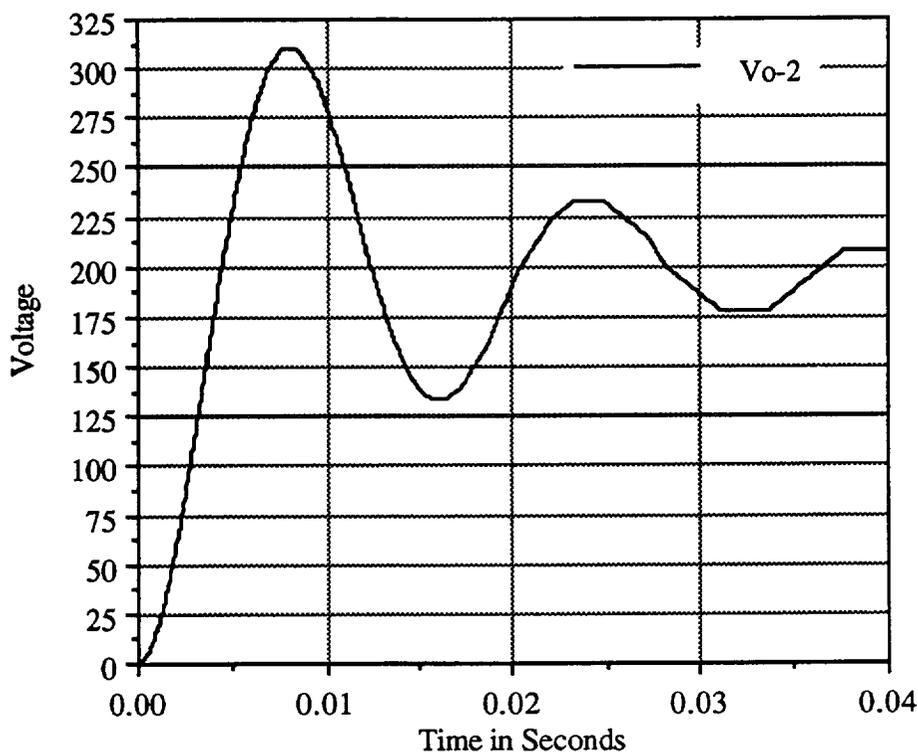


Fig. 6. Output voltage for Simulation 2.

CONCLUSION

Two different simulation methods are presented. The method in Simulation 1 is a more traditional approach that relies on the modeling of each mode of operation of the dc-dc power converter. This is an effective method of modeling and has been used by several of authors [2,3,4]. However, for a large power system the number of models needed to describe the different modes of a variety of switching converters could become very large [1]. The modeling method presented in Simulation 2 tries to produce a single model of a switching converter. This method of modeling is done by producing a weighted average of the different linear models. Anytime an average is taken some distortion in time will occur because averaging is a form of high pass filtering. The comparison of the two different modeling methods in Figure 7 does show time distortion but the voltage magnitudes appear to be basically the same. More work in developing the theory and many more simulations must be done but the average method does appear to have some merit.

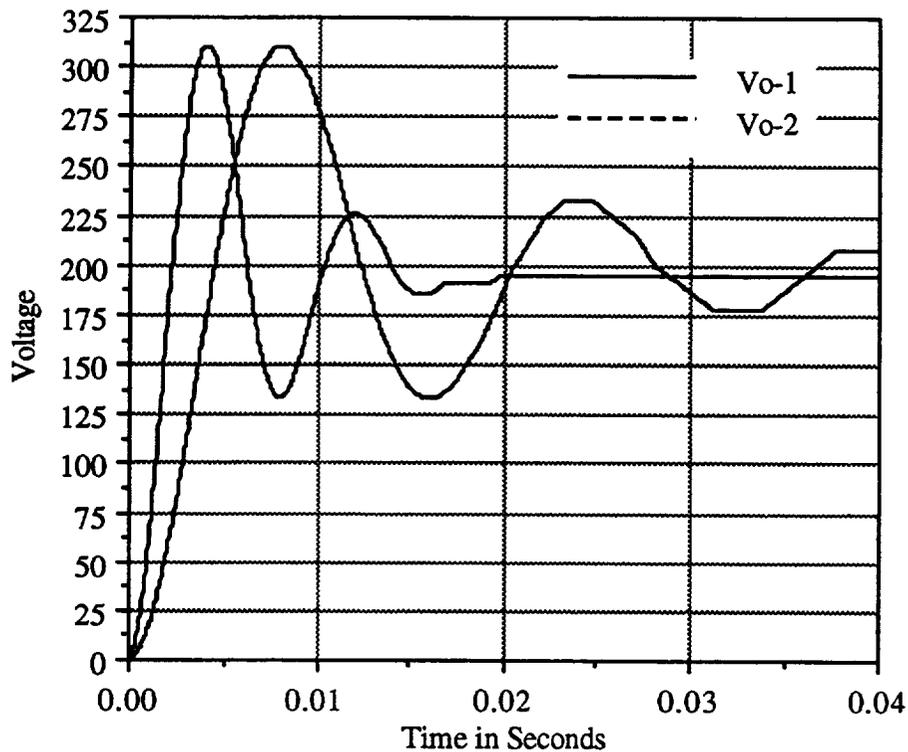


Fig. 7. Output voltage for Simulation 1 and 2.

REFERENCES

- [1] R.M. Nelms and L.L. Grigsby, "Simulation of DC Spacecraft Power Systems," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 25, No. 1, January 1989.
- [2] B.H. Cho, J.R. Lee, and F.C. Lee, "Large-Signal Stability Analysis of Spacecraft Power Processing Systems," *IEEE Transactions on Power Electronics*, Vol. 5, No. 1, January 1990.
- [3] B.H. Cho and F.C. Lee, "Modeling and Analysis of Spacecraft Power Systems," *IEEE Transactions on Power Electronics*, Vol. 3, No. 1, January 1988.
- [4] J.R. Lee, B.H. Cho, S.J. Kim, and F.C. Lee, "Modeling and Simulation of Spacecraft Power Systems," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 24, No. 3, May 1988.

56-72

11169

N95- 32424

**Image Analysis Used to Count and Measure Etched Tracks
from Ionizing Radiation**

Final Report
NASA/ASEE Summer Faculty Fellowship Program--1994
Johnson Space Center

Prepared By:	George E. Blanford. Ph.D., Cindy K. Schulz
Academic Rank:	Professor Graduate Student
College and Department:	University of Houston- Clear Lake Natural and Applied Sciences Houston, Texas 77058
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Earth Science and Solar System Exploration
Branch:	Planetary Missions and Materials
JSC Colleague:	David S. McKay
Date Submitted:	August 5, 1994
Contract Number:	NGT-44-005-803

ABSTRACT

We have developed techniques to use digitized scanning electron micrographs and computer image analysis programs to measure track densities in lunar soil grains and plastic dosimeters.

Tracks in lunar samples are formed by highly ionizing solar energetic particles and cosmic rays during near surface exposure on the Moon. The track densities are related to the exposure conditions (depth and time). Distributions of the number of grains as a function of their track densities can reveal the modality of soil maturation. We worked on two samples identified for a consortium study of lunar weathering effects, 61221 and 67701. They were prepared by the lunar curator's staff as polished grain mounts that were etched in boiling 1 N NaOH for 6 h to reveal tracks. We determined that back-scattered electron images taken at 10% contrast and ~50% brightness produced suitable high contrast images for analysis. We used the *NIH Image* program to cut out areas that were unsuitable for measurement such as edges, cracks, *etc.* We ascertained a gray-scale threshold of 25 to separate tracks from background. We used the computer to count everything that was two pixels or greater in size and to measure the area to obtain track densities. We found an excellent correlation with manual measurements for track densities below $1 \times 10^8 \text{ cm}^{-2}$. For track densities between $1 \times 10^8 \text{ cm}^{-2}$ to $1 \times 10^9 \text{ cm}^{-2}$ we found that a regression formula using the percentage area covered by tracks gave good agreement with manual measurements. We determined the track density distributions for 61221 and 67701. Sample 61221 is an immature sample, but not pristine. Sample 67701 is a submature sample that is very close to being fully mature. Because only 10% of the grains have track densities less than 10^9 cm^{-2} , it is difficult to determine whether the sample matured *in situ* or is a mixture of a mature and a submature soil.

Although our analysis of plastic dosimeters is at an early stage of development, results are encouraging. The dosimeter was etched in 6.25 N NaOH at 70°C for 16 h. We took 200x secondary electron images of the sample and used the *NIH Image* software to count and measure major and minor diameters of the etched tracks. We calculated the relative track etch rate from a formula that relates it to the major and minor diameters. We made a histogram of the number of tracks versus their relative etch rate. The relative track etching rate is proportional to the linear energy transfer of the particle. With appropriate calibration experiments the histogram could be used to calculate the radiation dose.

INTRODUCTION

Counting and measuring etched tracks from ionizing radiation can be used for various scientific purposes such as to determine radiation exposure ages and radiation dosage. Last summer we showed that computer image analysis could be use for counting tracks. This summer we continued to refine this process and we used image analysis to make morphological measurements of tracks in plastics that are used to determine radiation dose.

Price and Walker (1962) discovered that very ionizing radiation, such as fission fragments and cosmic rays, produces a trail of damage in dielectric materials that can be etched with a reagent to form visible tracks (cf. Fleischer *et al.*, 1975; Durrani and Bull, 1987). Their discovery has led to practical applications such as Nuclepore filter paper and cosmic ray dosimeters used by astronauts. Scientific applications include fission track dating of geological samples and cosmic ray-solar energetic particle weathering effects on lunar samples. From the beginning quantitative scientific results have followed from counting tracks on micrographs and by micrographically measuring track morphological characteristics. The sophistication and ready availability of image processing software can reduce this tedious labor.

Exposure of regolith grains to radiation and meteoroid impacts on the Moon, asteroids, some planets and satellites, and interplanetary dust particles produce measurable forms of "weathering." Research has shown that these measurable effects correlate in lunar soils (McKay *et al.*, 1991). Nevertheless, the correlations are very crude because the weathering effects on the Moon are usually measured as a bulk average for a given soil. Most weathering measurements are not very useful for making quantitative predictions of exposure age or even giving a relative measure of maturity for the soil. Furthermore, regolith soils mature by at least two distinct processes: by *in situ* weathering and by mixing. Bulk average measurements cannot distinguish the maturation processes. To improve our understanding of space weathering, we should find these correlations on a grain by grain basis. During the ASEE summer program, we concentrated principally on one form of weathering, namely the formation of tracks in individual soil grains.

Humans in space are exposed to a much more severe radiation environment than workers in nuclear plants. In low earth orbit the exposure is primarily from the inner Van Allen belt of trapped solar radiation. This is especially harsh in the South Atlantic Anomaly where the belt dips lower in altitude than on average. Consequently it is very important to monitor exposure to radiation for humans in space. One method to do this is to have astronauts wear personnel dosimeters. One form of dosimeter that is used is a sandwich of plastic sheets which record tracks of ionizing radiation.

Etching lunar soil grains and plastic sheets in a suitable reagent reveals tracks by producing pits at the track locations. We used a scanning electron microscope (SEM) to make digital images of the etched surfaces of polished grain mounts and the plastic sheets. We developed procedures to rapidly measure track densities with image processing software.

The refinement of track counting techniques will be used to study the maturation of lunar soils as part of a consortium study of various maturation effects. We worked on lunar samples 67701 and 61221 that have been selected by the consortium for the study of lunar soil maturation effects.

We also began exploring the use of computer image analysis to count and to measure diameters of etched tracks in plastics. The major and minor diameters of the tracks are related to the linear energy transfer (LET) of the ionizing particle. By measuring the LET of each track we can add them to find the radiation dose.

Using computer analysis to measure tracks will reduce the tedious measurement of manually counting tracks or of individually measuring them microscopically. This will make track measurements a much more useful tool for understanding radiation environments.

MATURATION MODES OF LUNAR SOILS 61221 AND 67701

EXPERIMENTAL DETAILS

Samples 61221 and 67701 had been sieved into various size fractions by David McKay's group. We had several hundred grains from the 90-150 μm size fraction prepared by the lunar curator's staff as a polished grain mount. The grain mount was etched in boiling 1 N NaOH solution for six hours. The grain mount was rinsed, dried, and coated with ~10 nm layer of AuPd to conduct the electron beam to ground. Most of this work had been done before the beginning of the summer program.

We obtained images on an ISI SEM. The sample was oriented perpendicular to the electron beam. The same condenser lens setting and aperture were used for all images. Nevertheless, the microscope is not equipped with a Faraday cup and we could not be sure of reproducing the same beam current exactly for each microscope session. The working distance knob was set at 8 mm, the focus knobs were set at 5 turns clockwise, and the image was brought into focus initially by adjusting the sample height. This procedure assures that magnification and resolution will be consistent from one session to another.

We determined magnification calibration with a stage micrometer. Last year we verified that it remained consistent within 1.5%. The SEM is capable of making conventional secondary electron images (SEI) and it is also equipped with a back-scattered electron (BSE) detector. Secondary electrons produce a gray scale micrograph that looks very much like a regular black and white photograph. If SEI were used, we felt that fairly sophisticated image processing would be necessary to use the computer to distinguish tracks from background. BSE images, however, naturally showed a high contrast between tracks and background. We purposely chose to exploit this property and took digital images that appeared to the naked eye to be almost binary with very little gray. Using the computer we could set the contrast and brightness to numerically reproducible settings. From our experience from last year, we thought that we could use the same contrast and brightness settings. However, last year's sample had been etched for 15 h. Consequently its tracks were much larger. When we used the contrast and brightness settings from last year, we greatly undercounted the sample. This meant that we had to search again for the proper settings. We chose a wide variety of contrast settings and adjusted brightness settings visually to reproduce a high contrast image. We chose to work at a contrast setting of 10%. The brightness settings ranged from 48.75% to 50.55%. At any one session the range was much smaller, but the range from session to session was greater because the beam current was not exactly reproducible.

We produced digital images using an eXL computer manufactured by Oxford Instruments, formerly Link Analytical. The computer has a proprietary operating system and software. The system is designed to be used with electron microscopes and it controls energy dispersive x-ray analysis as well as digital imaging. Digital images were collected as a Kalman average for 90 s. The images were 512 x 512 pixels at a 256 gray-scale (8 bit). We primarily worked at a magnification of 10000x, but we worked as low as 5000x for grains with low track densities. The images were up loaded to the building 31 network (node: SN-Titan). From the building 31 network the files were downloaded to a UHCL Macintosh (node: regolith).

The images were analyzed using *NIH Image* software, a public domain program. This software is faster than the eXL software and we do not effectively tie up the SEM doing our analyses. We created a mask for each image to obscure parts of the image we did not wish to analyze such as areas off the edge of the grain, large cracks, *etc.* We set a binary threshold at a gray scale of 25, and had the computer count everything that had a size of two pixels or greater. *NIH Image* also returned the total area covered by tracks.

CALIBRATION

Figure 1 shows a correlation diagram of track density measurements using image analysis with conventional measurements from a photomicrograph. The correlation is excellent for track densities below $1 \times 10^8 \text{ cm}^{-2}$. However, above track densities of

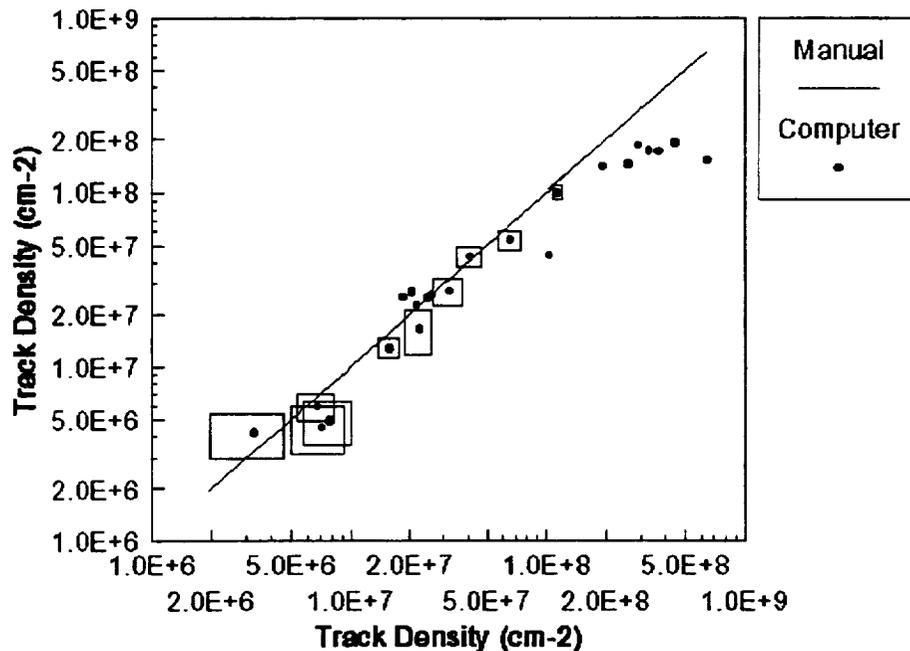


Figure 1. Graph of track densities in lunar soil grains from samples 61221 and 67701. The ordinate has values determined from counts using computer software. The abscissa has values determined by manual counting. Rectangles represent one standard deviation uncertainty in the measurements and are shown only for a representative set of points.

$1 \times 10^8 \text{ cm}^{-2}$ the image analysis technique shows saturation as we had found last year. We found that the area covered by the tracks was proportional to the number of tracks. We calculated a regression line that correlates the percentage area of the tracks with the track density. In Fig. 2 we show a graph of manually counted track density versus track density determined using percentage area and the regression equation. When we present histograms of the two samples we indicate which measurements result from computer counting, which from the area regression line, and which were equivalent using both methods.

RESULTS

Track density distributions for samples 61221 and 67701 are shown in Figs. 3 and 4. Blanford *et al.* (1979) (or McKay *et al.*, 1991) have discussed the relationship of the track density distribution to the modality of soil maturation. The two figures clearly show that the two soils have different radiation histories. Soil 61221 is immature which means that most of its soil grains have not been exposed for very long ($< \sim 10^4$ y) within the top millimeter of the lunar surface. That there are a relatively small number of grains with

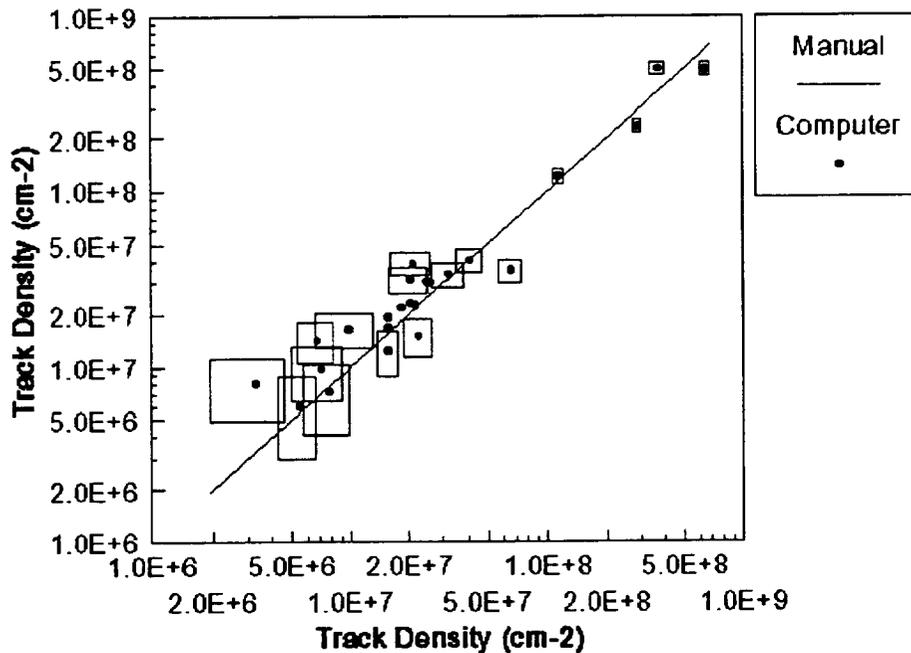


Figure 2. Track densities determined from the percentage area of the image covered by tracks versus track densities determined from manual counting. Rectangles represent one standard deviation uncertainty in the measurements and are shown only for a representative number of points.

track densities $<10^7 \text{ cm}^{-2}$ indicates that this soil has been within $\sim 1 \text{ cm}$ of the lunar surface for a relatively long time ($\sim 10^6 \text{ y}$). Soil 67701 is submature, but nearly mature. Because only 10% of the soil grains have track densities $<10^8 \text{ cm}^{-2}$, it is difficult to tell whether the soil is bimodal consisting of two mixed components (submature type II) or it has matured *in situ* (submature type I). The maturity designations for 61221 and 67701 agree with the designations from the ferromagnetic resonance index I_s/FeO . It is known that the darkening of the albedo of lunar soils with increased maturity is an effect that is dominated by the smallest fractions of the lunar soil (Fischer *et al.*, 1994). We plan to repeat these measurements with the 10-20 μm and 20-45 μm fractions to determine whether or not these smaller fractions give qualitatively the same information as the 90-150 μm fraction.

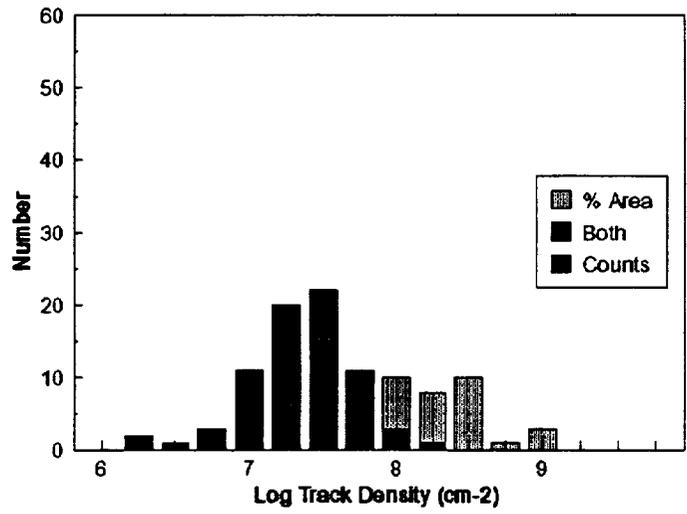


Figure 3. Solar flare and cosmic ray track density distribution for the 90-150 μm fraction of lunar soil 61221 (102 grains). Measurements determined by computer counting alone are shown as solid, measurements determined by percentage area regression are shown by the lightest shading, and measurements that were the same for both methods are shown by the intermediate shading.

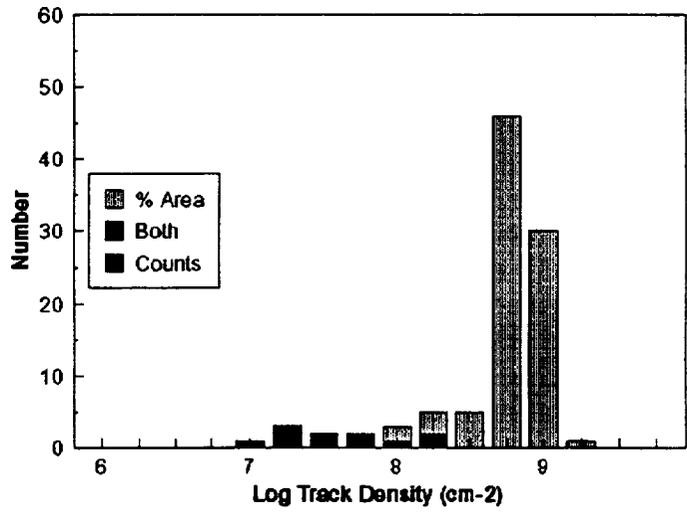


Figure 4. Solar flare and cosmic ray track density distribution for the 90-150 μm fraction of lunar soil 67701 (98 grains). Measurements determined by computer counting alone are shown as solid, measurements determined by percentage area regression are shown by the lightest shading, and measurements that were the same for both methods are shown by the intermediate shading.

ANALYZING PLASTIC TRACK DOSIMETERS USING IMAGE ANALYSIS

EXPERIMENTAL DETAILS

Experimental work for this project is at a very preliminary stage. A personnel dosimeter that had been flown on STS 59 was provided by Dr. Gautum Badwhar. It contained 4 sheets of CR-39 plastic (allyl diglycol carbonate) interleaved with thin sheets of Makrofol (bisphenol-A polycarbonate). We etched part of one of the CR-39 sheets in 6.25 N NaOH at 70°C for 16 h. We coated the sample with ~10 nm AuPd for electron microscopic observation.

We took both SEI and BSE images with varying contrast and brightness. In this case the SEI images proved to be the most useful. Because we cannot set the contrast and brightness of SEI images with the digital accuracy of the BSE images, we took a series of micrographs using various settings that gave about the same visual appearance. We found that there needs to be some adjustment of the threshold for image analysis, but the counting and measuring were not strongly dependent on the conditions used. This should lead to very reproducible results. However, more experimentation will be necessary to establish conditions reliably.

Figure 5 shows a 200x SEI image of the etched CR-39 dosimeter with the counting record on the right. Besides counting the tracks the *NIH Image* software also can measure the major and minor diameters of the tracks. Durrani and Bull (1987) give the following formulas for the major diameter D and the minor diameter d

$$D = \frac{2h(V^2 - 1)^{1/2}}{V \sin \theta + 1}$$

$$d = 2h \left(\frac{V \sin \theta - 1}{V \sin \theta + 1} \right)^{1/2}$$

where h is the amount of surface removed during etching, V is the relative etch rate V_T/V_B , and θ is the angle the track makes with the surface of the plastic. V_T is the etch rate along the track and V_B is the bulk etch rate. The relative etch rate V is the critical measurement for dosimetry because it is directly related to the linear energy transfer (LET) of the particle (Price *et al.*, 1967; O'Sullivan *et al.*, 1971). Manipulating the formulas for d and D , one finds that the relative etch rate V is given by

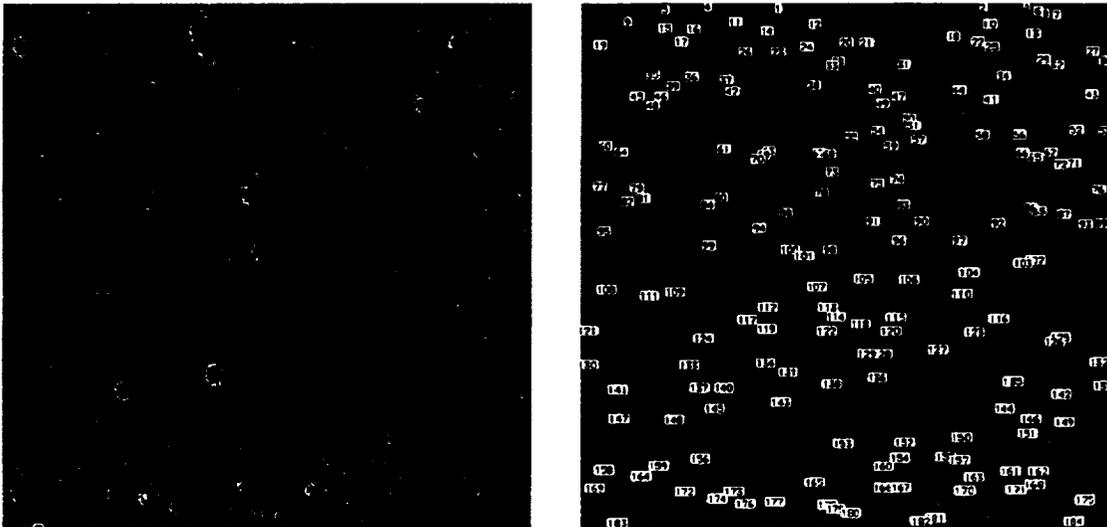


Figure 5. On the left is a secondary electron image of an etched CR-39 plastic dosimeter exposed on mission STS 59. The image on the right shows the tracks that were counted.

$$V = \sqrt{1 + \frac{D^2}{h^2 \left[1 - \left(\frac{d}{2h} \right)^2 \right]}}$$

Normally h is measured during each experiment, but we did not plan far enough ahead and did not make this measurement. We estimated it instead from He and Price (1992) as being $21 \mu\text{m}$ for our etching conditions. Figure 6 shows a histogram of the number of tracks as a function of their relative etching rate. Because the relative etching rate is directly related to LET, if we had a calibration for this plastic, the dose could be calculated by multiplying the number in each column by the appropriate LET for that column and adding them together. Most of the tracks represented in this histogram are probably protons and α -particles. However, the three tracks on the right of the histogram are very likely heavier ions. Many images would have to be analyzed to include these heavy ions with good statistical accuracy and thereby give a good measure of the dose.

CONCLUSION

During this summer program we have found that using image analysis we can count and measure tracks with as great a reliability as by using manual means. We also found that changing etching conditions changes image analytical conditions. Establishing the correct analytical conditions is quite time consuming. For time efficient measurements one must first establish the desired etching conditions and then use them consistently for

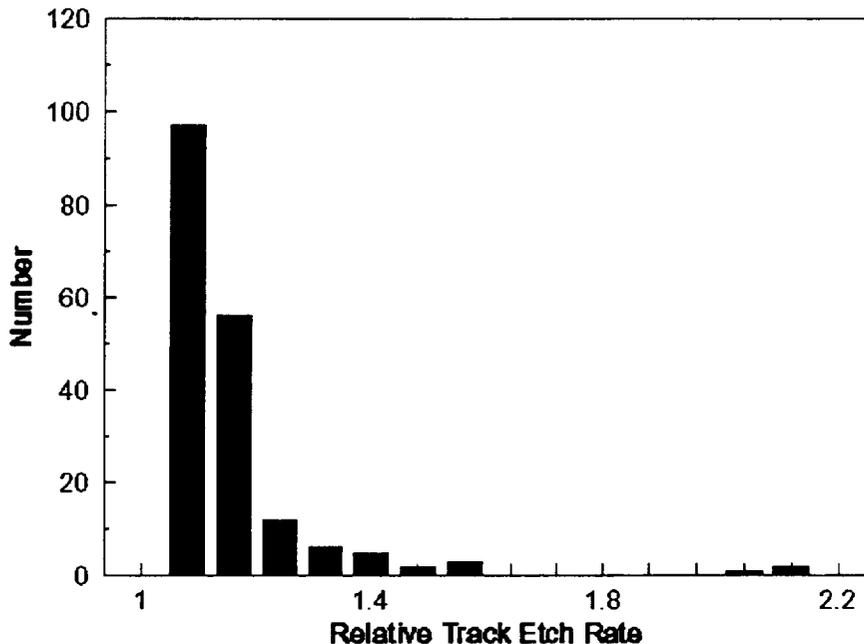


Figure 6. A histogram of the number of tracks as a function of their relative etching rate. The relative etch rate is directly related to the linear energy transfer (LET). By multiplying the number of tracks by the appropriate LET the dose can be obtained.

future work. For working with plagioclase grains in lunar samples we now have established two sets of image analytical conditions for two etching conditions. We are now able to analyze lunar plagioclase samples rapidly for the two cases.

We have established preliminary conditions for counting and measuring major and minor track diameters in CR-39 plastic. We have not yet optimized these conditions. Furthermore the measurement of major and minor track diameters is not the most accurate method of measuring the relative track etch rate V . It is better to use the track length to obtain V (Durrani and Bull, 1987). This requires measuring the track's projected length and its depth. Scanning electron micrographs are not suitable for this, whereas transmission optical micrographs are. Work will continue to try to do these measurements with digital images taken on a stereo optical microscope. If the goal of these measurements were to measure the charge-energy spectrum of cosmic rays, then measuring track lengths would be necessary. However, for space dosimetry the measurement of diameters may be adequate because most of the ionizing particles are protons. Proton tracks are small and diameter measurements may be adequate. Nevertheless, the heavy ions make a significant contribution to the total dose even though they are small in number. This is the result of the LET being proportional to the square of

the atomic number. At this time we can not definitively answer whether we can simply measure the major and minor diameters or if we must also measure lengths and depths.

REFERENCES

- Blanford G.E., Blanford J., and Hawkins J.A. (1979) Irradiation stratigraphy and depositional history of the Apollo 16 double drive tube 60009/10. *Proc. Lunar and Planetary Sci. Conf. 10th*, 1333-1349.
- Durrani S.A. and Bull R.K. (1987) *Solid State Nuclear Track Detection*. Pergamon Press, Oxford.
- Fischer E.M., Pieters C.M., and Pratt S.F. (1994) Modeling the space weathering-induced optical alteration of lunar soils: First results. In *Lunar and Planetary Science XXV*, p. 371-373. Lunar and Planetary Institute, Houston.
- Fleischer R.L., Price P.B., and Walker R.M. (1975) *Nuclear Tracks in Solids: Principles and Applications*. University of California Press, Berkeley, CA.
- He Y.D. and Price P.B. (1992) Sensitivity study of CR-39 plastic track detectors. *Nucl. Tracks Radiat. Meas.* **20**, 491-494.
- McKay D.S., Heiken G., Basu A., Blanford G., Simon S., Reedy R., French B.M., and Papike J. (1991) The lunar regolith. In *Lunar Sourcebook: A User's Guide to the Moon* (G. Heiken, D. Vaniman, and B.M. French, ed.), pp. 285-356. (Cambridge University Press, Cambridge).
- Morris R.V., Score R., Dardano C., and Heiken G. (1983) *Handbook of Lunar Soils*. JSC 19069, NASA Johnson Space Center, Houston.
- O'Sullivan D., Price P.B., Shirk E.K., Fowler P.H., Kidd J.M., Kobetich E.J. and Thorne R. (1971) High resolution measurements of slowing cosmic rays from Fe to U. *Phys. Rev. Letters* **26**, 463-466.
- Price P.B. and Walker R.M. (1962) Electron microscope observation of etched tracks from spallation recoils in mica. *Phys. Rev. Letters* **8**, 217-219.
- Price P.B., Fleischer R.L., Peterson D.D., O'Ceallaigh C., O'Sullivan D., Thompson A. (1967) Identification of isotopes of energetic particles with dielectric track detectors. *Phys. Rev.* **164**, 1618-1620.

4470

Unified Approach For Incompressible Flows
Final Report
NASA/ASEE Summer Faculty Fellowship Program - 1994
Johnson Space Center

Prepared By: Tyne-Hsien Chang, Ph.D., P.E.
Academic Rank: Associate Professor
College and Department: Texas A&M University at Galveston
Department of Maritime Systems Engineering

NASA/JSC
Directorate: Engineering
Division: Navigation Control and Aeronautics Division
Branch:

JSC Colleague: C. P. Li, Ph.D.
Date Submitted: August 9, 1994
Contract Number: NGT-44-005-803

ABSTRACT

A unified approach for solving incompressible flows has been investigated in this study. The numerical CTVD (Centered Total Variation Diminishing) scheme used in this study was successfully developed by Sanders and Li for compressible flows, especially for the high speed. The CTVD scheme possesses better mathematical properties to damp out the spurious oscillations while providing high-order accuracy for high speed flows. It leads us to believe that the CTVD scheme can equally well apply to solve incompressible flows. Because of the mathematical difference between the governing equations for incompressible and compressible flows, the scheme can not directly apply to the incompressible flows. However, if one can modify the continuity equation for incompressible flows by introducing pseudo-compressibility, the governing equations for incompressible flows would have the same mathematical characters as compressible flows. The application of the algorithm to incompressible flows thus becomes feasible.

In this study, the governing equations for incompressible flows comprise continuity equation and momentum equations. The continuity equation is modified by adding a time-derivative of the pressure term containing the artificial compressibility. The modified continuity equation together with the unsteady momentum equations forms a hyperbolic-parabolic type of time-dependent system of equations. Thus, the CTVD schemes can be implemented. In addition, the physical and numerical boundary conditions are properly implemented by the characteristic boundary conditions.

Accordingly, a CFD code has been developed for this research and is currently under testing. Flow past a circular cylinder was chosen for numerical experiments to determine the accuracy and efficiency of the code. The code has shown some promising results.

INTRODUCTION

GOVERNING EQUATIONS

The two-dimensional incompressible-flow equations expressed in conservative variables, p , ρu , and ρv are as follows:

$$\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = 0 \quad (1)$$

$$\frac{\partial \rho u}{\partial t} + \frac{\partial \rho u^2}{\partial x} + \frac{\partial \rho uv}{\partial y} = -\frac{\partial p}{\partial x} + \left(\frac{\mu}{\rho}\right)\left(\frac{\partial^2 \rho u}{\partial x^2} + \frac{\partial^2 \rho u}{\partial y^2}\right) \quad (2)$$

$$\frac{\partial \rho v}{\partial t} + \frac{\partial \rho uv}{\partial x} + \frac{\partial \rho v^2}{\partial y} = -\frac{\partial p}{\partial y} + \left(\frac{\mu}{\rho}\right)\left(\frac{\partial^2 \rho v}{\partial x^2} + \frac{\partial^2 \rho v}{\partial y^2}\right) \quad (3)$$

where u and v are velocity components in the x and y directions, respectively. P is the static pressure. μ denotes the dynamic viscosity. ρ is the fluid density. The incompressible governing equations (1) through (3) are mathematically classified as elliptic partial differential equations while compressible governing equations are hyperbolic partial differential equations. Because of the mathematical difference between the hyperbolic and elliptic partial differential equations, a well-developed numerical schemes for compressible flows can not apply directly to solve incompressible flows. However, if one modifies the continuity equation given in equation (1) by introducing artificial compressibility, the resulting incompressible governing equations are of hyperbolic type. The modified continuity equation including a time-derivative of the pressure term containing the artificial compressibility is given in equation (4).

$$\frac{\partial p}{\partial t} + \frac{\partial \beta \rho u}{\partial x} + \frac{\partial \beta \rho v}{\partial y} = 0 \quad (4)$$

where β is known as the pseudo-compressibility constant and has the dimension as velocity square.

Equations (2) through (4) are transformed onto generalized curvilinear coordinates, ξ and η given by

$$\xi = \xi(x, y, t) \tag{5}$$

$$\eta = \eta(x, y, t) \tag{6}$$

The governing equations are then given by

$$\frac{\partial Q}{\partial \tau} + \frac{\partial(E - E_\eta)}{\partial \xi} + \frac{\partial(F - F_\xi)}{\partial \eta} = 0 \tag{7}$$

where

$$Q = \frac{1}{J} \begin{bmatrix} \rho \\ \beta \rho u \\ \beta \rho v \end{bmatrix} \tag{8}$$

$$E = \frac{1}{J} \begin{bmatrix} \beta \rho (U - \xi_t) \\ \rho u U + \xi_x \rho \\ \rho v V + \xi_y \rho \end{bmatrix} \tag{9}$$

$$F = \frac{1}{J} \begin{bmatrix} \beta \rho (V - \eta_t) \\ \rho u V + \eta_x \rho \\ \rho v V + \eta_y \rho \end{bmatrix} \tag{10}$$

$$F_v = \frac{\mu}{J} \begin{bmatrix} 0 \\ (\xi_x^2 + \xi_y^2) U_\xi + (\xi_x \eta_x + \xi_y \eta_y) U_\eta \\ (\xi_x^2 + \xi_y^2) V_\xi + (\xi_x \eta_x + \xi_y \eta_y) V_\eta \end{bmatrix} \quad (11)$$

$$E_v = \frac{\mu}{J} \begin{bmatrix} 0 \\ (\xi_x \eta_x + \xi_y \eta_y) U_\xi + (\eta_x^2 + \eta_y^2) U_\eta \\ (\xi_x \eta_x + \xi_y \eta_y) V_\xi + (\eta_x^2 + \eta_y^2) V_\eta \end{bmatrix} \quad (12)$$

U and V are contravariant velocities. J is the Jacobian of the transformation, the metrics of the transformation are

$$\xi_x = \frac{\partial \xi}{\partial x}, \xi_y = \frac{\partial \xi}{\partial y} \quad (13)$$

and

$$\eta_x = \frac{\partial \eta}{\partial x}, \eta_y = \frac{\partial \eta}{\partial y} \quad (14)$$

U and V are contravariant velocities and given as follows:

$$U = \xi_x u + \xi_y v, V = \eta_x u + \eta_y v \quad (15)$$

The transformed governing equation given in equation (7) is then solved by a finite volume method. Accordingly, the conservation laws given below in integral form is used in the formulation.

$$\frac{\partial}{\partial t} \int Q dV + \int D \cdot dA = 0 \quad (16)$$

)

D is derived from equation (7) and a second-order flux tensor defined in ξ and η .

NUMERICAL SCHEME

A CTVD scheme developed by Sanders and Li (Ref. 1) is adopted to determine the interfacial variables for the finite volume formulation. The CTVD differencing has distinctively desired properties in overcoming the spurious oscillations and odd- and even-point decoupling in the solution which are caused by the use of central differencing. This scheme has demonstrated the superiority in reducing CPU time while providing up to fourth order accuracy. In addition, the implementation of these algorithms is simple and no tuning parameters are needed.

By applying the interfacial variables derived from the CTVD scheme to the conservation laws given in equation (16) numerical fluxes are evaluated on the cell faces based on the spatial discretization. Equation (16) is further reduced to a system of time dependent differential equations by the method of lines. The system of differential equations can be integrated by a number of relaxation methods. In this study, the explicit Runge-Kutta method is employed for solutions.

The implementation of wall and far-field boundary conditions is crucial to the numerical scheme. The flows are treated as subsonic. Therefore, the proper far-field boundary conditions must be imposed. Characteristic boundary conditions are used and integrated into CTVD scheme to evaluate fluxes.

NUMERICAL EXPERIMENTS

A CFD code based on the CTVD scheme has been developed to solve two-dimensional incompressible flows. Numerical testings have been conducted to solve flow past a circular cylinder problem. The far-field boundary conditions are imposed at a distance 10 times of radius from the center of the cylinder. The initial conditions are given as follows:

$$\begin{aligned} \rho &= 2.0 \text{ lb/ft}^2 \\ U_0 &= 1.0 \text{ ft/sec.} \\ \rho &= 1.0 \text{ slug/ft}^3 \end{aligned}$$

$R=1.0$ ft., is the radius of the cylinder. $\beta=5,10,30,40$, and $50 \text{ ft}^2/\text{sec}^2$ are utilized in numerical testings. The O-grid system is used to discretize the flow domain. All numerical testings are based on a grid of 61×82 nodes. A periodic boundary condition is imposed around the branch cut. The maximum CFL number can be used in each testing is 0.6.

According to the numerical results, β value plays very important role in the numerical calculation. When β is greater than 10 and less than 40, the solution converges. A typical second-order pressure field is given in Figure 1 after 500 iterations.

The numerical testings have not finished yet. More testings are need to investigate properly imposing wall boundary conditions. How to increase CFL number in the calculation is also a subject for more research.

CONCLUDING REMARKS

A research CFD code has been developed for incompressible flows using the CTVD scheme. The numerical results have shown that the CTVD scheme can apply to incompressible flows by the use of a modified continuity equation. A converged solution can be obtained by choosing a proper pseudo-compressibility constant. A further investigation on the CFL number is necessary in order to increase the efficiency of the method. The implementation of diagonal ADI method to solve the reduced set of ordinary differential equations would increase the efficiency of the method.

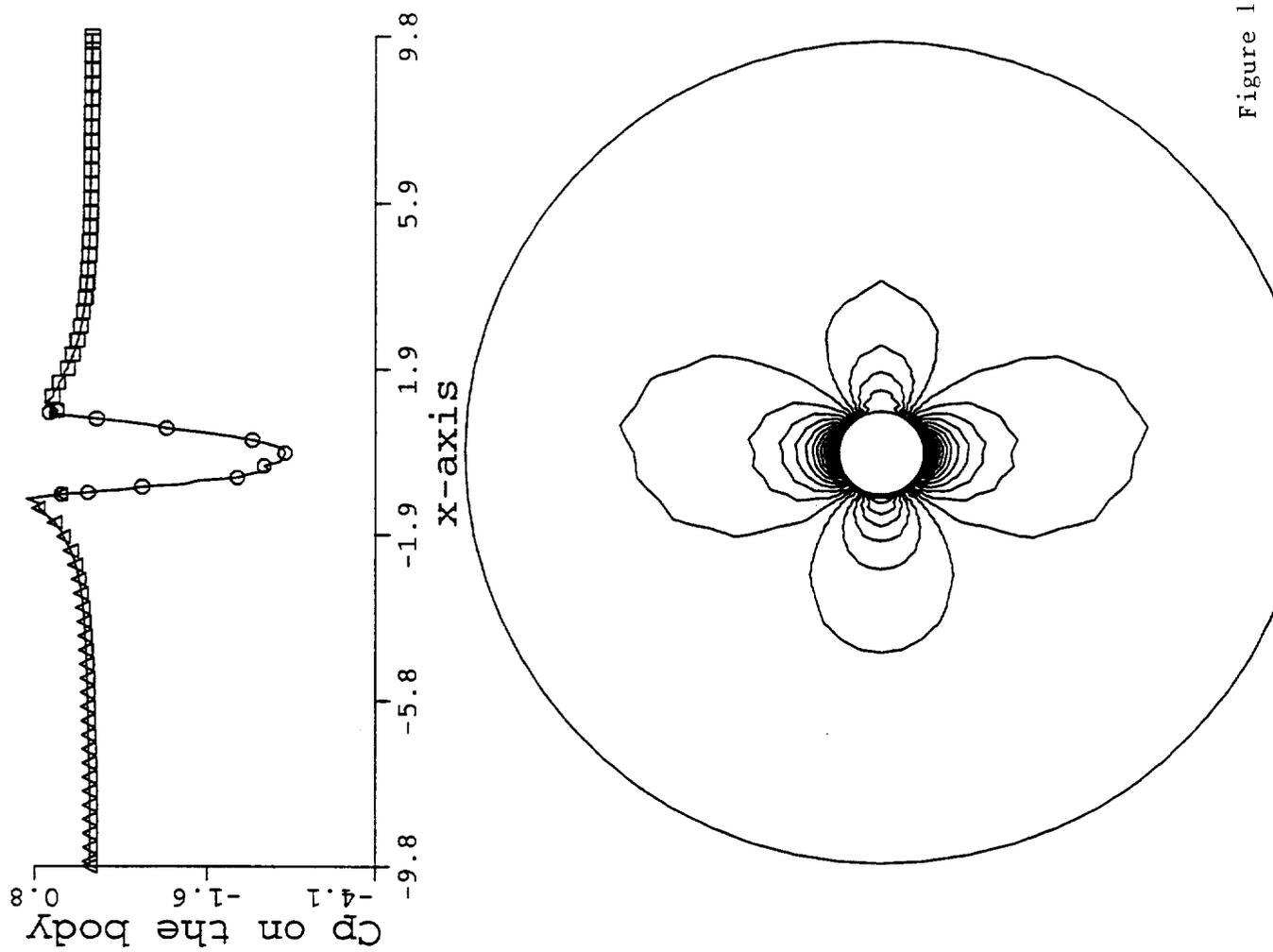
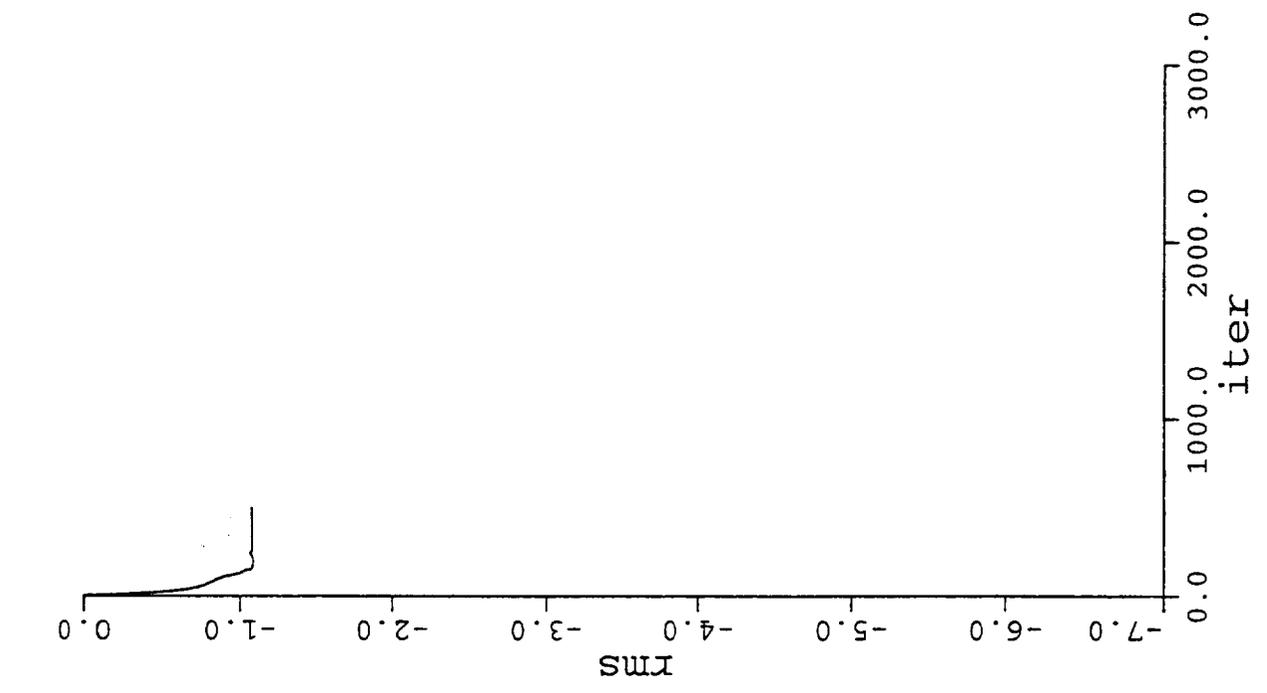


Figure 1

REFERENCES

1. Sanders, R. and Li, C. P., "A Variation Nonexpansive Central Differencing Scheme for Nonlinear Hyperbolic Conservation Laws," Proc. Tenth Inter. Conf. Computing Methods in Applied Sciences and Engineering, Versailles, France, Feb. 1992, pp. 511-526.

53-39

4971

N95-32426

**CHARACTERIZING THE FRAGMENTATION RING ON
WITNESS PLATES**

Final Report
NASA/ASEE Summer Faculty Fellowship Program -- 1994
Johnson Space Center

Prepared By: Yuan-Liang A. Chen, Ph.D.,
Academic Rank: Associate Professor,
College and Department: Oklahoma Baptist University
Division of Natural Science
Shawnee, Ok 74081

NASA/JSC

Directorate: Space and Life Sciences

Division: Solar System Exploration Division

Branch: Planetary Sciences, SN 4

JSC Colleague: Friedrich Horz

Date Submitted: August 5, 1994

ABSTRACT

Hole-saw ring patterns on the witness plate are studied systematically. Aluminium alloy 1100 and 6061, inconel, teflon, and lead of different thickness were utilized for target material. The ring pattern are observed at $D_p/T=1.98$ to 4.02 for aluminium 1100 target, at $D_p/T=1.55$ to 10.51 for aluminium 6061 target, at $D_p/T=4.41$ to 9.98 for inconel target, at $D_p/T=1.33$ to 3.94 for teflon target, and at $D_p/T=12.5$ to 20.89 for lead target. The ring diameter showed a decreasing trend when D_p/T ratio increased. A analytical model is introduced. The crater distributions on the ring were investigated. The majority of the craters are in the neighborhood of $1100\ \mu\text{m}$ to $1400\ \mu\text{m}$ size. Many broken string like craters were observed on the ring as well as inside the ring.

INTRODUCTION

Impact experiments have been performed with a 5 mm Light Gas Gun (LGG) at the Experimental Impact Laboratory of the Solar System Exploration Division, SN4, with the objective to learn about the collisional fragmentation, dispersion and deceleration of hypervelocity projectiles as they pass through targets of variable thickness. Such experiments are of basic nature, yet they relate directly to the design of cosmic dust flight instruments and to the protection of space craft in near Earth orbit.

Consistent with the SN-interests in cosmic dust instruments, soda lime glass projectiles of variable diameters (50, 150, 1000, 3175 μm diameter) were accelerated in the LGG to some 6 km/s and they impacted and penetrated targets of various material and thickness. This project studied aluminum alloys (1100, 6061), teflon, inconel and lead targets. A so-called "witness plate" (aluminum or copper) was positioned behind these targets to intercept the entire debris cloud that exits penetrated targets; the fragments composing this cloud are of high speed and caused a complex pattern of secondary impacts on the witness plate. The geometry of these secondary crater patterns is the subject of this study. Specially, as target thickness approaches that of projectile diameter, a characteristic distinct ring of fragment-impacts is observed on the witness plate that contrasts markedly with irregular fragment dispersions at more massive target, and with distinctly radial deposits resulting from the penetration of thinner targets. The ring pattern was introduced with the descriptive term "**Hole-Saw Ring**" by F. Horz et al. (1) and were reported by Stimp et al.(2), Pietkutowsky (3) and others for various projectile and target combinations and impact velocities. Many theoretical models (4,5) study the fragmentation of spheres and associated debris cloud. They suggest that the outer shell of projectile yields smaller and faster fragments while the more massive and slow fragments are due to the core of the projectile. At the time of this project being performed, there is no theory or model that derived these uniform fragment sizes and dispersion angles from the hole-saw ring. This distinct hole-saw ring occurs over a very limited range of D_p/T conditions. It appears to be an excellent boundary condition to test theoretical and computational models on collisional disruption of spherical objects.

The Project is divided into two parts: Phase I--Extract the necessary data from the witness-plates, such as the radius of the ring; the number and the size distribution of the craters in relation to the impact conditions. Study the physics rules that govern these phenomena. Phase II-- Use a hydrocode (CTH) to model the rings and establish a reliable tool to further investigate this special fragmentation case. During the ASEE/NASA summer program, we concentrated on the first part of the Project. The results from the Phase I will be utilized to continue Phase II of this project.

HOLE-SAW RING DIAMETERS ON THE WITNESS PLATE

This project intends to contribute to an improved understanding of hole-saw ring pattern on the witness plate. We investigated the witness plates due to various diameter soda-lime glass sphere projectiles impacted on different target materials. The projectile's diameter and target thickness ratios, D_p/T , were altered while the impact velocity was kept a constant. Four sizes of glass spheres were accelerated to 2 km/s - 6 km/s using the 5 mm LGG at the Experimental Impact Laboratory of the SSED, SN4. The 3.175 mm soda-lime glass sphere was selected for the purpose of a more uniform projectile material and better production of the hole-saw ring. We employed aluminum alloy 6061-T6 with thickness of 4.064 mm, 5.334 mm, 6.35 mm, 8.128 mm, 10.016 mm, and 16.002 mm as the target. The copper witness plate with various stand-off distance "L" was utilized to collect the information for analysis.

An Analytical Model

A debris-cloud model by J. Lawrence(6) is employed to study the relationship between the hole-saw ring diameter and the ratio of D_p/T .

The nondimensional mass ratio, M , is defined as

$$M = \frac{3\beta^2 \rho_t T}{2\rho_p D_p} = \frac{m_t}{m_p}, \quad (1)$$

where the β is the ratio of the hole diameter to the projectile diameter, ρ_p is the density of the projectile, ρ_t is the density of target, T stands for the thickness of the target, m_t is the mass removed from the target, and m_p is the mass of the projectile.

A nondimensionalized impact velocity, V_p , is also defined as

$$V_p = \frac{u_p^2}{2\xi_b}. \quad (2)$$

where ξ_b is actually a nondimensional kinetic energy, ξ_b is considered an average energy per unit mass needed to break up or vaporize the materials involved in the penetration process. This term is integrated over the space and time involved in the penetration and debris generation process.

The half angle of the debris expansion, θ , is determined from $\theta = \sin^{-1}(u_e/u_c)$, where u_e is the debris expansion velocity, and u_c is the debris cloud center-of-mass velocity. The half angle of the debris expansion, θ , can also be written as

$$\theta = \sin^{-1} \left(M - \frac{(1+M)^2}{V_p} \right)^{1/2}. \quad (3)$$

The hole-saw ring diameter, D, can be calculated by

$$D = 2 L \tan \theta. \quad (4)$$

Combine Eq.(3) and (4), the ring diameter is

$$D = 2 L \left(\frac{V_p M - (1+M)^2}{V_p - V_p M + (1+M)^2} \right)^{1/2}. \quad (5)$$

Replace the constants in the Eq.(1) by κ , and rewrite it in the form

$$M = \frac{3 \rho_t}{2 \rho_p} \left(\frac{D_h}{D_p} \right)^2 \left(\frac{T}{D_p} \right) = \kappa \left(\frac{D_h}{D_p} \right)^2 \left(\frac{T}{D_p} \right) = \kappa X, \quad (6)$$

and substitute this equation into Eq.(5) to obtain

$$D = 2 L \left(\frac{-\kappa^2 X^2 - (2+V_p)\kappa X - 1}{(2+V_p)\kappa^2 X^2 - (2-V_p)\kappa X + 1} \right)^{1/2}, \quad (7)$$

a constant V_p for the fixed impact velocity and projectile/target material and thickness. Eq.(7) offers a mathematical relationship between the hole-saw ring diameter and the geometric variable X.

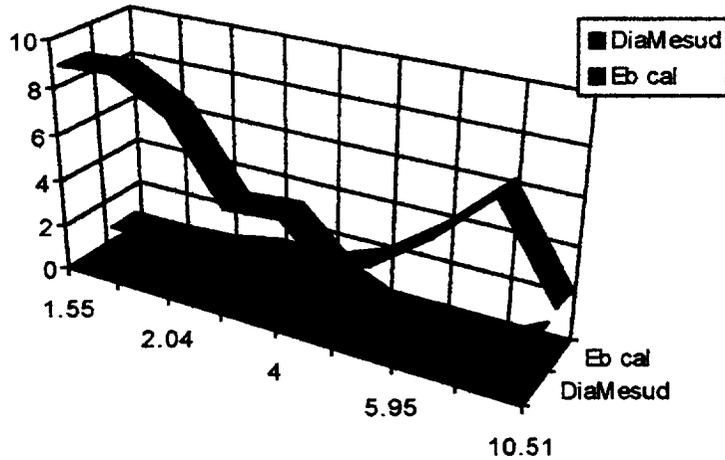
Experimental Data Analysis

Results of the measurements of the ring diameters and the related information are presented in the table of Hole-saw Ring Data. The ring pattern starts at $D_p/T = 1.59$ and disappears at about $D_p/T = 15$. The lead targets began showing the ring pattern at $D_p/T = 1.27$, however the circular pattern has no specific craters that formed the rim of the ring. The rings at this regime have a radically diffused shape. The well defined hole-saw rings started at $D_p/T = 12.50$ and up to $D_p/T = 20.89$ in our observations. For the teflon target, we have employed different impact velocity, from 2.3 km/s to 6.3 km/s, and various target thickness. At low velocity, 2.3 km/s, the ring pattern started to appear at approximately $D_p/T = 3.90$, and faded away at $D_p/T = 63.5$. The true hole-saw ring has not ever appeared in this impact velocity regime. At 4 km/s velocity, some diffused ring started to show up at $D_p/T = 1.57$, a very nice ring was founded at $D_p/T = 6.35$, and then it disappeared at $D_p/T = 31.75$. The target plates of aluminum alloy 1100 showed the ring pattern from $D_p/T = 1.98$ to $D_p/T = 4.02$.

HOLE-SAW RING DATA
Measured and Computed Values

Target Material	Impact Velocity	D_p (cm)	Shot number	L Standoff	D_p/T	D_t/D_p (cm)	ξ_b computed (10^{10})	$D_{measured}$ (cm)	ρ_t/ρ_p
Inconel	5.94	0.3175	1109	12.06	4.41	1.61	0.292	2.306	3.24
Inconel	5.76	0.3175	1110	12.06	5.68	1.72	0.227	6.783	3.24
Inconel	5.96	0.3175	1120	12.06	7.82	1.44	0.405	3.691	3.24
Inconel	6.03	0.3175	1119	12.06	9.98	1.34	0.507	3.678	3.24
Lead	6.38	0.3175	65	12.06	12.5	1.57	0.770	5.990	4.11
Lead	6.1	0.3175	66	12.7	20.89	1.51	0.784	2.637	4.11
Teflon	4.03	0.3175	84	13.0	6.35	1.51	1.20	1.201	0.87
Teflon	6.0	0.3175	501	12.46	12.5	1.30	3.44	1.355	0.87
Teflon	6.25	0.3175	34	12.06	1.94	2.68	0.704	6.970	0.87
Teflon	6.31	0.3175	35	12.06	3.97	1.89	1.80	3.115	0.87
Teflon	6.32	0.1000	534	4.76	1.33	3.20	0.436	3.654	0.87
Teflon	6.28	0.1000	547	4.76	3.94	2.16	1.23	3.267	0.87
Teflon	6.25	0.1000	536	4.76	3.94	--	--	2.963	0.87
Al1100	3.85	0.3175	1166	12.06	3.18	1.57	0.853	2.390	1.11
Al1100	5.16	0.3175	1159	12.06	3.11	1.86	1.02	3.826	1.11
Al1100	5.87	0.3175	789	11.35	1.98	2.76	0.456	6.768	1.11
Al1100	5.89	0.3175	793	6.35	1.98	2.78	0.447	4.328	1.11
Al1100	5.86	0.3175	794	6.03	2.40	2.32	0.725	3.624	.11
Al1100	6.03	0.3175	795	6.03	3.01	2.17	0.916	3.010	1.11
Al1100	5.95	0.3175	790	11.35	4.02	2.14	0.943	4.060	1.11
Al1100	5.90	0.3175	740	13.02	4.00	2.11	0.966	3.670	1.11
Al6061	5.92	0.3175	1065	12.06	1.55	2.33	0.743	8.830	1.11
Al6061	6.00	0.3175	1480	12.06	1.98	2.43	0.680	8.951	1.11
Al6061	5.70	0.3175	1483	48.24	1.98	2.43	0.625	32.53	1.11
Al6061	6.06	0.3175	1057	12.06	2.04	2.05	1.10	7.678	1.11
Al6061	5.75	0.3175	1465	12.06	3.13	1.92	1.17	4.399	1.11
Al6061	5.90	0.3175	1479	48.24	3.13	1.90	1.29	22.23	1.11
Al6061	5.86	0.3175	1471	12.06	4.00	1.77	1.49	4.593	1.11
Al6061	5.90	0.3175	1477	48.24	4.00	1.72	1.64	18.31	1.11
Al6061	5.95	0.3175	1484	96.48	4.00	1.80	1.50	35.78	1.11
Al6160	5.65	0.3175	1470	12.06	5.00	1.64	1.66	2.436	1.11
Al6061	5.83	0.3175	1475	48.24	5.00	1.63	1.82	12.41	1.11
Al6061	5.72	0.3175	1469	12.06	5.95	1.52	2.02	1.342	1.11
Al6061	5.20	0.3175	1476	48.24	5.95	1.46	1.83	3.978	1.11
Al6061	6.00	0.3175	1481	12.06	7.81	1.36	2.78	0.929	1.11
Al6061	5.86	0.3175	1482	48.24	7.81	1.33	2.78	4.014	1.11
Al6061	5.91	0.3175	1058	12.06	10.51	1.22	3.25	2.323	1.11

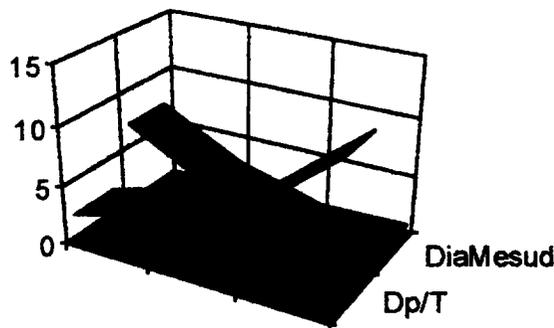
The Aluminum 6061 target has been studied with a systematic change of D_p/T values from 1.55 to 10.51. We tried to control the projectile impact velocity as close as we can to a fixed speed at 6 km/s. The stand-off distance from the target to the witness plate was altered for the different D_p/T values. This procedure helped to uncover the center portion of the ring in much more detail. A major part of the projectile material was melted and the melted material spread on the witness plate produced a web like pattern inside the ring. When the stand-off distance was increased, the web configuration was enhanced and linked the craters on the rim of the hole-saw ring into a circle. The craters inside the ring due to the secondary impact are mostly less than $50 \mu\text{m}$ for a D_p/T ratio less than 4. When the D_p/T ratio increased the smaller craters disappeared, and the ring diameter decreased as well.



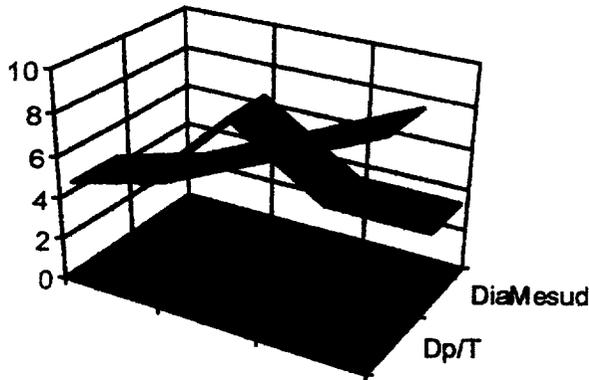
When the D_p/T ratio increases the diameter of the ring decreases and the average energy per unit mass, ξ_b , increases to the maximum value of 3.0×10^{10} erg/gm at about $D_p/T = 8.5$.

For Teflon target, the graph shows the same pattern of the D_p/T - Diameter relationship. The inconel target behaved quite differently.

Teflon



The ring diameter increases with respect to D_p/T ratio increases. After reached a maximum value, it decreased as D_p/T ratio increased.



For aluminium 1100 alloy, the diameter of the ring reaches a maximum value at a early stage, then start to drop with respect to the increasing of D_p/T ratio. After reaching a minimum value it started to increase.

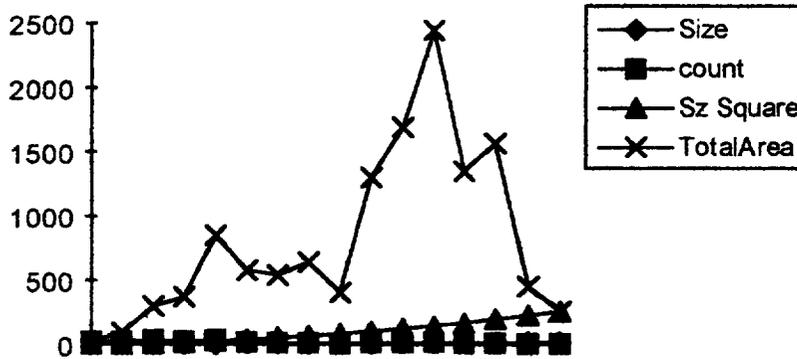
Aluminium 1100



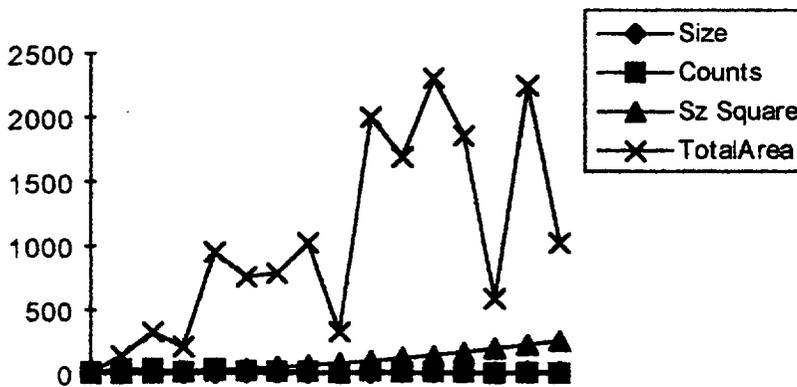
Crater Distribution On The Ring

The craters observed on the hole-saw ring can be classified into four categories: (1) craters produced by multiple impacts, (2) craters produced by melted projectiles which formed a unusual shape, (3) long string like craters, and (4) single impact created craters. Most the craters that can be seen by the naked eye belong to the first three groups. We have rarely seen the clean, single large craters. The craters in the group (2) caused the most difficulty when we counted the craters and tried to measure the crater diameter. It was found that all the craters on the ring were produced by the projectiles. The melted projectile material, soda-lime glass, was observed in the craters. At $D_p/T = 4$,

the increase of the stand-off distance shifted the maximum crater occupied area from about 1500 μm craters to a double peak at 1100 μm and 1400 μm craters. The same behavior happened for the D_p/T ratio at 5. The later one shifted from 1200 μm to 1100 μm and 600 μm . The crater distributions are shown in the following charts.



Crater Distribution at $D_p/T= 4$



Crater Distribution, $D_p/T= 4$

The first chart above is the witness plate located at 12.06 cm from the aluminium 6061 target. The second chart is the witness plate located at 28.24 cm from the target plate.

A crater equation by A. Watts (7),

$$\left(\frac{d_c}{d_p}\right) = 1.0857 \left(\frac{\rho_p}{\rho_t}\right)^{0.2857} \left(\frac{\rho_t}{Y_t}\right)^{0.2857} \left(\frac{c_t}{c_p}\right)^{0.2857} \frac{u_o^{0.5714}}{\left(1 + \left(\frac{\rho_p}{\rho_t}\right)^{1/2}\right)^{0.5714}}, \quad (8)$$

was employed to calculate the fragment mass from the projectile which created the hole-saw ring. The fragment mass distribution is very similar to the crater distribution. Based

on this calculation, the total mass of the fragments that produced the craters on the ring was only a small fraction of the projectile mass.

DISCUSSION AND CONCLUSION

The analytic model was used to calculate the functional relationship between the diameter of hole-saw ring and the D_p/T ratio by assuming that the average breakup energy per unit mass was constant. We selected a known ring diameter value to calculate the ξ_b value, and used the calculated result as a constant to calculate the ring diameter under a different D_p/T ratio in the same impact velocity regime. This procedure shows that the constant ξ_b value is not valid. Our ring diameters calculated by using Eq.(7) differ from the experiment result very much. The melting string and multiple impacted crater could be causes of the deviation.

A wide range of the D_p/T ratio for different target material was observed in our experiment. The ring patterns are observed at $D_p/T=1.98$ to 4.02 for aluminium 1100 targets, at $D_p/T=1.55$ to 10.51 for aluminium 6061 targets, at $D_p/T=4.41$ to 9.98 for inconel targets, at $D_p/T=1.33$ to 3.94 for teflon targets, and at $D_p/T=12.5$ to 20.89 for lead targets.

During the ASEE/NASA summer program period, we have utilized the CTH hydrocode to simulate this complicated phenomenon. The specific ring patterns have not been reproduced successfully. Further investigation is needed. The special hole-saw ring pattern will serve as a very good boundary condition for the further understanding of the hypervelocity impact phenomena.

REFERENCES

1. F. Horz, M. Cintala, R. Bernhard and T. See, Dimensionally Scaled Penetration Experiments: Aluminium targets and glass projectiles 50 μm to 3.2 mm in diameter. *Int. J. Impact Engng* Vol. 15, No. 3, pp. 257-280 (1994).
2. A. J. Stilp, V. Hohler, E. Schneider and K. Weber, Debris cloud expansion studies. *Int. J. Impact Engng* 10, 543-559 (1990).
3. A. J. Pietkutowsky, A simple dynamic model for the formation of debris cloud. *Int. J. Impact Engng* 10, 453-472 (1990).
4. X. Weng and C. H. Yew, Hypervelocity impact of two sphere. *Int. J. Impact Engng* 8, 229-240 (1989).
5. D. E. Grady and S. L. Passmann, stability and fragmentation of ejecta in hypervelocity impact. *Int. J. Impact Engng* 10, 197-212 (1989).
6. R. J. Lawrence, A simple model for the optimization of stand-off hypervelocity particle shields. *Int. J. Impact Engng*. Vol. 5, pp. 451-461 (1987).
7. A. Watts, D. Atkinson and S. Rieco, Dimensional scaling for impact cratering and perforation. NASA contractor report, NCR 188259, March 16, 1993.

59-34

4972

Baseline Experimental Investigation of an Electrohydrodynamically Assisted Heat Pipe

Final Report
1994 NASA/ASEE Summer Faculty Fellowship Program
Johnson Space Center

Submitted by:
A. B. Duncan¹

Assistant Professor
University of Illinois at Chicago

NASA/JSC

Directorate: Engineering

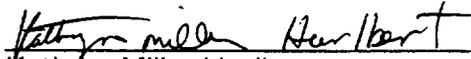
Division: Crew and Thermal Systems

Branch: Thermal Systems

JSC Colleague: Kathryn Miller Hurlbert

Date Submitted: August 4, 1994

Contract Number: NGT-44-005-803


Kathryn Miller Hurlbert


Allen B. Duncan

¹Direct All Correspondence to:
Dr. A. B. Duncan
Department of Mechanical Engineering (M/C 251)
University of Illinois at Chicago
Engineering Research Facility
842 W. Taylor St.
Chicago, IL 60607-7022

ABSTRACT

The increases in power demand and associated thermal management requirements of future space programs such as potential Lunar/Mars missions will require enhancing the operating efficiencies of thermal management devices. Currently, the use of electrohydrodynamically (EHD) assisted thermal control devices is under consideration as a potential method of increasing thermal management system capacity. The objectives of the currently described investigation included completing build-up of the EHD-Assisted Heat Pipe Test bed, developing test procedures for an experimental evaluation of the unassisted heat pipe, developing an analytical model capable of predicting the performance limits of the unassisted heat pipe, and obtaining experimental data which would define the performance characteristics of the unassisted heat pipe.

The information obtained in the currently proposed study will be used in order to provide extensive comparisons with the EHD-assisted performance observations to be obtained during the continuing investigation of EHD-Assisted heat transfer devices. Through comparisons of the baseline test bed data and the EHD assisted test bed data, accurate insight into the performance enhancing characteristics of EHD augmentation may be obtained. This may lead to optimization, development, and implementation of EHD technology for future space programs.

INTRODUCTION

The power demand and thermal management requirements of future space programs will require improving the efficiencies of thermal management devices. Currently, electrohydrodynamically assisted heat pipes are being considered for implementation in spacecraft thermal management systems. Margo and Seyed-Yagoobi (1993) observed increases of up to 71% in forced and free convection heat transfer by utilizing EHD assistance in a convection loop. The intent of the current investigation was to obtain baseline (i.e., without the operational EHD pump) performance data for a EHD-assisted heat pipe in order to accurately quantify the enhancing characteristics of the EHD pump.

Electrohydrodynamic pumping occurs when an electric field interacts with free electric charges in a dielectric fluid. The electric field induces an electromotive force on the free electric charges, dragging the liquid and providing a pumping effect on the liquid medium. This phenomena is illustrated in fig. 1.

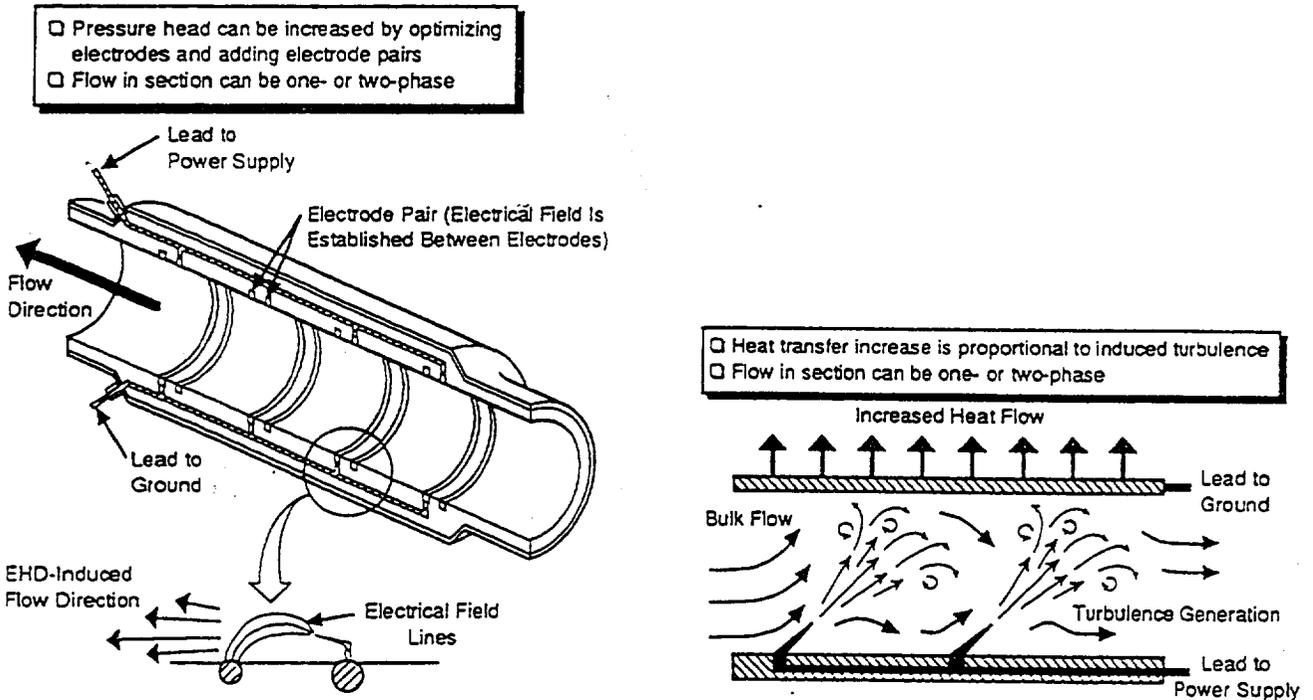


Fig. 1. Principles of the Electrohydrodynamic Pump

EXPERIMENTAL INVESTIGATION

A schematic of this EHD-Assisted Heat Pipe Test Bed utilized in the current investigation is presented in fig. 2. Build-up of the EHD-assisted heat pipe test bed has been completed. As previously stated, the current investigation is intended to obtain baseline performance data for the heat pipe operating without EHD assistance. In order to do this, the maximum heat transport capacity of the unassisted heat pipe will be experimentally determined. Using freon 113 as the heat pipe working fluid, heater power input, condenser temperature, and heat pipe adverse tilt will be varied in order to determine the operating characteristics of the unassisted heat pipe. This will be accomplished by measuring the temperature profiles of both the liquid and vapor flow channels during steady-state operation. A test matrix for this procedure is illustrated in Table 1. Upon completion of this portion of the experimental investigation, data will be compiled and presented in an orderly fashion for further analysis.

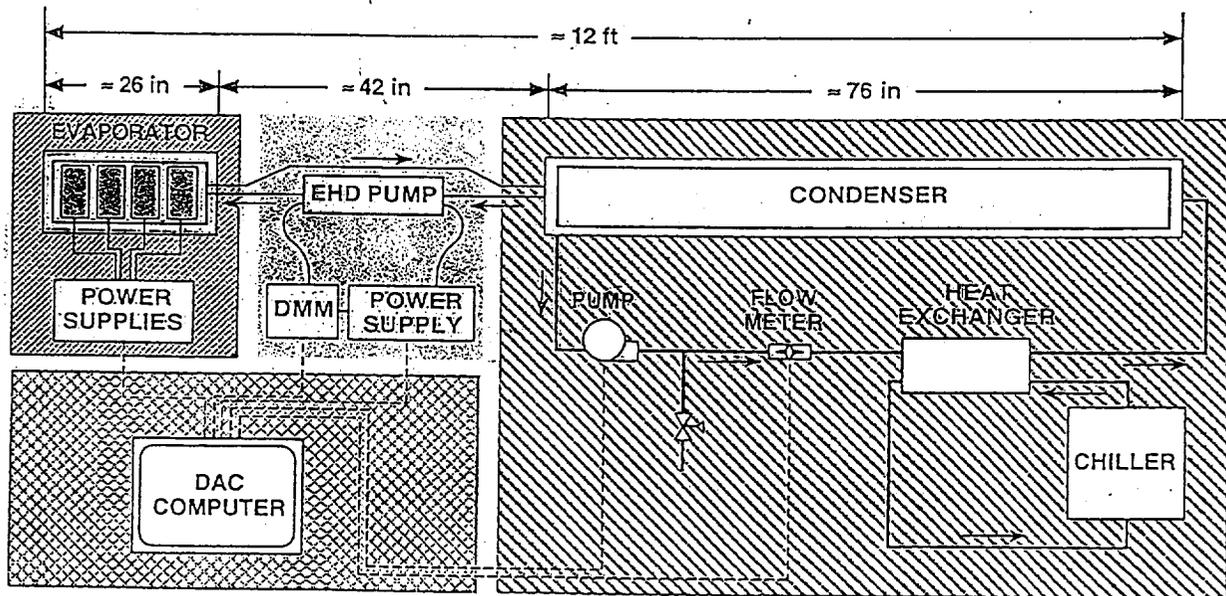


Fig. 2 EHD-Assisted Heat Pipe Schematic

Test Point	Test Stand Inclination (Deg)	Total Power (W)	Evaluation parameter
1	0 to 3.60°, 0.36° increments	100	RTD 21 to 44, TC 45 - 50
2	0 to 3.60°, 0.36° increments	200	RTD 21 to 44, TC 45 - 50
3	0 to 3.60°, 0.36° increments	300	RTD 21 to 44, TC 45 - 50
4	0 to 3.60°, 0.36° increments	400	RTD 21 to 44, TC 45 - 50
5	0 to 3.60°, 0.36° increments	500	RTD 21 to 44, TC 45 - 50
6	0 to 3.60°, 0.36° increments	600	RTD 21 to 44, TC 45 - 50
7	0 to 3.60°, 0.36° increments	700	RTD 21 to 44, TC 45 - 50
8	0 to 3.60°, 0.36° increments	800	RTD 21 to 44, TC 45 - 50
9	0 to 3.60°, 0.36° increments	900	RTD 21 to 44, TC 45 - 50
10	0 to 3.60°, 0.36° increments	1000	RTD 21 to 44, TC 45 - 50
11	0 to 3.60°, 0.36° increments	1100	RTD 21 to 44, TC 45 - 50
12	0 to 3.60°, 0.36° increments	1200	RTD 21 to 44, TC 45 - 50
13	0 to 3.60°, 0.36° increments	1300	RTD 21 to 44, TC 45 - 50
14	0 to 3.60°, 0.36° increments	1400	RTD 21 to 44, TC 45 - 50
15	0 to 3.60°, 0.36° increments	1500	RTD 21 to 44, TC 45 - 50
16	0 to 3.60°, 0.36° increments	1600	RTD 21 to 44, TC 45 - 50

Table 1. Experimental Investigation Test Matrix

ANALYTICAL INVESTIGATION

An analysis of the EHD-assisted heat pipe was performed in order to predict the baseline performance characteristics. Using classical closed-form heat pipe analysis (Chi, 1976), the performance limitations of the freon-charged heat pipe were calculated.

Generally, there are five possible performance limitations for an operating heat pipe. The "sonic" limit is characterized by choked vapor flow in the heat pipe. The boiling limit occurs when nucleate boiling in the heat pipe evaporator, produced from high heat flux levels, creates an unstable region which may not be sufficiently wetted, leading to dry-out. The entrainment limit occurs when the counterflowing liquid and vapor produce a "tearing off" of liquid, also leading to evaporator dry-out. The viscous limit occurs when the vapor pressure of the working fluid is not great enough to drive vapor flow, a limit often associated with low temperature heat pipes. The capillary limit is that limit which occurs when the available capillary pumping pressure in the liquid phase is not great enough to overcome the other pressure drops associated with heat pipe operation such as the pressure drop in the liquid and vapor flow paths and the hydrostatic pressure drop.

An analysis of the operating limits of the Grumman Monogroove heat pipe used in the EHD-assisted heat pipe test bed has indicated that the primary performance limit associated with the test bed is the capillary limit. An analysis of the capillary limit of the EHD-assisted heat pipe test bed, similar to that performed by Ochterbeck and Peterson (1990) has been carried out in the current investigation.

Development of the Analytical Model

The capillary limitation of the Grumman Monogroove heat pipe may be defined by two separate pressure balance statements. First, since the driving pressure difference for return of the liquid to the evaporator is defined by the liquid/vapor interface in the evaporator, the pressure drop across this interface must be equal to the sum of the pressure drops on a path taken from this evaporative interface to the point of liquid replenishment in the evaporator. This may be stated mathematically as follows:

$$\Delta P_{wallcapillary} = \Delta P_{vapor} + \Delta P_{liquid} + \Delta P_{wallwick} + \Delta P_{tilt} + \Delta P_{headdia}$$

In addition to this requirement, the height of the liquid in the axial groove of the heat pipe must be sufficient to continuously supply the wall wicks with liquid. In other words, the liquid height in the axial groove must be great enough to replenish the leading edge of the wall wicks. This may be expressed:

$$\Delta P_{groove.capillary} > \Delta P_{vapor} + \Delta P_{liquid} + \Delta P_{tilt}$$

Each of the terms in the two pressure balance statements may be determined through analysis of the test bed heat pipe.

The pressure drop across the wall capillary in the heat pipe evaporator may be determined from the equation of Young and Laplace (Adamson, 1990).

$$\Delta P_{\text{wall, capillary}} = \frac{2\sigma \cos(\theta + \alpha_w)}{W_w}$$

where θ is the wetting angle of the fluid, α_w is the groove taper angle, and W_w is the wall wick width at the height of the meniscus. The pressure drop across the groove meniscus is:

$$\Delta P_{\text{groove, capillary}} = \frac{2\sigma \cos(\theta + \alpha_g)}{W_g}$$

The hydrostatic pressure drop associated with the heat pipe orientation may be stated as

$$\Delta P_{\text{ult}} = \rho_{\text{liquid}} g h.$$

ρ_{liquid} represents liquid density while g is the acceleration due to gravity and h is the adverse tilt of the heat pipe.

The pressure drop associated with flow of the liquid phase may be approximated by considering it as steady-state incompressible pipe flow. As shown in Chi (1976)

$$\Delta P_{\text{liquid}} = \frac{2(f \text{Re})\mu_l Q L_{\text{eff}}}{\rho_l H_{fg} A_l D_l^2}$$

where L_{eff} is the effective length of the heat pipe, f is the Fanning friction factor associated with the flow, Re is the Reynolds' number, Q is the transported heat, A_l is the area of the liquid flow path, D_l is the hydraulic diameter of the liquid flow path, and H_{fg} is the latent heat of vaporization of the working fluid. A similar expression has been derived for the vapor flow (Chi, 1976).

In the wall-wick structure, similar assumptions were used to calculate the pressure drop term. The wall wick pressure drop may be calculated as

$$\Delta P_{\text{wallwick}} = \frac{32\mu_l \pi D_v Q}{8\rho_l H_{fg}} \left[\frac{1}{(nA_w D_w^2 L)_{\text{evap}}} + \frac{1}{(nA_w D_w^2 L)_{\text{cond}}} \right]$$

where n is the number of grooves per unit length.

Utilization of the individual pressure drop terms into the governing capillary limit equations allows for the determination of the maximum heat transport capacity. The following assumptions were used in developing the computer model which is presented in Appendix A.

- i) The heat pipe operated isothermally, allowing for the calculation of fluid properties at one temperature.
- ii) Vapor flow was incompressible (verified by checking the Mach number).
- iii) The heat pipe was optimally charged.
- iv) The freon fully wet the aluminum surface.
- v) The evaporation and condensation of the freon was uniform across the surfaces of the evaporator and condenser.

Results of the Analytical Model

The maximum heat transport capacity as a function of temperature of the heat pipe is illustrated in fig. 3. As indicated in fig. 3, a maximum heat transport value of 420 Watts is predicted at an adiabatic operating temperature of 125 °C. Fig. 4 illustrates the affects of adverse tilt on the performance limit of the heat pipe test bed. At the optimal operating temperature, maximum heat transport capacity is expected to decrease from 420 W to nearly 55 W as the condenser height is increased from 0 mm to 50 mm.

These predictions seem quite reasonable when compared with other analytical models of the Grumman Monogroove heat pipe (Ochterbeck and Peterson, 1990). The results indicate that the available power to the evaporator of the EHD-assisted heat pipe test bed will be quite adequate (1600 W), as will be the adverse tilt height capability (250 mm).

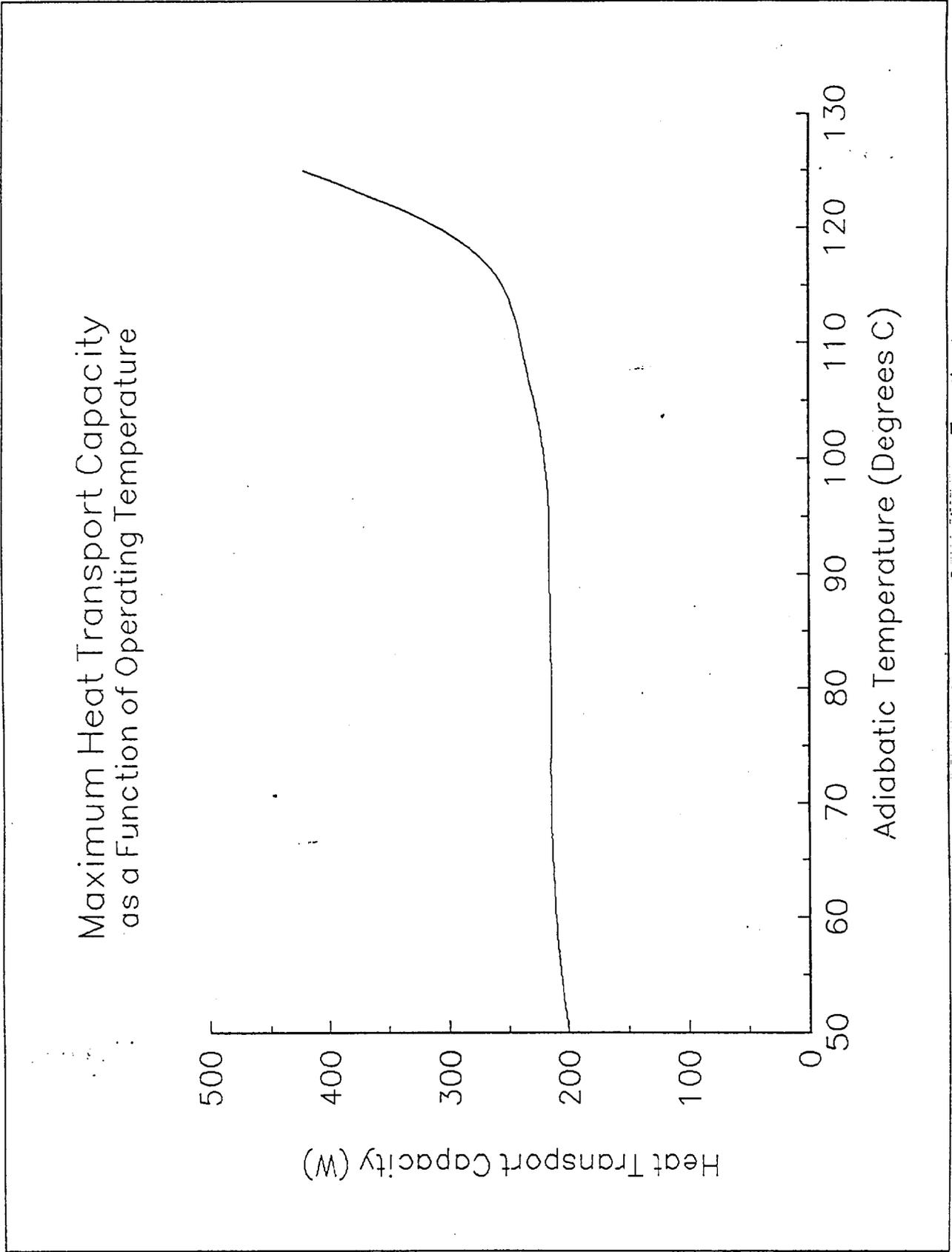


Fig. 3 Maximum Heat Transport Capacity as a Function of Operating Temperature

Maximum Heat Transport Capacity (1.25 Degrees C) as a Function of Adverse Tilt

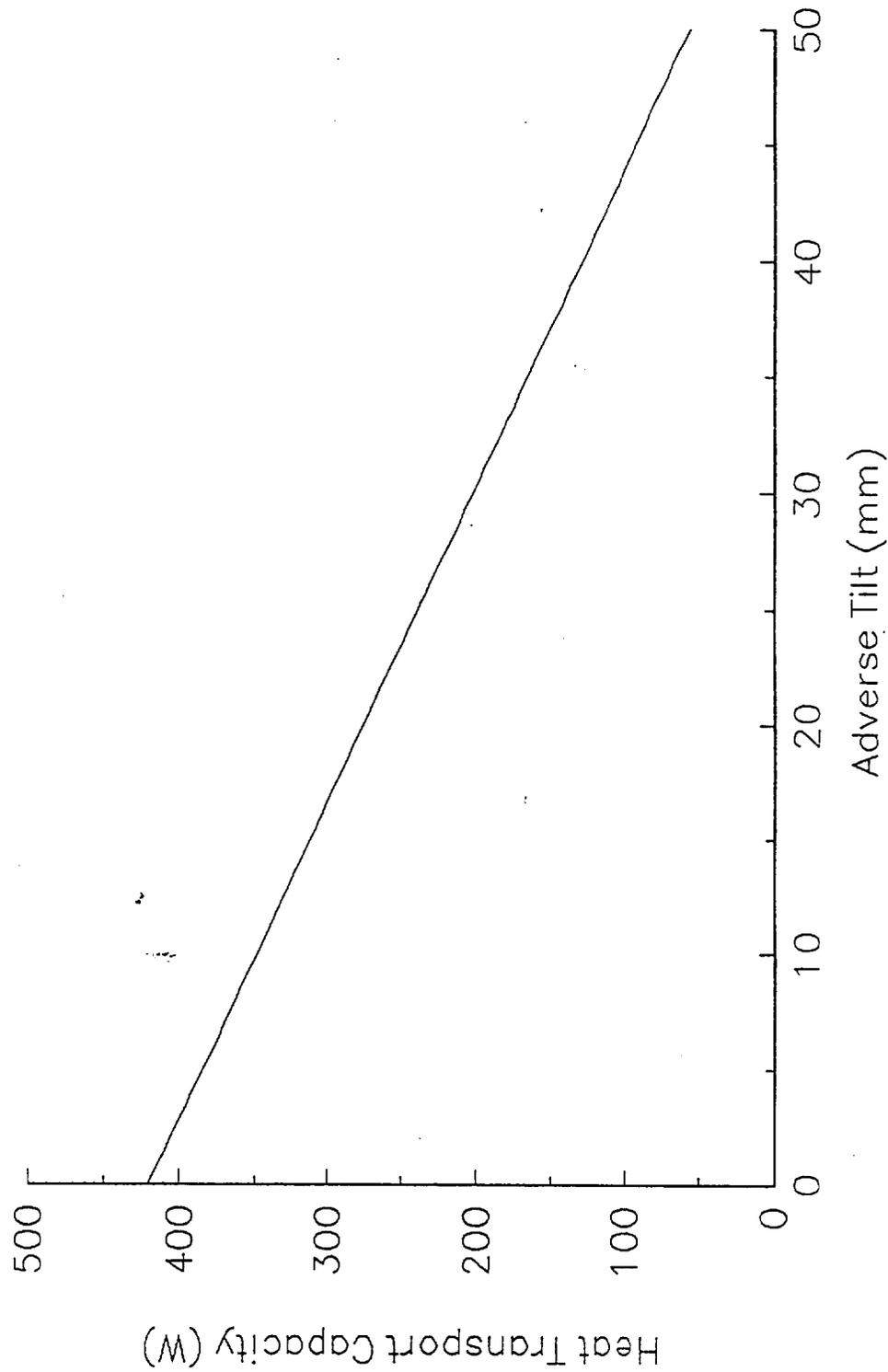


Fig. 4 Maximum Heat Transport Capacity as a Function of Adverse Tilt

CONCLUSIONS

The EHD-Assisted Heat Pipe Test Bed has been successfully fabricated and is ready for comprehensive experimental evaluation. An analytical model of the unassisted heat pipe has been developed which suggests that the capabilities of the experimental facility are well within those which will be required for evaluation. The analytical model has predicted that the maximum heat transport capacity of the heat pipe will be capillary limited and will be approximately equal to 420 W. Additionally, the affects of adverse tilt on the heat transport capacity of the unassisted heat pipe have been modeled. Results suggest that the maximum heat transport capacity of the unassisted heat pipe will decrease by more than 80% at an adverse tilt of 50 mm.

The experimental data to be obtained from the unassisted heat pipe will provide accurate insight into the performance characteristics of the heat pipe. These experimental results should be compared with the results of the modeling effort and an attempt should be made to account for any significant discrepancies. The combined results of the analytical and experimental investigations should lay a solid foundation for evaluation of the heat pipe utilizing EHD assistance.

REFERENCES

Adamson, A. W., 1990, *Physical Chemistry of Surfaces*, 5th Edition, John Wiley and Sons Publishing, New York, NY.

Chi, S. W., 1976, *Heat Pipe Theory and Practice*, McGraw-Hill Book Company, New York, NY.

Margo, B. D., and Seyed-Yagobbi, J., 1993, "Heat Transfer Enhancement Under Various Orientations Resulting From Attractive Mode Induction Electrohydrodynamic Pumping," American Society of Mechanical Engineers, HTD-Vol. 248, Symposium on Fundamentals of Heat Transfer in Electromagnetic, Electrostatic, and Acoustic Fields.

Ochterbeck, J., and Peterson, G. P., "Development of a Generalized Analytical Model for High Capacity External Artery Heat Pipes," Final Report , Submitted to McDonnell Douglas Space Systems Division, Houston, TX 77058, Contract # A96D-J714-STN-KHA-903296.

APPENDIX A
ANALYTICAL MODEL
FORTRAN PROGRAM

program properties

This program is intended for use in the EHD Heat Pipe Investigation
Freon 113 is used in a modified Grumman monogroove heat pipe
Authors: Edouard Motte and Allen Duncan

Variable Definitions:

DPWC = delta p wall capillary
DPMC - delta p monogroove capillary
T - Temperature, degrees C
Vrho - vapor density, kg/m**3
Lrho - liquid density, kg/m**3
Hfg - Latent heat of vaporization, kJ/kg
sig - surface tension, N/m
Cp - vapor specific heat, kJ/kg*k
muv - vapor viscosity, microPascals*s
mul - liquid viscosity, micropascals*s
Kl - liquid thermal conductivity, W/mK
M - molecular weight, kg/kmol
Real T, Vrho, Lrho, Hfg, sig, Cp, muv, mul, Kl, M
Real DPWC, DPMC, DPT, DPL, DPV, Le, Lc, La, Lover, Leff
Real h, Dv, Dl, Q
Real fRel, fRev, Rev
Le = 0.635
Lc = 1.854
La = 0.889
Lover = 3.378
Dv = 15.24e-3
Dl = 10.16e-3
write(*,*)'input h'
read(*,*)h
fRel = 16.
fRev = 16.
open (unit = 4, file = 'output.dat', status = 'unknown')
M = 187.38
Write(*,*)'Input Temperature (C)'
read(*,*)T
T= 125.
write(4,*)'T= ', T
Leff = (Le + Lc)/2. + La
write(4,*)'Leff',Leff
c Subroutine Lden calculates liquid density
Call Lden(T,Lrho)
c Subroutine Vden calculates vapor density (saturation conditions?)
Call Vden(T, Vrho)
write(4,*)'out of vden', T, Vrho
c Subrouting Latent calculates the heat of vaporization
Call Latent(T,Hfg)
c Subroutine Sigma calculates liquid surface tension
Call Sigma(T, sig)
c Subroutine SpecificHeat calculates the vapor specific heat
Call SpecificHeat(T,Cp)
c Subroutine muvapor calculates the viscosity of the vapor
Call muvapor(T,muv)
c Subroutine muliquid calculates the viscosity of the liquid
Call muliquid(T, mul)
c Subroutine Kliiquid calculates the thermal conductivity of the liquid
Call Kliiquid(T,Kl)
write(4,*)'T = ', T
write(4,*)'Liquid density (kg/m**3)= ', Lrho
write(4,*)'vapor density (kg/m**#)= ', Vrho

```

write(4,*)'Heat of Vaporization (kj/kg) = ', Hfg
write(4,*)'surface tension (N/m)', sig
write(4,*)'Vapor specific heat (kj/kg*k)', Cp
write(4,*)'Vapor viscosity (Pa*s)', muv
write(4,*)'Liquid viscosity (Pa*s)', mul
write(4,*)'Liquid thermal conductivity (W/mK)', Kl
c Subroutine DPwallCap calculates the pressure drop across the liquid
c vapor interface of the wall and monogroove capillaries
  Call DPwallCap(sig, DPWC, DPMC)
  write(4,*)'delta p capillary (Pa)', DPWC
  write(4,*)'delta p monogroove (Pa)', DPMC
c Subroutine DPtilt calculates the pressure drop due to adverse tilt
  call DPtilt(Lrho, h, DPT)
  write(4,*)'Lrho', Lrho
  write(4,*)'h', h
  write(4,*)'DPT', DPT
c Subroutine DPhydro represents the hydrostatic loss associated with the wall
c groove flow
  call DPhydro(Lrho, Dv, DPH)
  write(4,*)'DPH', DPH
c Subroutine Dpliq calculates the pressure drop in the laminar liquid flow
c channel
  Call Dpliq(fRel, mul, Lrho, Leff, Hfg, Dl, DPL)
c subroutine Dpvap calculates the pressure drop associated with the vapor flow
  call Dpvap(fRev,muv, Leff, Vrho, Hfg, Dv,DPV)
c subroutine Heat calculates the heat transport capacity
  call Heat(DPWC, DPMC, DPT, DPH, DPV, DPL, Q)
  call Reynolds(Vrho, Q, Dv, Hfg, muv, Rev)\
  write(*,*)'Q = ', Q
  * If(Rev.gt.2300.) Call TurbHeat(DPWC, DPMC, DPT, DPH, DPL,
  * Q, Leff, Rev,Dv,Vrho, Hfg)
  stop
  end
Subroutine Lden (T,Lrho)
Real T, Lrho
Lrho = 1619.662 - (2.607 * T) - (0.0021 * (T**2))
return
end
Subroutine Vden(T, Vrho)
Real T, Vrho
Vrho = 1.2296 + (0.05363 * T) + (0.0010399 * T**2.) +
% 1.21529e-5*(T**3.)
return
end
Subroutine Latent(T, Hfg)
Real T, Hfg
Hfg = 158.0593 - 0.2675*T - 0.0006 * (T**2)
return
end
Subroutine sigma (T, sig)
real T, sig
sig = 0.0188 - 0.0001 * T
return
end
Subroutine SpecificHeat(T,Cp)
Real T, Cp
Cp = 0.6170 + 0.0010 * T
return
end
Subroutine muvapor(T,muv)
real T, muv
muv = 8.2651 + (0.0635*T) - 0.0002 * (T ** 2)
muv = muv *1.e-6
return
end
Subroutine muliquid(T, mul)

```

```

real T, mul
mul = 990.6387 - 16.7189*T + 0.1862*(T**2) - 0.0009*(T**3)
mul = mul*1.e-6
return
end
Subroutine Kliquid(T,Kl)
real T, Kl
Kl = 0.0801 - 0.0002*T
return
end
Subroutine DPWallCap(sig, DPWC, DPMC)
Real sig, DPWC, DPMC, theta, alphaW, Wwe, Wwc, Wm
Real Rwe, Rwc, Depth
theta = 0.
c alphaW is the taper angle
c Wall wick widths at the height of the meniscus
c evaporator and condenser, respectively
c Wwe represents the wall wick depth at the height of the meniscus, evaporator
Wwe = 0.014e-3
c Wwc - wall wick width, height of meniscus, condenser
Wwc = 0.165e-3
c Rwe - wall wick root width, evaporator
Rwe = 0.014e-3
c Rwc - wall wick root width, condenser
Rwc = 0.089e-3
c Depth - depth of the wall wick grooves
Depth = 0.196e-3
c WM = monogroove width
WM = 0.25e-3
alphaW = Atan((Wwe - Rwe)/(2*Depth))
write(4,*) 'alphaw (rad)', alphaW
write(4,*) 'Wwe', Wwe
write(4,*) 'theta (rad)', theta
write(4,*) 'sigma (N/m)', sig
write(4,*) 'cos(theta + alphaw)', cos(theta + alphaw)
DPWC = ((2*sig*cos(theta + alphaw))/Wwe)
DPMC = 2.*sig*cos(theta)/WM
return
end
Subroutine DPtilt(Lrho, h, DPT)
real Lrho, h, DPT, g
g = 9.81
DPT = (Lrho*g*h)
return
end
subroutine DPhydro(Lrho, Dv, DPH)
real Lrho, Dv, DPH, g
g = 9.81
write(4,*) 'hydro', Lrho, Dv, g
DPH = Lrho*g*Dv
return
end
subroutine DPliq(fRel, mul, Lrho, Leff, hfg, Dl, DPL)
real fRel, mul, Lrho, Leff, hfg, Dl, DPL
DPL = (2.*fRel*mul*Leff)/(Lrho*Hfg*3.14159*(Dl**4.))
write(4,*) 'frel', fRel
write(4,*) 'mul', mul
write(4,*) 'Lrho', Lrho
write(4,*) 'leff', Leff
write(4,*) 'hfg', hfg
write(4,*) 'Dl', Dl
write(4,*) 'DPL', DPL
return
end
Subroutine Dpvap(fRev, muv, Leff, Vrho, Hfg, Dv, DPV)
Real fRev, muv, Leff, Vrho, Hfg, Dv, DPV

```

```

write(4,*)'fRev',fRev,'muv',muv,'Leff', Leff,'Vrho', Vrho
write(4,*)'Hfg', Hfg,'Dv',Dv
DPV = 2.*fRev*muv*Leff/(Vrho*3.14159*Hfg*(Dv**4.))
write(4,*)'dpv',DPV
return
end
Subroutine Heat(DPWC, DPMC, DPT, DPH, DPV, DPL, Q)
Real DPWC, DPMC, DPT, DPH, DPV, DPL, Q
Q = (DPWC-DPT-DPH)/(DPV+DPL)
write(4,*)'Q=', Q
return
end
Subroutine Reynolds(Vrho, Q, Dv, Hfg, muv, Rev)
Real Vrho, Q, Dv, Hfg, muv, Rev
Rev = (Vrho*Q*Dv)/(Hfg*muv)
write(4,*)'vapor reynolds no', Rev
return
end
Subroutine TurbHeat(DPWC, DPMC, DPT, DPH, DPL, Q,
* Leff, Rev,Dv, Vrho, Hfg)
Real DPWC, DPMC, DPT, DPH, DPL, Q, DPVT, F, Leff, Rev
Real A, B, C, Q1, Q2, Hfg, D, E
write(4,*)'DPWC', DPWC, 'DPMC', DPMC, 'DPT', DPT, 'DPH', DPH
write(4,*)'DPL',DPL,'Q',Q, 'Leff', Leff, 'Rev', Rev
write(4,*)'Dv', Dv, 'Vrho', Vrho, 'Hfg', Hfg
F = 2./((2.236*(LOG(Rev))-4.639)**2.)
write(4,*)'fturb',F
D = 8.*F*Leff
E= Vrho*(Hfg**2.)*(3.14159**2.)*(Dv**6.)/16.
DPVT = D/E
A = DPVT
B = DPL
C = -1.*(DPWC-DPT-DPH)
write(4,*)'a',a,'b',b,'c',c
Q1 = ((-1.*B) +SQRT((B**2.)-(4.*A*C))) /(2.*A)
Q2 = ((-1.*B) -SQRT((B**2.)-(4.*A*C))) /(2.*A)
write(4,*)'Q1', Q1
write(4,*)'Q2', Q2
return
end

```

516-54
4913

N95-32428

BICYCLE ERGOMETER INSTRUMENTATION TO DETERMINE MUSCLE AND BONE FORCES DURING EXERCISE

Final Report
NASA/ASEE Summer Faculty Fellowship Program--1994
Johnson Space Center

Prepared By: Fernando Figueroa, Ph.D.
Academic Rank: Assistant Professor
College and Department: Tulane University
Department of Mechanical Engineering
New Orleans, LA 70118

NASA/JSC

Directorate: Space and Life Sciences
Division: Medical Sciences
Branch: Space Biomedical Research Institute
JSC Colleague: Linda C. Shackelford, M.D.
Date Submitted: July 28, 1994
Contract Number: NGT-44-005-803

ABSTRACT

It is hypothesized that bone loss experienced by astronauts in zero gravity conditions may be curtailed by appropriate exercise. According to Wolf's law, bone regenerates when muscles produce stresses by pulling on the bone during daily activity and/or exercise on Earth. To use this theory to prevent or decrease bone loss, one needs to quantify musculoskeletal loads and relate them to bone density changes. In the context of the space program, it is desirable to determine musculoskeletal loads during exercise (using the bicycle ergometer in this case) so that one may make similar measurements on Earth and in space. In this manner, load measurements on Earth may be used as reference to generate similar loads during exercise in space. The work reported in this document entails a musculoskeletal load measurement system that, when complete, will provide forces at muscle insertion points and other contact points, on bone. This data will be used by Dr. Beth A. Todd, who is also a SFF working with Dr. Shackelford, as input to a finite element model of bone sections to determine stress distributions.

A bicycle ergometer has been instrumented to measure parameters needed to determine musculoskeletal forces during exercise. A primary feature of the system is its compactness. It uses small/light sensors without line-of-sight requirements. The system developed includes sensors, signal processing, a data acquisition system, and software to collect the data. The sensors used include optical encoders to measure position and orientation of the pedal (foot), accelerometers to determine kinematic parameters of the shank and thigh, load cells to measure pedal forces on the sagittal plane, and EMG probes to measure muscle activity. The signals are processed using anti-aliasing filters and amplifiers. The sensors' outputs are digitized using 30 channels of a board mounted inside a 486 class PC. A program sets the data acquisition parameters and collects data during a time period specified by the user. The data is put directly into a file on the hard disk in binary form. The 30 channels are sampled at 200 KHz, and each 30 channel scan is done at a rate of 1000 Hz. The instrumented ergometer has been flown in the KC-135 zero-gravity (zero-g) flight to collect information needed to determine musculoskeletal forces under these conditions. Similar information has been collected in 1-g conditions for comparison with the results from the zero-g case. At this time, the sets of data from both experiments are being processed. An existing methodology will be used to determine the kinematic parameters of the shank and thigh using accelerometer and encoder data. This methodology was developed during the fellow's previous NASA/ASEE fellowship and thanks to a Director's Grant. In the future, a methodology to determine the musculoskeletal forces using Newton's Law of Motion and optimization techniques will be developed to determine forces exerted by particular muscles.

INTRODUCTION

It has been determined that astronauts lose bone from the trabecular regions during space flight ¹. This loss is significant even when the permanence in zero-g conditions is as short as two weeks. Because of absence of gravitational forces, it is hypothesized that the reason for bone loss is the decrease of muscle pull activity on the bone during normal daily chores and during exercise. Therefore, one way to quantify bone remodeling activity is to determine the loading history on the bone. One may generate bone loading history during exercise on earth to determine the types of loads required for bone maintenance, and use this data as a template for musculoskeletal loads that must be generated in space. Therefore, musculoskeletal loads in space must also be determined. The instrumented ergometer system described in this report was developed with the objective of determining musculoskeletal forces during exercise on earth and in space.

The system developed uses a set of light and compact sensors that do not have line-of-sight requirements. The sensors measure motion parameters, loads, and electromyographic activity, which will be used to ultimately determine forces exerted by individual muscles. The methodology to determine the kinematics has already been developed. The methodology to determine musculoskeletal forces has been outlined, but has not yet been fully developed.

The sections that follow give a detailed description of each element of the system and its operation, and a description of the experiments performed including all the information needed to interpret the data collected. The task of turning the experimental data into musculoskeletal loads will be briefly outlined, and should be done in the near future.

SYSTEM CONFIGURATION AND OPERATION

A graphical representation of the instrumented ergometer system is shown in Figure 1. The load cells and encoders were connected directly to the data acquisition (DAQ) board. The EMG probes were passed through a recorder to amplify and filter to avoid aliasing. The recorder was used only because there was no other anti-aliasing filtering system available. The accelerometer signals were amplified using factory made low noise amplifier units (approximate gain of 100). A detailed description of each element of the system is provided in the experiment section.

The measurements provided by the above instrumented ergometer system will be used to determine forces exerted by individual muscles using the following two

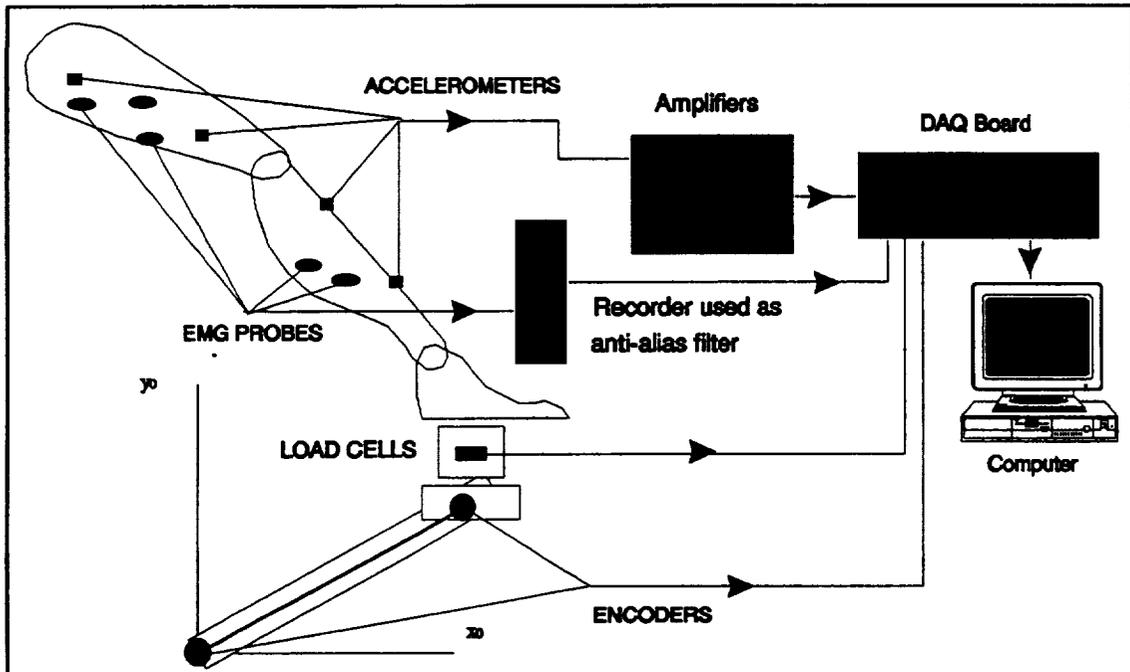


Figure 1.- Instrumented Ergometer System Configuration

step procedure: (1) Use Newton-Euler Equations of Motion of each body segment to determine a set of equations that relate the forces exerted by muscles on the segment bone and forces and torques applied by other bone neighboring sections at joints and contact points; this set of equations is indeterminate, meaning that there will be more unknowns than equations^{2,3,4,5,6,7,8}. (2) Generate additional equations using optimization methods (optimize a criteria function)^{9,2}, heuristic knowledge about muscle activity during portion of a cycling exercise⁷, muscle force sharing indicated by muscle, tendon, and moment arm¹⁰, and other methods. Verification of the results will be done using the EMG measurements.

Part (1) of the procedure above requires the determination of the accelerations (angular and linear) of each body segment. A procedure to determine these accelerations using accelerations measured by the accelerometers and encoders has been developed and successfully simulated¹¹. As mentioned earlier, the methods for the determination of forces is being developed using techniques in part (2) of the procedure outlined in the previous paragraph. The rest of the report describes the experiment and the data collected with two purposes: (1) to analyze qualitatively the values of loads, accelerations, and EMG activity during exercise, and (2) to prepare the data for use in the determination of musculoskeletal loads.

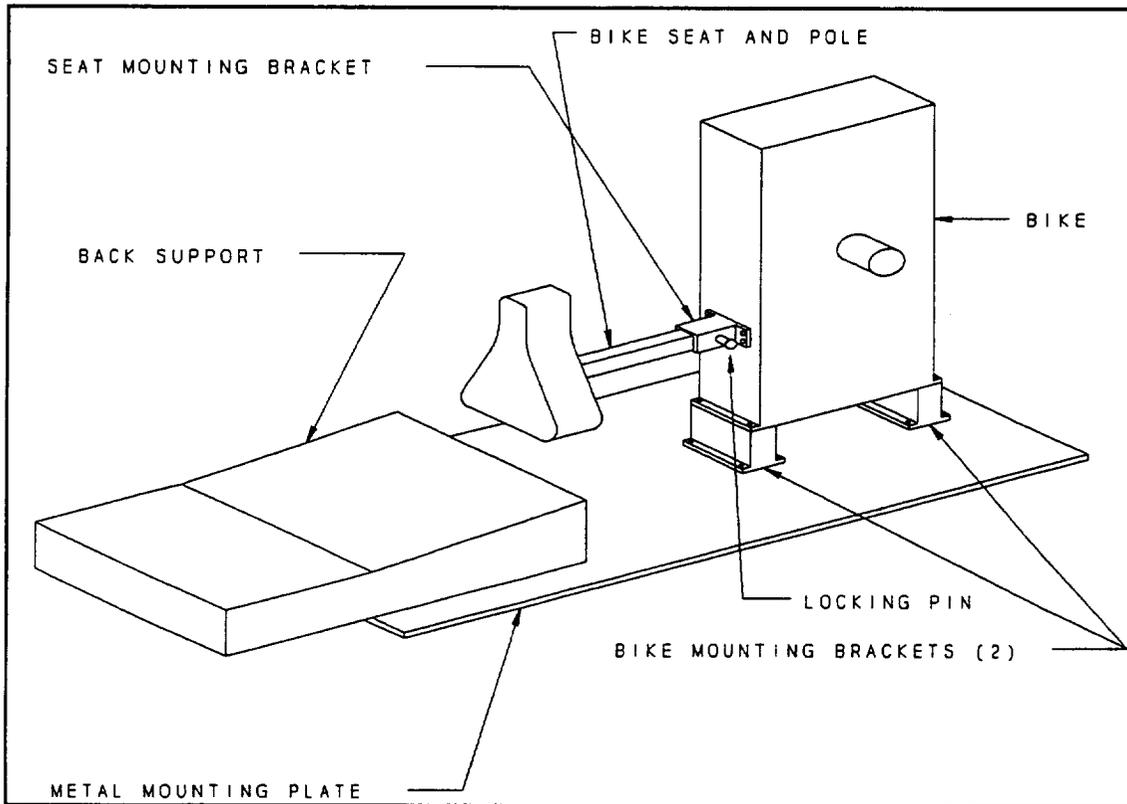


Figure 2.- Bicycle Ergometer mounting

EXPERIMENT

The experiment consisted in running the program after all sensors were mounted and connected. Each run corresponded to the duration of zero-g conditions during one parabolic flight path and lasted 20 or 25 seconds. Before the program was started, the pedal and crank initial positions were marked as a reference. Also, a file name and the experiment duration were input prior to each run. The ergometer assembly and mounting is shown in Figure 2. The overall experiment physical layout is shown in Figure 3. The data for each run was saved in the format shown in Figure 4. Figure 4 represents one scan of all 30 channels used. Data corresponding to the second scan starts in position 31 and ends in position 60. The rest of the data follows the same organization. Each position is filled by a two byte binary representation of an integer.

The experiment runs included various modes of exercise. Table 1 shows a list of the experiment data files and their descriptions.

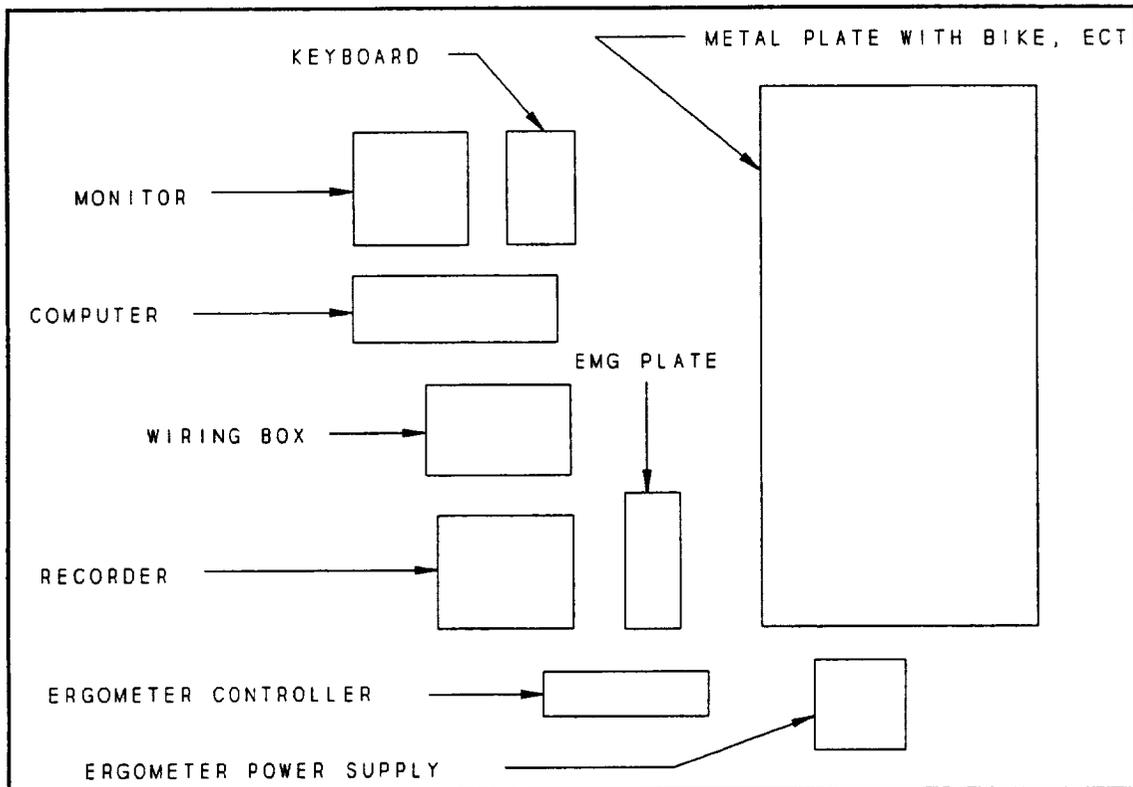


Figure 3.- System layout during the experiment

	ACCEL	EMG	ACCEL	L. C.	P. ENC.	C. ENC.
DIFFERENTIAL PAIR	8 ... 15,		48 ... 51,	52 ... 55,		
CHANNEL #	0 ... 7,	16 ... 23,	24 ... 27,	28 ... 31,	32 ... 34,	35 ... 37
DAQ GAIN	1	1	1	10	1	1
	1 ... 8,	9 ... 16,	17 ... 20,	21 ... 24,	25 ... 28,	27 ... 30

POSITION FROM THE BEGINNING OF THE FILE FOR ONE SCAN

L. C. Load Cell
 P. ENC. Pedal Encoder
 C. ENC. Crank Encoder

Figure 4.- Organization of the Data File

Table 1.- EXPERIMENT DATA FILES

Filename	Description	Init. Pos.
p1.dat ... p7.dat g1.dat, g2.dat	Standard cycling (pedaling) Standard cycling on ground	47° from horiz.
r2.dat ... r10.dat	Resistive (squats)	43° from vert.
pu*.dat	pull ups	
ha*.dat	with shoulder harness	
s*.dat	squats with shoulder harness	
gpu*.dat,	Pull throughs on ground	
gpr*.dat	Presses on ground	
gtr*.dat	toe raises	
gs*.dat	shoulder harness	
gsc1.dat	shoulder harness standard cycling	

A complete description of each sensor system and their operating parameters are presented in the following sections.

Accelerometers

A total of 12 one dimensional accelerometers were used. These were bundled in groups of three to measure three orthogonal components of the acceleration at a point. Thus, three orthogonal components of acceleration were measured at four distinct points, two points on the shank and two points on the thigh (see Figure 1). Table 2 shows a detailed description of the accelerometers and their calibration. Each accelerometer was amplified by approximately 100 prior to digitization. The calibration includes the effects of this amplification. The gain shown is that of the data acquisition system. The nomenclature used to label the table columns translates as follows:

Ch = DAQ channel used

Scn = Position within the 30 channel scan

Gain = DAQ gain setting.

Table 2.- ACCELEROMETER PARAMETERS

ID	Ch.	Scn	Gain	Calibration (Volts)					
				→ 0g	↓ 1g	↑ -1g	↑ - →	→ - ↓	C -2g
SDX	02	3	1	0.36	0.73	.01	-0.35	-0.37	-0.72
SDY	25	18	1	0.23	0.62	-0.09	-0.32	-0.39	-0.71
SDZ	24	17	1	.17	.56	-0.27	-0.44	-0.39	-0.83
SUX	03	4	1	-0.23	.17	-0.69	-0.46	-0.40	-0.86
SUY	26	19	1	1.19	1.45	.83	-0.36	-0.26	-0.62
SUZ	07	8	1	.54	.88	.10	-0.44	-0.34	-0.78
TDX	04	5	1	-0.41	.09	-0.93	-0.52	-0.50	-1.02
TDY	27	20	1	-0.36	.04	-0.86	-0.50	-0.40	-0.90
TDZ	05	6	1	.28	.82	-0.31	-0.59	-0.54	-1.13
TUX	01	2	1	-0.29	.08	-0.71	-0.42	-0.37	-0.79
TUY	00	1	1	-1.05	-0.66	-1.40	-0.35	-0.39	-0.74
TUZ	06	7	1	-1.76	-1.40	-2.17	-0.41	-0.36	-0.77

Load Cells

Load cells measure forces on the sagittal plane on both pedals. Table 3 describes the parameters of these sensors. The x-right load cell was changed for the ground experiments and the table shows its sensitivity.

EMG Probes

These probes measure muscle activity on various muscles. Table 4 describes the parameters of these sensors.

Table 3.- LOAD CELL PARAMETERS

Ld. Cell	Ch.	Scn.	Gain	Sens.	Sign →	Sign ↑
x right	31	24	1	.800 mV/lb .9778 (gnd)	+ x-dir (pull foot to front)	
y right	29	22	1	.343 mV/lb		+ comp.
x left	30	23	1	.992 mV/lb	+ x-dir	
y left	28	21	1	.28 mV/lb		- comp

Encoders

Encoders measure the crank position and the right pedal position. The parameters of the encoders are shown in Table 5.

SOFTWARE

The software for data acquisition was developed using LabWindows/CVI from National Instruments, Inc., Austin, TX. The software and data acquisition board reside in a 50 MHz-486 IBM compatible PC with 16MB of RAM. The data acquisition board used is the AT-MIO-64F-5 board with 64 single ended analog inputs. The program writes the digitized data directly to disk and is run from the LabWindows/CVI environment. Inputs to the program include an ASCII text channel-gain file called "gain30.dat" that resides in the CVI directory, the name of the file for the data, and the scanning frequency, which are input by the user. The scanning frequency used was 1000 Hz.

DATA COLLECTION

Data was collected during the KC-135 zero-g flight and later on ground. During the flight, a data file was created almost for each 20 or 25 seconds of zero gravity time during the parabolic flight paths. The organization of each data file is detailed in Figure 4, and descriptions of each data file are shown in Table 1.

Table 4.- EMG PROBE PARAMETERS

Probe	Ch.	Scn.	Gain	Performance at $\pm 6V$		
				60Hz	500Hz	1000Hz
1 peroneous longus	16	9	10	352	357	358
2 lateral gastrocnemius	17	10	10	349	355	356
3 medial gastrocnemius	18	11	10	341	346	347
4 soleus	19	12	10	347	352	353
5 flexor halluc	20	13	10	365	371	372
6 peroneous brevis	21	14	10	358	364	365
7 extensor digitorum	22	15	10	347	353	354
8 tibialis anterior	23	16	10	363	368	369

DATA PROCESSING

Reduction of the data collected during the experiment has not yet been done. Dr. Shackelford's group, including myself, is currently in the process of extracting relevant information from the experiment's data files.

PENDING TASKS

The pending tasks to determine musculoskeletal forces include:

1. Interpret pedal and crank encoder information.
2. Using the positions of the accelerometers extracted from x-ray images of the subject's leg, define accelerations in the sagittal plane at each accelerometer mounting location on the shank and thigh. Extract acceleration components on the sagittal plane.
3. Use accelerations at the accelerometer locations and the geometry associated with the body section to determine angular velocity and acceleration of the section, and the acceleration of the center of mass.
4. Use the Newton-Euler equations of motion and other optimization and heuristic equations to determine forces exerted by particular muscles.

Table 5.- ENCODER PARAMETERS

Encoder	Ch.	Scn.
Ped. A	32	25
Ped. B	33	26
Ped. X	34	27
Crnk. A	35	28
Crnk B	36	29
Crnk X	37	30

5. Provide musculoskeletal forces and points of application for input to a finite element analysis of the section bone.

CONCLUSIONS AND RECOMMENDATIONS

A bicycle ergometer has been instrumented to measure parameters needed to determine musculoskeletal forces during exercise. This system, which includes hardware and methodologies, was conceived as a tool to help determine the causes for bone loss by astronauts that remain in zero-g conditions during space missions, and also as a tool to synthesize countermeasure exercises to decrease bone loss. The first prototype system was built and tested in zero-g conditions and on ground. The measurement system appears to have operated satisfactorily, and the data is being readied for use with methodologies to determine forces exerted by muscles and bone. These forces will be used as input to a finite element analysis model currently being developed by Dr. Beth A Todd¹² who also works with NASA Colleague Dr. Shackelford.

REFERENCES

1. Oganov, V.S., Grigoriev, A.I., Voronin, L.I., Rakhmanov, A.S., Bakulin, A.V., Schneider, V., LeBlanc, A. "Bone mineral density in cosmonauts after 4.5-6 month long flights aboard Orbital Station Mir". *Aerospace and Environmental Medicine*. 5,6:20-24, 1992.
2. Redfield, R., and Hull, M. L. (1986A). Prediction of pedal forces in bicycling using optimization methods. *J. Biomechanics*, Vol. 19, No. 7, pp. 523-540.
3. Anderson, F. C., Ziegler, J. M., Pandy M. G., and Whalen, R. T. (1993), Numerical computation of optimal controls for large-scale musculoskeletal systems. *Advances in Bioengineering, ASME Winter Annual Meeting*. BED-Vol. 26, pp. 519-522.
4. Yang, Y., Yahia, L. H., and Feldman, A. G. (1993). A versatile dynamic model of human arm. *Advances in Bioengineering, ASME Winter Annual Meeting*. BED-Vol. 26, pp. 527-529.
5. Abdel-Rahman, E., and Hefzy, M. S.(1993). Three-Dimensional dynamic modeling of the tibio-femoral joint. *Advances in Bioengineering, ASME Winter Annual Meeting*. BED-Vol. 26, pp. 315-318.
6. Ericson, M. O., Ekholm, J., Svensson, O., and Nisell, R. (1985). The forces of ankle joint structures during ergometer cycling. *J. of the American Orthopaedic Foot and Ankle Soc.*, Vol. 6, No. 3, pp. 135-142.
7. Harrison, R. N., Lees, A., McCullagh, P. J. J., and Rowe, W. B. (1986). A bioengineering analysis of human muscle and joint forces in the lower limbs during running. *J. of Sports Sciences*, Vol. 4, pp. 201-218.
8. Figueroa, J. Fernando "Loading, electromyograph, and motion during exercise", Final Report, NASA/ASEE Summer Faculty Fellowship Program, Eds: Hyman, W.A. and Goldstein, S.H., 1993, NASA CR-188271, pp. 11-1-11-11.
9. Seireg, A., and Arvikar, R. J. (1973). A mathematical model for evaluation of forces in lower extremities of the musculo-skeletal system. *J. Biomechanics*, Vol. 6, pp. 313-326.
10. Hoy, M. G., Zajac, F. E., and Gordon, M. E. (1990). A musculoskeletal model of the human lower extremity: the effect of muscle, tendon, and moment arm on the moment-angle relationship of musculotendon actuators at the hip, knee, and ankle. *J. Biomechanics*, Vol. 23, No. 2, pp. 157-169.

11. Liu, Shih-Ching, A method to determine the kinematics of the lower limbs of a subject pedaling a bicycle using encoders and accelerometers, M. S. Thesis, Tulane University, Mechanical Engineering Department, May 1994.
12. Todd, Beth A., "Finite Element Modeling of the Lower Extremities," Final Report, NASA/ASEE Summer Faculty Fellowship Program, Eds: Bannerot, R., 1994, NASA CR-?, pp. 29-1-29-?.

APPENDIX

Parts List

The sensors and instrumentation used in the system are listed below. This list is provided so that the reader may access detailed information about these parts by reviewing the manufacturers' literature.

Name	Spec.	Description	Manufacturer
Crank Encoder	ERO-1324	512 lines/rev. optical encoder	Heidenhain, Traunreut, Germany.
Pedal Encoder	M20051221031	512 lines/rev. optical encoder	Dynapar Corp., Gurnee, IL
Accelerometers	EGAXT-10	±10g range strain gauge acc.	Entran Devices, Inc., Fairfield, NJ
Pedal Load Cell (shear)	ELF-TC1000-250	250 lbs range load cell	Entran Devices, Inc., Fairfield, NJ
Pedal Load Cell	ELH-TC590-500	500 lbs range load cell	Entran Devices, Inc., Fairfield, NJ

S11-32
4274

N95-32429

**VALIDATION OF SPACE/GROUND ANTENNA CONTROL ALGORITHMS
USING A COMPUTER-AIDED DESIGN TOOL**

Final Report
NASA/ASEE Summer Faculty Fellowship Program -- 1994
Johnson Space Center

Prepared by:	Dr. Rex E. Gantenbein
Academic rank:	Associate Professor
College and department:	University of Wyoming Dept. of Computer Science Laramie WY 82071-3682
NASA/JSC	
Directorate:	Engineering
Division:	Tracking and Communications
Branch:	Electromagnetic Systems
JSC colleague:	Andrew Benjamin
Date submitted:	5 August 1994
Contract number:	NGT-44-005-803

ABSTRACT

The validation of the algorithms for controlling the space-to-ground antenna subsystem for Space Station Alpha is an important step in assuring reliable communications. These algorithms have been developed and tested using a simulation environment based on a computer-aided design tool that can provide a time-based execution framework with variable environmental parameters.

Our work this summer has involved the exploration of this environment and the documentation of the procedures used to validate these algorithms. We have installed a variety of tools in a laboratory of the Tracking and Communications Division for reproducing the simulation experiments carried out on these algorithms to verify that they do meet their requirements for controlling the antenna systems. In this report, we describe the processes used in these simulations and our work in validating the tests used.

INTRODUCTION

Designing algorithms for real-time control of communications devices is difficult, but even more so is validating the correctness of that design. The consequences of design errors are high in many applications of this type, especially where manned missions are dependent on the system. The users of such systems are often justifiably concerned about the possibility of design errors being propagated through the development process into the completed and deployed system. Design errors can be a significant problem in making computer-driven control systems dependable. In many cases, the manifestations of design faults become noticeable only after implementation, at which point correction can be expensive due to the amount of rework that may be required to eliminate the problem.

One approach proposed to combat this problem is validation of a computer system design prior to implementation. Design validation can improve the quality and reduce the cost of a computer system by eliminating design faults before implementation begins. Since a design cannot be "tested" in the usual manner, other techniques based on formal logic or simulation must be employed to assure both the developers and the users that the system, as designed, will behave according to their expectations and needs.

In this summer's project, the author has explored ways to validate the algorithms developed for controlling the Space-to-Ground Antenna (SGANT) subsystem to be used on board Space Station Alpha for communicating with the Tracking and Data Relay Satellite (TDRS) system. These designs are very complex and thus prone to design faults that might be overlooked using traditional testing methods. The goal of this work is to determine effective ways of validating these designs prior to their implementation to assure the highest quality systems at all stages of the development process.

PROJECT OVERVIEW

The SGANT system is being designed by the Satellite and Communications Systems Division of SPAR Aerospace Ltd. of Quebec, Canada. Dynacon Enterprises Ltd., in Ontario, Canada, is designing three control algorithms for this system under contract to SPAR. The three algorithms control slew, search, and pull-in/tracking for the SGANT system.

To verify the performance of these algorithms and demonstrate their conformance to the requirements specified by

SPAR, Dynacon developed two simulation facilities. These are the Final Verification Simulator (FVS) and the Probability of Successful Search Simulator (PSS) [Dynacon92]. The FVS is a collection of simulations based on a set of common subsystem models and represents a comprehensive model of the SGANT system. The PSS simulator addresses the statistical performance of the TDRS search and acquisition functions in SGANT.

The FVS simulator includes a discrete-time simulation of the control algorithms and body dynamics of the system as deployed both in micro-gravity and in full gravity (for evaluation in a ground-based lab). This simulator produces a time-based history of the algorithm behavior under various parameters. This history is used as input to the PSS simulator, which executes a series of searches over this history to determine the likelihood of successful acquisition of the TDRS signal. Together, the two simulators provide a means of verifying that the control algorithms, as designed, are capable of meeting the required 90% or more probability of a successful search.¹

VALIDATION ENVIRONMENT

The goal of the summer project was to recreate the environment used by Dynacon to validate the SGANT control algorithms. This required the use of a variety of different tools, which were installed on the SUN Sparcstation (commlab) housed in the Communications Laboratory in Building 14 at JSC. In this section we describe briefly the tools used.

MATRIX_x

The FVS was implemented using the SystemBuild software of MATRIX_x, an automated design tool that provides graphic representation of a variety of system building blocks as well as simulation by discrete or continuous time execution is [ISI92]. This engineering system design tool supports the development of graphics-based designs that can be executed on data tables that represent sensor inputs, as well as a hierarchy of components that can be packaged into "super-

¹As pointed out in [Dynacon92], the probability of successful search is an estimate of the probability of acquisition, which is based on the ability of the system both to locate the TDRS target and to pull it in and track it. The computation required to simulate the second probability is, according to the documentation, "too prohibitive at this time to allow a statistically significant estimate" of acquisition.

blocks" in the design to provide composability and incremental development. It also allows user code blocks in programming languages (such as FORTRAN) to be incorporated into a simulation. This feature was used to insert the control algorithms, as implemented, into the simulation for validation.

GNU C compiler

The PSS simulator was written in the C programming language, in conformance with the ANSI C standard. The C compiler provided with the SUN operating system on commlab was not an ANSI compiler, however, but conformed to the older standard of C. Normally, there are some syntactic differences between ANSI and non-ANSI C that can be easily resolved, but in this particular instance, conversion of the PSS programs to non-ANSI (SUN-compatible) C code resulted in the program producing incompatible output.

In an attempt to quickly resolve this conflict, we installed version 2.6.0 of the GNU C compiler, which we retrieved by anonymous FTP from Massachusetts Institute of Technology. This compiler conforms to the ANSI standard syntax and, after installation, was used to compile the PSS programs on commlab. The procedures for retrieving and installing GNU C are included as Appendix A.

As documented later in the report, this compiler was still not able to reproduce the PSS behavior exactly. For this reason, a full, commercial ANSI C compiler was purchased and installed in August 1994. Testing of this compiler's results against those from Dynacon is still being completed.

GNUPLOT

Since the data sets produced by the FVS and PSS simulators are so large (depending on the duration of the simulated run and the size of the time frame, as will be described), it is necessary to have some plotting tools available to look at and summarize the data. The MATRIX_x simulations provide plots of selected output data, but it was inconvenient to invoke this package for a simple plot of data. For this reason, we installed another package available by anonymous FTP from M.I.T., GNUPLOT. This package is not part of the GNU project as is GNU C, but provides a simple environment for producing two- and three-dimensional plots. We found this facility easy to use and very handy for quick analysis of data during this project. The procedures for retrieving, installing, and running GNUPLOT are included as Appendix B.

VALIDATION PROCEDURES

What we attempted to do in this project was to validate the results reported in [Dynacon92] by recreating the simulations in the Communications Laboratory. This involved the following steps:

- running the FVS simulations under MATRIX_x to produce the time history for the tracking and crosstracking error;
- extracting the tracking and crosstracking data into a file that can be used by the PSS simulator;
- running the PSS simulator on the tracking/crosstracking data file; and
- analyzing the results.

In this section, we describe the procedures needed to recreate the simulations. A script for performing simulations is included as Appendix C.

Running the FVS simulation in MATRIX_x

The FVS simulator is driven by batch runscripts executed from a custom MATRIX_x executable file. This executable is currently installed as "fvsim_course3.1" in the directory /usr2/home/designlab/ben/course on commlab.² When new versions of MATRIX_x are installed, or if this file becomes corrupted, it will be necessary to install it. The instructions for making a new custom executable are included as Appendix D.

The "fvsim_course3.1" file is executed to invoke the MATRIX_x core. At the "<>" prompt, the batch runscript for the simulation desired can be started with the MATRIX_x core `exec` command. The runscripts used in the Dynacon simulations are documented in Section 4 ("Algorithm Verification") of the final design report [Dynacon92]. There are several parameters in the runscripts that can change the FVS simulation and affect the results. The most important of these are listed in Table I below.

The runscripts add these parameters and other data values

²This directory is the "home" directory for the simulation process. Throughout this document, we will use a "." to refer to this directory when discussing file structures.

Table I. MATRIX_x runscript parameters.

VARIABLE	VALUE	DESCRIPTION
choice	1,2,3,4,5	select algorithms to run (typically 1)
serial	string	used to name output files
totaltime	integer	simulation time (seconds)
DToutput	decimal	time frame (seconds)
autotrack	0, 1	autotracking simulation: off (0) / on (1)
slew	0, 1	slew simulation: off (0) / on (1)
search	0, 1	TDRS search simulation: off (0) / on (1)
track	0, 1	antenna track simulation: off (0) / on (1)
TDRS	0, 1	custom (0) TDRS motion or start at zenith (1)
Sr	3.5, 26.5	TDRS mean signal level (should match the level used in PSS simulation)
CMSON	0, 1	CMS Estimate noise: off (0) / on (1)
x/y/zrot	decimal	rotation disturbances
x/y/ztran	decimal	translation disturbances
seed1/2/3/4	integer	set CMS position and velocity noise
do_sim<x>	_big, _medium, _slew, _slew2, 2, <blank>	select runscript to execute ("define" at start of the batch file must provide pathname for the runscript)

for the FVS simulation to the MATRIX_x data stack, then invoke another runscript to actually perform the simulation. These runscripts are stored in the directory ./udc_files with the ".udc" extension. Some additional files are also included in this directory that are invoked from the udc runscripts.

Depending on the runscript invoked (and the size of the ensuing simulation), the time required for the simulation can run from just under two hours to approximately 48 hours (for a full simulation). It is advisable to consider specifically what parameters are needed prior to execution. (In Appendix C, we describe the runscripts and parameters needed to prepare data for the PSS simulator.)

Extracting the data for the PSS simulator

Once the FVS simulation has completed (messages describing the percentage of completion will appear; when the simulation has reached 100%, the "<>" prompt will be displayed), a number of variables will have been created by the simulation. These variables are stored in the N-by-57 matrix "y" on the MATRIX_x stack, where N is determined by the time duration parameter "totaltime" and the time frame parameter "DToutput" in the batch runscript. In addition, this data has been saved in a file in the ./output directory, under the file name "out_<serial>" where <serial> is also defined in the runscript. Data from the FVS simulation can be recovered from these file by invoking the MATRIX_x core and loading this file.

Each row of this matrix represents the values of a variable computed at each time frame in the simulation. Of particular interest for the PSS simulation are rows 15 and 16, which contain the values for the track (TK) and cross-track (XTK) boresight tracking angle error created by the search simulation in the FVS. The 35-second spiral search portions of these two vectors should be saved (as an ASCII file, using the MATRIX_x command "fsave") in a separate file for input to the PSS simulator.

This file will contain the values of TK and XTK as data pairs. However, MATRIX_x may not always format the file in two columns, so it is necessary to pass this file through a "filter" program, written by Dynacon, prior to its input to the PSS program. This is a C program that simply reads in the data and writes every third data pair, starting with the first, to standard output, formatting them into two columns. NOTE: MATRIX_x also attaches some labeling information to the front of the saved file (typically three lines of text). The current filter program will not accept these three lines, so

the file must be edited by hand prior to running the filter program. Appendix E contains C code that could be added to the filter program to automatically skip over these lines.

Running the PSS simulation

The TK/XTK data from the verifications simulations performed by Dynacon reside as files with the name "pss_<n>" in the directory ./PSS/csim, where <n> is the simulation run number as described in [Dynacon92]. This directory is also where the filter program resides as "filter.c" for the source code and "filter" for the executable.

Filtering the "pss_<n>" files produces a spiral data file that represents the time history of the simulation for tracking TDRS. Each of the spiral data files has the name "spiraldat<n>" to correspond with the files from which it was produced. These files are used as input to the PSS simulator.

The PSS simulator is composed of four separate C source files: "input.c", "poa.c", "sim.c", and "pwrl.c". Prior to running a PSS simulation on a spiral data file, it is necessary to modify "input.c" to set the number of simulation runs, the spiral data set number <n>, and the mean TDRS signal level (low 3.5 dB, high 26.5 dB). These values should be assigned to the variables "n_runs", "spiral", and "mean", respectively.

In addition, the simulation uses the C random number generator to help determine the probabilistic behavior of the search. In order to exactly recreate the results of a particular Dynacon simulation, the seed for the random number generator (which normally is set to the process id of the executing simulation), must be initialized to the same value as the Dynacon run. This is accomplished by editing the file "poa.c" and setting the variable "pid" to the value used in the original run. This value is recorded in the outputs of the Dynacon simulations in ./PSS/csim under the file names "pss_<n>_<level>.plot", where <n> is the spiral data set number and <level> is "low" or "high" corresponding to the mean TDRS signal level used in the simulation.

Analyzing the outputs of the PSS simulation

The output from the Dynacon PSS simulations is summarized in Appendix F for the corresponding PSS simulation run number, spiral data set number and signal level. (Also included is the random number generator seed used for each simulation, as described in the previous section.) The last column in the table in the Appendix represents the probability of successful

search for a maximum pull-in radius of 0.6 degrees. The specifications call for this value to be 90% or greater.

The actual output from the PSS simulator contains a range of probabilities for pull-in radius values from 0.1 to 0.8 degrees. Obviously, as the radius increases, the probability of successful search will increase. The results were plotted by Dynacon using the XPLOT tool, producing postscript-format-
ted files stored as "pss_<n>_<level>.ps" in the ./PSS/csim directory. They can also be plotted using GNUPLOT. (NOTE: the output files contain a header line describing the spiral data set number, signal level, number of runs, and seed. This line must be manually removed prior to plotting with GNUPLOT.) Examples of these output files and the associated plots are shown in Appendix G.

EVALUATION OF EXPERIENCES

It is difficult to say exactly how well we were able to verify the results of the FVS/PSS simulation during this summer's work. Part of the problem was that we were unable to collect data from Dynacon until well into July, so that our time to work with their simulations and data was effectively halved.

Once we had loaded the information from Dynacon onto commlab (which, incidentally, involved installing additional disk space to hold the data and provide room to run the simulations), our first work was to try to run the PSS simulator on the spiral data sets provided by Dynacon. The PSS simulator was compiled under an ANSI C compiler at Dynacon that was not available on commlab. It was fairly simple to convert the syntax of the PSS code from ANSI to "old style" C, but when we ran the simulator, the probabilities produced were always zero! After some investigation, we concluded that the reason for this was in the way the compilers handled random number generation. To a degree, this was verified when we retrieved and installed GNU C (which conforms to the ANSI standard syntax) and ran the original programs. The probabilities produced were non-zero and agreed (for simulations at the low signal level, at least) with those documented by Dynacon. We are still investigating the discrepancies in the high signal level results.

One ongoing problem was the use of the constant "RAND_MAX" in the PSS simulator code. This constant, which represents the maximum integer to be generated by the random number facility, is defined as one less than the maximum integer by the ANSI standard (so, on commlab, this value would

be $2^{31}-2$, or 2,147,483,646). The constant was not part of the GNU C environment, so it was inserted into the source code in the file "sim.c". Since the low-level signal simulations seemed to be reproducible, we believe that this works, but if an ANSI C compiler is installed, this definition should be removed.

Once we had addressed the problems with the PSS simulator, we turned to the FVS simulator and attempted to see how to create the data files that the PSS simulator used as input. The documentation on this process included in the report was incomplete; so a number of "problem reports" had to be created for Dynacon to fill in the gaps. The procedure used for sending these reports to Dynacon over Internet is included as Appendix H. Eventually, these procedures were collected together (see Appendix C) so that we could attempt to recreate the simulations.

At this time, we believe we understand the procedures by which the simulations were run, but we have not yet been able to completely recreate and validate the results. We expect that this project will require more time and effort. Among the lessons that can be learned from this project is the value of good organization in documenting a project. It may be that the documents provided by Dynacon were not meant to be a "user's manual" for the simulation system but rather a demonstration of the algorithms' compliance with their requirements through unplanned (some would prefer "artistic") testing. There is no indication of a test plan included in the documentation available here, so it is not at all clear that such a plan exists. While there is little reason to doubt that the testing was done and that the algorithms do indeed meet their requirements, there is little evidence at this time that unequivocally demonstrates that they do so.

REFERENCES

[Dynacon92] SSF Space/Ground Antenna Control Algorithms: Final Design Report, Volume II: Validation and Verification. Part A -- Simulator Description and Validation, Algorithm Verification Overview. Dynacon Enterprises Limited, 1992.

[ISI92] SystemBuild/WS: User's Guide for Version 3.0, Part 000-0051-002. Integrated Systems, Inc., 1992.

92-32
11975

N95-32430

The Composite Classification Problem in Optical Information Processing

Final Report
NASA/ASEE Summer Faculty Fellowship Program—1994
Johnson Space Center

Prepared By:	Eric B. Hall, Ph.D.
Academic Rank:	Assistant Professor
College and Department:	Southern Methodist University Dept. of Electrical Engineering Dallas, Texas 75275
NASA/JSC	
Directorate:	Engineering
Division:	Tracking and Communications
Branch:	Tracking Techniques
JSC Colleague:	Richard D. Juday
Date Submitted:	August 5, 1994
Contract Number:	NGT-44-005-803

ABSTRACT

Optical pattern recognition allows objects to be recognized from their images and permits their positional parameters to be estimated accurately in real time. The guiding principle behind optical pattern recognition is that a lens focusing a beam of coherent light modulated with an image produces the two-dimensional Fourier transform of that image. When the resulting output is further transformed by the matched filter corresponding to the original image, one obtains the autocorrelation function of the original image, which has a peak at the origin. Such a device is called an optical correlator and may be used to recognize and locate the image for which it is designed. (From a practical perspective, an approximation to the matched filter must be used since the spatial light modulator (SLM) on which the filter is implemented usually does not allow one to *independently* control both the magnitude and phase of the filter.) Generally, one is not just concerned with recognizing a single image, but is instead interested in recognizing a variety of rotated and scaled views of a particular image. In order to recognize these different views using an optical correlator, one may select a subset of these views (whose elements are called training images) and then use a composite filter that is designed to produce a correlation peak for each training image. Presumably, these peaks should be sharp and easily distinguishable from the surrounding correlation plane values. In this report we consider two areas of research regarding composite optical correlators. First, we consider the question of how best to choose the training images that are used to design the composite filter. With regard to quantity, the number of training images should be large enough to adequately represent all possible views of the targeted object yet small enough to ensure that the resolution of the filter is not exhausted. As for the images themselves, they should be distinct enough to avoid numerical difficulties yet similar enough to avoid gaps in which certain views of the target will be unrecognized. One method that we introduce to study this problem is called probing and involves the creation of artificial imagery. The second problem we consider involves the classification of the composite filter's correlation plane data. In particular, we would like to determine not only whether or not we are viewing a training image, but, in the former case, we would like to determine *which* training image is being viewed. This second problem is investigated using traditional M -ary hypothesis testing techniques.

INTRODUCTION & BACKGROUND

A Review of Quadratic Classifiers Consider a random vector X taking values in \mathbb{R}^d with a multivariate¹ Gaussian density function

$$f(x) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right) \quad (1)$$

where $\mu = E[X] = [\mu_i]_{i=1}^d$ and $\Sigma = E[(X - \mu)(X - \mu)^\top] = [\sigma_{ij}]_{i,j=1}^d$. Note that this distribution is completely characterized by $d + \frac{1}{2}d(d+1)$ parameters. We sometimes denote this distribution by $N(\mu, \Sigma)$ where the covariance matrix Σ is a symmetric, nonnegative definite matrix. In writing expression (1), we have assumed that Σ is in fact *positive* definite, and hence invertible. If, instead, Σ is singular then $\alpha^\top X = 0$ a.s. for some nonzero vector α from \mathbb{R}^d . In such a case, we say that the distribution of X is degenerate. If $E[|\alpha^\top X|^2] > 0$ for each nonzero $\alpha \in \mathbb{R}^d$ then the covariance matrix Σ is positive definite. The nonnegative square root of the value $(x - \mu)^\top \Sigma^{-1}(x - \mu)$ in the exponent of (1) is called the Mahalanobis distance from x to the mean μ . Note that a set of points with equal Mahalanobis distance to the mean forms a hyperellipsoid in \mathbb{R}^d .²

Our interest lies with the *a posteriori* density function

$$p(t_i|x) = \frac{p(x|t_i)P(t_i)}{p(x)}$$

where t_i denotes an image with an *a priori* probability of $P(t_i)$. After taking the natural log of the *a posteriori* density function and neglecting the terms that do not change with i ,³ we obtain a discriminant function of the form $g_i(x) = \ln p(x|t_i) + \ln P(t_i)$. Assume now that $p(x|t_i)$ is a multivariate Gaussian density function with mean vector μ_i and covariance matrix Σ_i . In this case, it follows that

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^\top \Sigma_i^{-1}(x - \mu_i) - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| + \ln P(t_i). \quad (2)$$

If $\Sigma = \sigma^2 I$, then equation (2) reduces to

$$g_i(x) = \frac{-\|x - \mu_i\|^2}{2\sigma^2} + \ln P(t_i) \quad (3)$$

¹The components of X are said to be jointly or mutually Gaussian. Whereas a sum of Gaussian random variables need not be Gaussian, and uncorrelated Gaussian random variables need not be independent, these useful properties do hold when the random variables are jointly Gaussian.

²For more information concerning the topics in this section, see [3] or [4].

³Throughout this discussion the g_i 's are equal up to additive, constant functions of i and/or multiplicative, positive, constant functions of i .

where $\|x - \mu_i\|^2 \equiv (x - \mu_i)^\top (x - \mu_i)$. Note that if the *a priori* probabilities are equal then this test assigns x to the category that has the nearest mean, where distance is determined using the previous norm. This type of classifier is called a *minimum distance classifier*. However, if we expand g_i a step further (retaining all but the $x^\top x$ term), equation (3) reduces to

$$g_i(x) = \alpha_i^\top x + \beta_i \quad (4)$$

where $\alpha_i = \mu_i/\sigma^2$ and $\beta_i = -\mu_i^\top \mu_i/2\sigma^2 + \ln P(t_i)$. A test of this form is called a *correlator detector* or *linear detector*. As indicated by equation (4), the decision boundaries induced by equation (3) are hyperplanes.

Next, consider the case in which Σ_i is equal to some constant matrix Σ for all i . In this case, equation (2) reduces to

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^\top \Sigma^{-1}(x - \mu_i) + \ln P(t_i).$$

If the *a priori* probabilities are equal then this test assigns x to the distribution whose mean has the smallest Mahalanobis distance to x . Expanding further, however, we again obtain a test function of the form

$$g_i(x) = \alpha_i^\top x + \beta_i$$

where this time $\alpha_i = \Sigma^{-1}\mu_i$ and $\beta_i = -\frac{1}{2}\mu_i^\top \Sigma^{-1}\mu_i + \ln P(t_i)$. Note that again we have obtained a linear classifier and that our decision boundaries are hyperplanes.

In the general case, equation (2) may be written as

$$g_i(x) = x^\top A_i x + \alpha_i^\top x + \beta_i$$

where $A_i = -\frac{1}{2}\Sigma_i^{-1}$, $\alpha_i = \Sigma_i^{-1}\mu_i$, and $\beta_i = -\frac{1}{2}\mu_i^\top \Sigma_i^{-1}\mu_i - \frac{1}{2}\ln |\Sigma_i| + \ln P(t_i)$. Note that this test function is quadratic and that the corresponding decision boundaries are hyperquadrics.

The Classification Problem & Bayesian Inference Consider N_0 classes of objects denoted by $\omega_0, \dots, \omega_{N_0}$. Assume that object class ω_k contains N_k training images denoted by $T_{1k}, T_{2k}, \dots, T_{N_k k}$. Let $N_T = \sum_{k=1}^{N_0} N_k$ denote the total number of training images. The standard classification problem seeks a partition of the “signal space” into N_T classification regions denoted by R_1, \dots, R_{N_T} . The *composite classification problem*, on the other hand, seeks a partition of the “signal space” into N_0 regions denoted by C_1, \dots, C_{N_0} ; that is, each region corresponds to a different object class where each object class can contain many different training images.

Let N_F denote the number of composite filters that are formed from the training images, and let N_M denote the number of “features” that are required by the chosen

correlation metric. Then, the dimension n of our signal space is given by $N_F N_M$. (In this context, a “signal” corresponds to the output of our composite filter. Our goal is to classify these outputs into their corresponding object classes. The goal of the standard classification problem would be to map each possible output to the training image that produced it. Note that, for the moment, we are assuming that only training images are available as inputs to our system.) Our goals, once we develop our test, will be to choose the training image groups and composite filters so that (1) the distributions of the test statistics can be estimated accurately with as few samples as possible, and, (2) the distributions of the test statistics do not significantly overlap. (The first goal has pragmatic motivations, and the second goal reflects the standard desire that our test be as close to singular as possible.)

Let X denote our signal; that is, let X be a random vector taking values in $\mathbb{R}^{N_M N_F}$ that represents the N_M outputs of each of the N_F filters. Let $p_{T_{ik}}(x)$ be a probability density function for X when the training image T_{ik} is used as the input. Let Π_{ik} denote the *a priori* probability that the input to our system is training image T_{ik} . We will use these *a priori* probabilities to develop a Bayesian test. One assumption at this point is, of course, that these probabilities both exist and are known. Further, we will assume that the sum of the Π_{ik} 's over all possible training images is unity. That is, as mentioned above, we will continue to assume that only training images are input into our system. Based upon these assumptions we have the following expression for a probability density function for the output of our system:

$$p_T(x) = \sum_{k=1}^{N_0} \sum_{i=1}^{N_k} \Pi_{ik} p_{T_{ik}}(x).$$

Note that this density is only exact if our inputs are exclusively training images. In a more general setting in which our inputs need not be training images we would hope that this density would be a close approximation of the true density of the output. (An interesting problem would be to investigate the behavior of this approximation as $N_T \rightarrow \infty$.)

The Standard Classifier For a standard classifier, we let $N_0 = N_T$; that is, our object classes are singleton sets each containing a single training image. In this case, we will denote T_{ik} by T_k and Π_{ik} by Π_k . According to the usual Bayes formula, we have

$$p(T_k|X = x) = \frac{p_{T_k}(x)\Pi_k}{p_T(x)}$$

where $p(T_k|X = x)$ denotes the conditional probability that T_k was input given that we observed x . Our Bayesian hypothesis test then is to assign x to T_k if and only if $p(T_k|X = x) \geq p(T_i|X = x)$ for all $i = 1, \dots, N_T$. (That is, we choose the training image corresponding to a maximum *a posteriori* distribution.)

The Composite Classifier For a composite classifier, N_0 is less than (usually *much* less than) N_T . Let $p_{\omega_k}(x)$ be a probability density function for X when the input is a training image from the object class ω_i . Let Λ_k be the *a priori* probability that the input will be from class ω_i . (Yet again, we are assuming that the input will always be a training image.) Note that

$$\Lambda_k = \sum_{j=1}^{N_k} \Pi_{jk}.$$

Also, note that

$$p_{\omega_k}(x) = \frac{1}{\Lambda_k} \sum_{j=1}^{N_k} p_{T_{jk}}(x) \Pi_{jk},$$

and further, using Bayes' formula, note that

$$p(\omega_k|X = x) = \frac{p_{\omega_k}(x)\Lambda_k}{p_T(x)} = \frac{1}{p_T(x)} \sum_{j=1}^{N_k} p_{T_{jk}}(x) \Pi_{jk} \quad (5)$$

where $p(\omega_k|X = x)$ is the conditional probability that the input is from class ω_k given that x is observed. Our test in this case assigns x to object class ω_k if and only if $p(\omega_k|X = x) \geq p(\omega_j|X = x)$ for all $j = 1, \dots, N_0$. That is, we assign x to class ω_k if and only if

$$\frac{p(\omega_k|X = x)}{p(\omega_j|X = x)} = \frac{\sum_{i=1}^{N_k} p_{T_{ik}}(x) \Pi_{ik}}{\sum_{i=1}^{N_j} p_{T_{ij}}(x) \Pi_{ij}} \geq 1$$

for all $j = 1, \dots, N_0$.

We will now assume that $p_{T_{jk}}(x)$ is an n -variate Gaussian density function with mean vector m_{jk} and covariance matrix C_{jk} where we recall that $n = N_M N_F$. Note that, in this case, $p_T(x)$ and $p_{\omega_k}(x)$ are *mixtures* of multivariate Gaussian densities, which, of course, can be far from Gaussian. Several problems present themselves at this point. First, the rather rueful distribution of X does not bode well for analytic solutions.⁴ Second, the parameters m_{jk} and C_{jk} are rarely known and hence often must be estimated. These estimates may then substituted into the test given above. Unfortunately, when this substitution is done, our test is generally no longer Bayesian, and hence, need no longer satisfy any desired optimality property such as minimum probability of error.⁵ Although the test statistic is difficult to work with, one possible

⁴To be precise, it is only our *approximation* of the distribution of X that is a mixture of Gaussian distributions unless we assume that our inputs will only be training images. Of course, there is no reason to expect that the true distribution of X is any less rueful than our approximation of that distribution.

⁵The likelihood ratio corresponding to such a test is sometimes said to be a *generalized likelihood ratio*, particularly when the unknown parameters are replaced with maximum likelihood estimates.

simplification involves coordinate transformations that simplify the calculation of the Mahalanobis distances in the exponents.

TRAINING IMAGES

In the previous section, we assumed that the input to our system was always a training image, that the output of our system given that the input was a training image was Gaussian, and that the output of our system given that the input was a training image from a particular object class was equal in distribution to a mixture of Gaussian distributions.

There are two types of non-training images. A *first category non-training image* is an unknown view of a known object. Ideally, this type of image will be close enough to an appropriate training image so that their distributions will have significant overlap. A *second category non-training image* is an image of an object that is not intended to be recognized. Ideally, this type of image should be far away from the training images so that its distribution will not have significant overlap with that of any training image.

Of course, as we approach the ideal situation, we are generally going to need an ever increasing supply of training images. While a larger number of training images would be helpful when the input is a non-training image, it would also increase the complexity of our system and it would decrease the performance when the input actually is a training image. What we need is: (1) a method of determining how many training images we need to meet some desired goal, and (2) a method of obtaining appropriate additional training images when such images are required. In the next section, we will consider both of these problems.

Probing In [2] the term *probing* is introduced and used to describe the creation of artificial images to improve and analyze pattern recognition algorithms. A deterministic pattern recognition algorithm is simply a function mapping the set of all possible images to some decision set. Ideally, one would choose this function by considering each possible image in turn and determining for each the appropriate decision that should be made if that image appears as the input to our system. Unfortunately, however, such a design procedure is generally not tractable due to the enormous number of possible input scenes. (For example, there are over $10^{39,000}$ different 128 by 128 pixel input images with 256 gray scales.) It is at this point that probing becomes useful since it allows one to intelligently sample this enormous image space in order to select images for which a pattern recognition scheme can be designed.

We have identified three different methods of probing that appear promising with regard to optical information processing. First, probing can be used to form a “map” of the image space. As an example, consider two images I_1 and I_2 from the image space consisting of all $N \times N$ pixel images composed of M gray scale levels. (Note

that such an image can be viewed as a point in the set $\{0, 1, \dots, M-1\}^{N^2}$.) For each value of $p \in [0, 1]$, let V_p be a random object taking values in $\{0, 1, \dots, M-1\}^{N^2}$ such that each pixel value in V_p is equal with probability p to the corresponding pixel value in I_2 and is equal with probability $1-p$ to the corresponding pixel value in I_1 .⁶ (Note that $V_0 = I_1$ and $V_1 = I_2$. Note also that if a particular pixel in I_1 and I_2 agree then they also agree with the corresponding pixel in V_p . Thus, the pixel variance in V_p is only positive for pixels at which I_1 and I_2 disagree.) As p increases from 0 to 1,⁷ our decision algorithm (when applied to V_p) will on average no longer recognize I_1 at some point p_1 and will begin instead to recognize I_2 at some point p_2 . These average values of p allow us to determine when images are “adjacent” and to recognize the possible existence of a “hole” between I_1 and I_2 . A map of this sort allows us to distinguish between two procedures that perform the same when the inputs are always training images. Further, we could possibly use a realization of V_p where $p_1 < p < p_2$ to fill such a hole in our set of training images. Of course, different realizations of V_p for some fixed choice of $p \in (0, 1)$ could be quite different. (A similar procedure is described in [7] where sections of training images are randomly selected and weighted to form what is called a synthetic reference object.)

Second, probing can be used to measure the robustness of a pattern recognition algorithm. For a training image T and a nonnegative value θ , let U_θ be an image (i.e. a random object taking values in $\{0, 1, \dots, M-1\}^{N^2}$) such that each pixel value in U_θ has a mean equal to the corresponding pixel value in T and has a variance equal to θ . Note that $U_0 = T$ a.s. The average value of θ at which T is no longer recognized provides an indication of the robustness of our algorithm with respect to that particular training image. Presumably this value should be large and should not vary widely among the different training images. Tests of this sort allow us to distinguish between two procedures that perform identically when the inputs are always training images.

Third, probing can be employed in which the pixel values in the synthesized image are *not* statistically independent. By introducing some sort of spatially localized dependence, we can analyze a situation in which a pixel in the synthesized image is more likely to be from a particular training image if its neighboring pixel values are also from that training image.

⁶In addition, one could further perturb the image by choosing for some given gray scale value M_0 , a realization of a random variable with a unimodal distribution on $\{0, 1, \dots, M-1\}$ centered at M_0 .

⁷Notice that from an information theoretic standpoint, the “uncertainty” is maximized when $p = 1/2$.

COMPOSITE CORRELATION

Synthetic Discriminant Functions The synthetic discriminant function approach uses a linear combination of training images to create a composite image that is cross correlated with the input to the system. The weights in the linear combination are chosen so that the cross correlation at the origin is the same for all training images from the same class. (The resulting filter is sometimes called an Equal Correlation Peak (ECP) Synthetic Discriminant Function (SDF).) For example, if we have N training images $s_1(x, y), \dots, s_N(x, y)$, then our composite image would be of the form

$$h(x, y) = \alpha_1 s_1(x, y) + \dots + \alpha_N s_N(x, y),$$

and the α_i 's would be chosen so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) s_i(x, y) dx dy = c_i$$

for $i = 1, \dots, N$ where the c_i 's are preselected constants. Modifications of the standard SDF approach exist that impose other constraints. For example, Minimum Variance SDF (MVSDF) minimizes the output noise variance and the Minimum Average Correlation Energy (MACE) filter attempts to produce sharp correlation peaks at the origin of the output.

Let $H(u, v)$ denote the Fourier transform of the composite image $h(x, y)$; that is, let

$$H(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) e^{-j2\pi(xu+yv)} dx dy.$$

Note that

$$h(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(u, v) e^{j2\pi(xu+yv)} du dv.$$

Also, let S_i denote the Fourier transform of the i th training image s_i for $i = 1, \dots, N$. Further, for an element x from \mathbb{R}^N or \mathbb{C}^N let \tilde{x} denote the corresponding element in $\mathbb{R}^{N \times N}$ or $\mathbb{C}^{N \times N}$ that is obtained by placing x along the diagonal and 0 elsewhere. That is, let

$$\tilde{x}_{ij} = \begin{cases} x_i & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Finally, assume that the input to our system is corrupted with an additive, zero mean, wide sense stationary noise process $N(x, y)$ with power spectral density $P_N(u, v)$. That is,

$$P_N(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{E}[N(x + \tau, y + \lambda) N^*(x, y)] e^{-j2\pi(\tau u + \lambda v)} d\tau d\lambda$$

where the asterisk denotes the complex conjugate operation. We will now list several popular performance criteria.

1. The Output Noise Variance

$$\text{ONV} = H^*(u, v)^T P_N(u, v) H(u, v)$$

2. The Average Similarity Measure

$$\text{ASM} = H^*(u, v)^T \left[\frac{1}{N} \sum_{i=1}^N (\tilde{S}_i(u, v) - M_S(u, v))^* (\tilde{S}_i(u, v) - M_S(u, v)) \right] H(u, v)$$

$$\text{where } M_S(u, v) = \frac{1}{N} \sum_{i=1}^N \tilde{S}_i(u, v)$$

3. The Average Correlation Energy

$$\text{ACE} = H^*(u, v)^T \left[\frac{1}{N} \sum_{i=1}^N \tilde{S}_i^*(u, v) \tilde{S}_i(u, v) \right] H(u, v)$$

4. The Average Correlation Height

$$\text{ACH} = \left| \frac{1}{N} \sum_{i=1}^N H(u, v)^T S_i(u, v) \right|$$

Topics for Further Research If the previous performance criteria were our only concern then our goal in choosing h would be to minimize ONV, ACE, and ASM, and to maximize ACH. Some immediate questions that arise are:

1. Is it possible to optimize any or all of these parameters simultaneously? (In general, the answer is no for the performance criteria listed above. However, this question and those that follow should be considered whenever additional performance criteria are included.)
2. If not, then can we fix one or more of the parameters and then optimize those that remain? Note that a similar procedure is used to obtain a Neyman–Pearson test. Since it is not generally possible to maximize the power and minimize the size of a test, a Neyman–Pearson test maximizes the power while keeping the size constant. (The proof of this result follows from standard techniques of variational calculus.)

3. There are numerous paradoxes that arise in the Neyman–Pearson theory. Do similar difficulties arise here? For example, in a Neyman–Pearson test, strange results can occur if the false alarm rate is chosen to be large in a test that is almost singular. Conversely, situations exist in which the power exceeds the false alarm rate by an arbitrarily small value.
4. An *optimal trade-off filter* is obtained by fixing all but one of the parameters and optimizing the other. Is this necessary? Might it be possible to fix fewer parameters and then optimize those that remain?

Minimum Euclidean Distance Optimal Filters Our goal in this section is to extend the results in [5] to include composite filters. In particular, we will first seek an algorithm by which the output of the MEDOF algorithm can be classified by statistical inference into training image classes. The following steps follow the procedure suggested in [1]:

1. Separate the training images into object classes. The first step is to select the training images and object classes. Each object class should correspond to a different object that we wish to recognize. The training images within each object class should be chosen to adequately describe the different expected orientations of the object they represent.

2. Create the filters by which these training images will be distinguished. One way in which this step could be achieved would be to create a composite image from each object class based upon the training images in that object class. These composite images could then be used to create composite filters whose combined outputs would comprise the components of our output random vector X . (That is, we would let $N_F = N_0$.)

3. Estimate the mean and covariance matrix for each class. We have assumed that the output of our system given that the input is a specific training image T_{jk} will be multivariate Gaussian with mean m_{jk} and covariance matrix C_{jk} . We will estimate these parameters via standard techniques. In particular, if we observe x_1, \dots, x_N when training image T_{jk} is our input then we will estimate m_{jk} via

$$\widehat{m}_{jk} = \frac{1}{N} \sum_{i=1}^N x_i,$$

and we will estimate C_{jk} via

$$\widehat{C}_{jk} = \frac{1}{N-1} \sum_{i=1}^N (x_i x_i^\top - \widehat{m}_{jk} \widehat{m}_{jk}^\top).$$

4. **Calculate the generalized weighted Gaussian sum for each class.** The Bayes test consists in selecting the object class for which equation (5) is largest. In our case, however, we must substitute the estimates we just obtained in place of the true parameters that are unknown. Thus, for each object class ω_k from the set of N_0 different object classes, we calculate the ratio

$$\frac{\sum_{j=1}^{N_k} \Pi_{jk} |\hat{C}_{jk}|^{-1/2} \exp\left(-\frac{1}{2}(x - \hat{m}_{jk})^\top \hat{C}_{jk}^{-1}(x - \hat{m}_{jk})\right)}{\sum_{i=1}^{N_0} \sum_{l=1}^{N_k} \Pi_{il} |\hat{C}_{il}|^{-1/2} \exp\left(-\frac{1}{2}(x - \hat{m}_{il})^\top \hat{C}_{il}^{-1}(x - \hat{m}_{il})\right)}. \quad (6)$$

5. **Choose the class with the largest weighted sum.** We select the object class for which the corresponding term found in expression (6) is the largest. Our test then announces that the input to our system belongs to this object class.

Caveats

1. The calculation of expression (6) is very computationally intensive. This problem can be lessened somewhat by appropriate coordinate transformations.
2. The insertion of the estimates in place of the true parameters may generally be expected to remove any optimality condition that the original test was designed to satisfy.
3. The procedure is based upon an initial assumption of normality that may or may not be justified. In particular, an important concern is the robustness of our test with regard to perturbations in the underlying distributions. Also, what modifications would be required in order to remove the assumption of normality?
4. The procedure requires one to possess the *a priori* probability associated with each training image. These probabilities are generally not known and hence must be either estimated or assumed. Again, the robustness of our procedure with respect to these values is an important concern.

CONCLUSION

We have presented an overview of our research in two areas of optical pattern recognition. First, we have considered the use of probing to map the image space and to measure the robustness of an optical correlator with respect to deviations in the input from training images. Second, we have considered the use of Bayesian inference in the design of composite correlators in which images are assigned to object classes consisting of a collection of training images.

REFERENCES

- [1] Bret F. Draayer, Gary W. Carhart, and Michael K. Giles, "Optimum classification of correlation-plane data by Bayesian decision theory," *Applied Optics*, Vol. 33, No. 14, May 10, 1994, pp. 3034-3049.
- [2] Kenneth Augustyn, "A new approach to automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 1, January 1992, pp. 105-114.
- [3] Richard O. Duda and Peter E. Hart, *Pattern Classification and Scene Analysis*, Wiley-Interscience, New York, 1973.
- [4] Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition*, Second edition, Academic Press, Boston, 1990.
- [5] Richard D. Juday, "Optimal realizable filters and the minimum Euclidean distance principle," *Applied Optics*, Vol. 32, No. 26, September 10, 1993, pp. 5100-5111.
- [6] Bahram Javidi and Joseph L. Horner, eds., *Real-Time Optical Information Processing*, Academic Press, Boston, 1994.
- [7] Jerome Knopp, "Spatial weighting of synthetic reference objects for joint correlation by random pixel mixing," *Proceedings of SPIE*, Orlando, Florida, April 16-18, 1990, Vol. 1295, Real-Time Image Processing II, pp. 138-145.

53-31

4976

**THE BACKGROUND AND THEORY
OF INTEGRATED RISK MANAGEMENT**

**Final Report
NASA/ASEE Summer Faculty Fellowship Program - 1994
Johnson Space Center**

Prepared by: John L. Hunsucker, Ph.D., P.E.

Academic Rank: Associate Professor

**College and Department: University of Houston
Department of
Industrial Engineering**

NASA/JSC

Program: Space Station

Office: Integrated Risk Management

JSC Colleague: James E. Van Laak

Date Submitted: August 5, 1994

Contract Number: NGT-44-005-803

ABSTRACT

While all good managers have always considered risk in their decision making, only recently have formal programs to do so been introduced. This report covers the logical structure behind the formulation of an integrated risk management plan (IRM). Included in the report are factors forcing the development of a formal plan to consider risk, the basic objective or purpose of an IRM, and desirable traits of such a plan. The report moves on to a discussion of background issues, seeks to formalize some definitions, and then discusses required information on threats. The report concludes with the steps for an IRM.

INTRODUCTION:

Any program should have a strong foundation that explains why the program is needed, what drives the program, what characteristics the program should possess, how the program will be used and by whom, and what the program hopes to accomplish. Said another way, every program needs a strong theoretical background that relates to not only its design but also to its usage. Action or activity demands reason. The intention of this paper is to present the beginning formulation of the background and theory of an integrated risk management program.

BACKGROUND AND PURPOSE

An integrated risk management plan, IRM, is related to helping a particular office or program incorporate risk into its decision making. The plan is tied to the entity that it supports.

Following a logical formation of ideas requires a statement of the purpose of an IRM. For reasons expressed later, a success statement is also necessary.

BASIC OBJECTIVE OF AN IRM PLAN

TO FACILITATE GOOD DECISIONS WHICH
INCORPORATE THE CONSIDERATION OF RISK
AND
MOVE TO THE FOREFRONT OF MANAGERIAL CONSIDERATION THOSE
ITEMS WHICH ARE REAL AND MEANINGFUL THREATS TO THE STATED
PURPOSE OF THE OFFICE OR PROGRAM

SUCCESS STATEMENT

THIS MODEL WILL BE SUCCESSFUL IF
IT ASSISTS IN THE DECISION PROCESS
THREATS ARE MOVED TO THE FOREFRONT
IT GETS INCORPORATED WIDELY THROUGHOUT THE SYSTEM WHICH IT
SUPPORTS

In essence, the objective statement expresses what the plan purports to do and the success statement helps to identify if the plan has successfully accomplished its objective. Note that while the two statements are similar and related, they are also somewhat different.

DRIVERS FOR IRM

All good managers have always considered risk. A reasonable question to ask then is what is new with IRM? A common question within the space station is what does IRM bring to the table? If it has always been done and good managers have always done it, why is it needed?

While it is true that risk has always been considered, what has been missing is a formal, structured program to consider risk. There are at least three drivers that mandate structure on the consideration of risk - complexity, consequence, and credibility.

Complexity

The complexity of technological decision making has increased significantly with time. Technological systems have more component parts and there are a larger number of choices for each of these parts. This increase in numbers, of course, increases the interaction between component systems. Said another way, society desires to do more than it has ever done and has more to do it with. As a specific example, the Space Station is arguably the most complex project ever undertaken by mankind.

Consequence

The consequences of technical decisions can have a much greater negative impact than ever before. One only needs to consider Chernobyl, Bhopal, and other well publicized incidents to realize that the wrong decision can wipe out a program, a product, a company, a culture, or even a civilization.

The dollar value of decisions is only one of many parameters which can be negatively impacted. The environment, culture, government, and even heredity are all candidates for damage from a bad technical decision. Never before in history has man had the capability to do so much damage and to do it so quickly.

Credibility

Given that complexity and consequence have increased, a conscientious manager is going to demand that subordinates present convincing arguments that risk has been considered. Otherwise, when this manager presents the case to upper management, credibility will be lost. As an aside, most managers would favor the consideration of a risk and perhaps the taking of a risk over the posture of never having considered the risk, particularly if the risk materializes into a negative incident.

Perhaps in the age of the slide rule, risk need not be considered formally. In the age of the computer, with rapid communications and advanced technologies, a formal program to consider risk is mandatory.

DESIRABLE TRAITS OF AN IRM PLAN

There are certain traits that are desirable in an IRM plan. These characteristics will help to identify component parts of a proposed plan.

Reproducible

Work effort that is produced by an IRM should be reproducible in the sense that given the same input and assumptions at a later date, then it should be possible to arrive at basically the same conclusions concerning risk. The logic should not decay with time.

As the program matures it is foreseeable that some early decision will be reconsidered at a much later date. At this point, the conclusions concerning risk and the logic and structure that led to these conclusions should be reproducible.

Transportable

Two different analysts, sitting in different locations and given the same data and assumptions, should arrive at somewhat the same conclusions regarding risk. As an example with the space station, analysts at JSC, KSC, MSFC, and Headquarters should have somewhat the same results given the same basic information.

Simple

Given that one of the criterion's for success is that the system be widely accepted, it must be as simple as possible. As an aside, many engineering managers have little or no training in statistics. As of this writing for example, two of five engineering undergraduate programs at the University of Houston do not require statistics. To this end, a complicated statistics package will have difficulty gaining wide spread acceptance. Simpler is better and the leaner the plan, the more likely is the acceptance. The longevity of the plan is also more than likely tied to its simplicity.

Based on the normal method of doing business

The supposition here is that good managers have always considered risk management. The IRM should just fit in naturally with what the managers are used to doing. For the most part this means that the IRM should follow the same breakdown that is used to structure the work. If a work break down structure is used, for example, then the IRM should follow the same WBS structure.

Managers will have enough to learn and do without having a system imposed on them that is totally different than what they normally do. The closer this system fits with the normal work of the managers then the more likely that it will be adopted and used.

Helps to remove or identify bias

Managers use intuition. Intuition is gained by experience. Experience introduces bias. Some people are risk takers and some are risk avoiders. All are influenced by past experience. Change the experience base and the underlying factors which influence decisions may become obscured. A

desirable trait of an IRM would be to eliminate bias in the decision process. This would seem to be virtually impossible. Thus the goal becomes one of identifying and understanding the bias.

Should be capable of evolution

With time, the concept of the work will change. The IRM must change with the work. As a prime example, early on, design and development are prime considerations. Hopefully, the plan will evolve with the work into the manufacturing or operations stage.

Verifiable

One of the real worries with risk management is whether all of the important issues have been considered. Has something been overlooked that shouldn't have been? A verification step helps to provide credibility and to increase confidence that nothing major has been missed.

BACKGROUND ISSUES

The following is a brief discussion of some of the background issues related to risk management. The intent here is to provide additional insight to the plan that is developed later.

Two different uses of RM

Risk management can be used either in alternative selection, a static use, or operations/development control, a dynamic use. In the static use, RM can be used to choose between alternatives. Risk, then, becomes another variable to consider in the selection among choices. In the dynamic version of RM, risk is used to predict how well a particular plan is going to work.

Much of the literature on risk is related to the dynamic use of RM. One of the characteristics of the dynamic state is the existence of data. Having data allows for the usage of such tools as control limits. The scarcity of data and data streams makes the consideration of risk in the static usage more complex.

Risk is temporal

The objective of a program, and the related definition of success, change with time. As the definition of success changes, so must the concept of risk. As a specific example, consider the space shuttle. For its first launch, success was more than likely defined as getting it up, getting it down, and not injuring anyone or anything. After fifty launches, one would suppose that the objective would be significantly more robust.

The concept of risk being tied to time is, of course, related to the trait concerning evolution discussed above. The RM program must have the capability of changing as the program changes.

Risk is hierarchical

It would be naive to suppose that the person at the top of a very complex program would define success in the same manner as the person in charge of some small sub-element many levels down in the structure. While these concepts of success are related, they are different. Thus the risk is considered differently and, to some extent, must be managed differently.

The risk manager is like a lifeguard

The question comes up, unfortunately too frequently, as to what does the risk manager bring to the table. After all, the functional manager has some capability or they would not be the manager. What do they need the risk manager for?

As a philosophical orientation to the subject, consider a lifeguard at the local swimming pool. They are not there to stop you from swimming in deep water. They are there to warn you when the water is getting deep and to assist you if you get in over your head and cannot cope. To some degree, the job of the risk manager is similar. They are not there to make the decision for the functional manager. They are there to warn of trouble and to assist in the event that trouble does indeed occur.

Exposure

This program, like any other innovative program, requires real, substantial involvement of upper management. If the upper management exposure is not real and tangible, then middle management will kill it off. When middle management sees that upper management has committed to the program in a real and tangible way, then they too will commit and the program has a chance.

DEFINITIONS

One of the difficulties in discussing risk is that the words, to a large degree, are known and understood by everyone. Unfortunately they are usually understood differently by different people. One of the first things to do with any risk management plan is to well define the terms and then to discipline everyone to use the terms as defined and only as defined. This, more than anything else, will help to abate the hours of discussion that the consideration of an IRM will promulgate.

The following list defines many of the terms used in the rest of this paper. While the list is not complete, it does contain the most common terms.

Threat - Any real or perceived state or condition which would or could have a negative impact on the stated purpose or success of an entity. Note that a perceived condition is important. If people feel that something is true, then their actions are influenced, whether the thing is true or not. Also note that threat is tied to purpose, objective, and success.

Risk - Throughout this paper, risk is used in the non-technical sense. To be more precise, to some authorities, risk is the analytical product of likelihood and consequence, the mathematical expectation. In this paper, it will be used only in the vague sense.

Prodromal event - An event which is a precursor of a condition, a warning.

Unfortunately, prodromes are those things most often seen in retrospect, looking backwards, after a calamity. Storm clouds are a prodrome for rain and wind. An open barn door is a prodrome for a stolen horse.

Parameter - Something which is measured with the intention of reflecting or showing a condition or state.

In the sense that it is used here, part of the task is to determine what parameters to measure so that they, to some degree, reflect the risk that is involved.

THREATS

Must be tied to a purpose or objective

When one discusses threats, the first question is and must be - a threat to what? If one does not know what is threatened, the concept becomes fuzzy and the possibility of successful abatement becomes less likely. Too often the discussion of threat or risk is reduced to a sort of "boogie man in the closet" issue. For a threat to be real, what is threatened must be known.

Threats can be self, up, down, or across threats

Consider the tree diagram of a typical WBS chart. An office in that diagram has offices above it and below it. It also has offices which are not in the same string of offices and which can only be reached by going up, across, and then down another string.

Self threats are those against the stated purpose or objective of the given office. These are threats which that office should know and understand best. To some degree, the office will also be familiar with up and down threats. It is the across threats which may be the most obscure. It is reasonable to suppose that the given office will have less knowledge about environments far from them. It is with threats of this sort that the IRM office may be most useful.

Required information on any threat

Other than the basic assumptions and determining what the threat is to, there is some desired information on any threat. Almost any manager, when presented with a statement of threat, will want to know at least four things: confidence, warning, likelihood, and consequence.

Confidence - What confidence does the analyst have in the figures or statements presented? Is the threat information based on solid firm information or is it based on a vague feeling? The confidence that the analyst has in the figures will, to some degree, be reflected in how seriously management considers the threat.

Prodromal events - Will the threat materialize with little or no warning or will there be time to react as events unfold? Threats which materialize with little warning represent, to some degree, a less palatable threat than one which gives you plenty of time to get ready.

Consequence/impact - If the threat materializes, what will be the end result or range of results? How bad can it get?

Likelihood - Here the consideration is one of probability. How probable is the threat to materialize?

As a special consideration, the determination of likelihood and of consequence is very difficult. It is at this point that the discussion of proposed risk management plans seems to become impaled on hours and hours of circular

reasoning. For this reason, and others, in what follows, we will use a five point scale to represent each of confidence, likelihood or consequence: very low, low, moderate, high, and very high. Obviously, these words will have different interpretations in different situations. The intent is to use the term and then to let discussion add meaning to the term.

USAGE

The question now becomes one of usage. Who will use the IRM and for what will it be used? Knowing the intended usage of a system is of paramount importance in designing the system.

The plan which is described in this paper will, of course, be used by everyone. However, the end user will be upper management. They are the ones charged with the most responsibility, thus they are the ones who have the end authority on whether risk has been considered and managed. To some degree, an IRM can be considered as a security blanket for upper management.

At this point, a word about numerical representations of risk is appropriate. The danger with any parameterized system is that too much will be read into a number that is generated. The only thing that can add definition and understanding to a generated number is usage and time.

As an example, suppose the risk of a particular entity was reported to be 0.7 on a 0 to 1 scale with 0 being low and 1 being high. What does 0.7 mean? How does it compare to someone else's 0.6 or 0.8? Only use and time can add the required definition. The most important use of the number is two-fold. One is to track the number over time. In our example, if the risk went to 0.8 in the next reporting period, most managers would understand that the risk posture was getting worse. Conversely, if the risk went to 0.6, most would understand that the risk was getting better. Tracking over time and looking at the trending of the number provides insight into the risk posture. The other use of a number is to give management the opportunity to explain the number. If the risk is reported as 0.7, the obvious question is why is it 0.7 and not 0.6 or 0.8. What is the explanation for the number? Again, time and use will add definition and reason to any parameter that is produced.

THE ENVIRONMENT

In order to understand the proposed plan, a brief discussion of the author's understanding of the environment is essential. No plan can be lifted directly from one environment and placed into another without modification. The literature is replete with examples of industries and companies who have tried this and failed.

The way the work at the space station seems to be structured is that the tasks are distributed among integrated product teams (IPT's). There are somewhere around 100 such teams, depending on how they are counted. If one thinks of a typical tree structure as in a WBS the IPT's are somewhat

similar. Every so often in the structure, they have introduced an Analysis and Integration team charged with the responsibility of integrating the levels below them. The intention of this structure is to empower the teams by forcing the decision making to as low a level as possible.

THE PLAN

The basic plan is to have the IPT's report their risk posture to upper level management on a routine and regular basis. If this is done often enough, then trending will develop and information will be gained. Due to the large number of teams, some subset will have to report each time. A typical plan would have a regular time slot each week for reporting, say two hours. Then the teams would be grouped in some natural ordering following the AIT structure and each would expect to report during their time slot when their particular grouping was presenting. Their turn would come up again every four or six weeks or whatever time was appropriate. Note that this method immediately franchises the risk management program. Teams will naturally turn to the office of IRM to get help.

The upper level manager reviewing the findings will add definition to any of the parameters generated simply by asking why the team assigned a certain number to a parameter. This system might well be chaotic at the start but time will add definition, stability and insight.

STEPS FOR THE REPORT

There are five basic steps that each team will have to take in order to generate the required information: background, threats, metric, reconciliation and transition. Each of these steps will provide a integral piece of their report.

Step 1: Background

In the background step there are three separate pieces of required information:

- Determine the purpose/objective of the IPT
- Write out the success statement of the IPT
- List the ground rules and assumptions of the IPT

The purpose of the background information is to insure that everyone has the same basic understanding of the basic function of the IPT, i.e., what they are supposed to do, how they will know if they have done it, and the ground rules covering the doing. Note that all three of these pieces of information will change with time.

Step 2: Threats - Top Ten Plus One

Each IPT needs to develop a top ten plus one threat list. The first part of this report contains a list of about ten threats. The term about ten is used to allow some flexibility in the number of threats listed. The intent is to not discourage a fairly significant eleventh or twelfth threat or to encourage the manufacture of a tenth threat.

The following information should be contained with each threat:

- What it is a threat to
- Whether the threat is up, down, self, or across
- The confidence in the analysis
- The consequences
- The likelihood
- Prodromal events

Each of the above items should be expressed in a paragraph form. The confidence, likelihood, and consequence should include in their paragraph a rating on the five point scale mentioned above. A brief justification for the rating should also be included.

The second part of this report should contain a dark horse. The dark horse is presented without any justification. It is the single thing in the managers mind that is worrying them but is still vague. The intent here is to encourage them to present unsubstantiated feeling of risk to management.

This report should be formulated by the IPT and go to the AIT and to the Office of IRM. One of the functions of the IRM would be to insure that the information is circulated to the concerned offices. The report should be presented, as previously discussed, to upper management on a regular basis.

Step 3: The Metric

The intent here is to present a metric which reflects the risk to the objectives and success statements. The metric is developed by each IPT and consists of three parts. The IPT evaluates the risk to themselves. Note that this is not a trade study but an evaluation of whether the IPT feels that they can do what they are supposed to do.

Three parameters are purposed, cost, schedule, and performance. The method is to fix two of the three and then to evaluate a confidence level in being able to accomplish the third, given the first two. The confidence is to be evaluated between 0 and 1 in 0.1 increments.

- Performance

Given your current schedule and budget the confidence that you will meet your technical performance goals is _____.

- Cost

Given your current schedule and technical performance goal, the confidence that you will meet your cost goals is _____.

- Schedule

Given your current budget and technical performance goal, the confidence that you will meet your schedule is _____.

There are any number of parameters that could have been proposed. However, in the interest of developing a simple program with a chance of acceptance, these three were considered the most important.

Again, use will provide definition to these parameters. Their most significant use will come with the tracking of the parameters over time. Trending will provide upper management with a large amount of information on risk and on abatement.

The tendency to combine the three metrics into one will only obscure information and should be avoided. The recommendation that each of the three be tracked separately is yet another reason to restricting ourselves to three metrics.

It is fully expected that each IPT will avail themselves of sophisticated methods to establish the value of the parameters. It may well even be true that each IPT uses a different method. However, over time and with experience, it is reasonable to suppose that some methods will prevail and become common. Here, the intent is to let use determine the method as opposed to method determining the use.

Step 4: Reconciliation

The top ten plus one list should be compared to the three metrics generated to see if the information plays well together. Logic and consistency are important. In essence, the IPT has verified their feelings on risk by working the problem two different ways, one in paragraph form and the other with the metrics. This part of the report should be in paragraph form and discuss how the two different parts support each other.

Step 5: Transition

Risk management is dynamic. The RM plan requires a periodic review to insure it is meeting its required objectives and to insure that the objectives themselves have not changed. As the program becomes more operational in nature, significant changes in the definition of success will occur forcing changes in the IRM plan. Step 5 in the report is a paragraph discussing how the office plans to deal with the change from a risk perspective.

CONCLUSIONS

As was stated earlier in the paper, this is a tentative beginning of the development of the foundations of risk management. Surely significant changes and modifications will occur with time. The question now becomes one of whether the implementation of such a program is worth the overhead that it brings. Can a program afford to implement an IRM plan?

One of the problems with technical decision making is that non-technical people understand neither the principles involved nor the process. Given this, they have to trust the technocrat to do what is right and correct. One thing is resoundingly clear, if the community as a whole loses their faith in the technocrats to make the "good" decision, they can shut the whole process down. If Congress loses its faith in NASA, they can stop the space station. If the public loses its faith in a company, they can force financial ruin. If the auditors and accountants lose their faith in the design team they can stop a project.

Given this and given the high risk of many technical decisions, the question more properly becomes one of can a company or program afford not to institute a formal risk management plan?

514-16
1/7/94

N95- 32432

**IMPLEMENTATION OF A SINGLE-STAGE-TO-ORBIT (SSTO)
MODEL FOR STABILITY AND CONTROL ANALYSIS**

**Final Report
NASA/ASEE Summer Faculty Fellowship Program--1994
Johnson Space Center**

Prepared By: Stephen A. Ingalls, MSAE

Academic Rank: Assistant Professor

College and Department: United States Military Academy
Department of Civil and
Mechanical Engineering
West Point, New York 10996

NASA/JSC

Directorate: Engineering

Division: Navigation, Control, and Aeronautics

Branch: Control and Guidance

JSC Colleague: G. Gene McSwain

Date Submitted: August 5, 1994

Contract Number: NGT-44-005-803

S15-32

4978

N95-32433

**SCANNERS, OPTICAL CHARACTER READERS, CYRILLIC
ALPHABET AND RUSSIAN TRANSLATIONS**

Final Report
NASA/ASEE Summer Faculty Fellowship Program 1994
Johnson Space Center

Prepared By:	Gordon G. Johnson, Ph.D.
Academic Rank:	Professor
College and Department:	University of Houston Department of Mathematics Houston, Texas 77204
NASA/JSC	
Directorate:	Information Systems
Division:	Technical Systems
Branch:	Client Server Systems
JSC Colleague:	Lui Wang
Date Submitted:	August 16, 1994
Contract Number:	NGT-44-005-803

REFERENCES

1. *Access to Space*. NASA Advanced Technology Team (Option 3), Volume 3. Final Report. 1993.
2. Englund, W. "SSV-001, 6-DOF Aero Database." Preliminary Aerodynamic Database. March 1994.
3. Hoak, D. E. et al. "USAF Stability and Control DATCOM." Data Compendium, 1978.
4. Lan, C. E. and J. Roskam. *Airplane Aerodynamics and Performance*. Roskam Aviation and Engineering Corporation, Ottawa, Kansas, 1988.
5. Lepsch, R. "Weight Statement - Level III, Unmanned SSV, Duel-Fuel, RD-701." Data Table, April 1994.
6. Rockwell International. "Operational Aerodynamic Data Book, Orbiter Vehicle Aerodynamic Data, Volume 3." Rockwell International, 1985.
7. Rockwell International. "Space Shuttle Ascent GN&C Data Book Appendix." Contract NAS9-18500. September 1992.
8. Roskam, J. *Airplane Flight Dynamics and Automatic Flight Controls*. Roskam Aviation and Engineering Corporation, Ottawa, Kansas, 1982.
9. Springer, A. "Initial Ascent and Entry Aerodynamic Database for the Wing/Body 001 (WB001) Reusable Launch Vehicle (RLV) Configuration." Memorandum, July 1994.

CONCLUSIONS

This study concludes the WB001's current level of lateral stability is insufficient. In particular, lateral trim solutions for desired cross-wind components in the landing environment were not possible. Instabilities in roll and pitch 'tuck' also exist, but appear less significant than deficiencies in the $C_{n\beta}$ derivative. Longitudinal trim solutions indicate steady-state glide path angles on the order of 12 degrees at Mach = 0.30, increasing to slightly greater than 20 degrees at $M = 0.60$. One attempt at improving the vehicle's lateral instabilities looked at sizing lateral control surfaces to a prescribed degree of inherent stability. A value comparable to subsonic stability of the orbiter ($C_{n\beta} = 0.100/\text{radian}$) was selected and required increasing tip-fin surface area 6.78 times (from around 215 to 1,457 ft²). This carries a commensurate increase in aircraft weight, highly undesirable in a SSTO configuration. Other methods of enhancing the vehicle's lateral stability are available and require further study.

(DATCOM). Comparison of LaRC tip-fin yaw moment to tip-fin side force at the same deflection resulted in a value for X_{tf} (moment arm of the tip-fin control surface to vehicle center-of-gravity) of 56.112 feet. Substituting a wing area of 4,211 ft² and span equal to 95.72 ft, equation 7 shows how increasing tip-fin surface area results in a linear increase in $C_{n\beta_tip-fin}$.

A vehicle $C_{n\beta}$ of 0.100/radian was selected and is representative of the orbiter's subsonic lateral stability in this derivative. WB001's current tip-fin surface area is 214.95 ft². Increasing area to obtain this degree of inherent stability requires a growth factor of 6.78 (total lateral control surface area required increases to 1,457 ft²).

This offers one method of improving vehicle lateral stability and cross-wind landing capability. Enhancing basic wing-body weathercock stability or employing alternative methods of improving tip-fin effectiveness without a size increase are equally viable options.

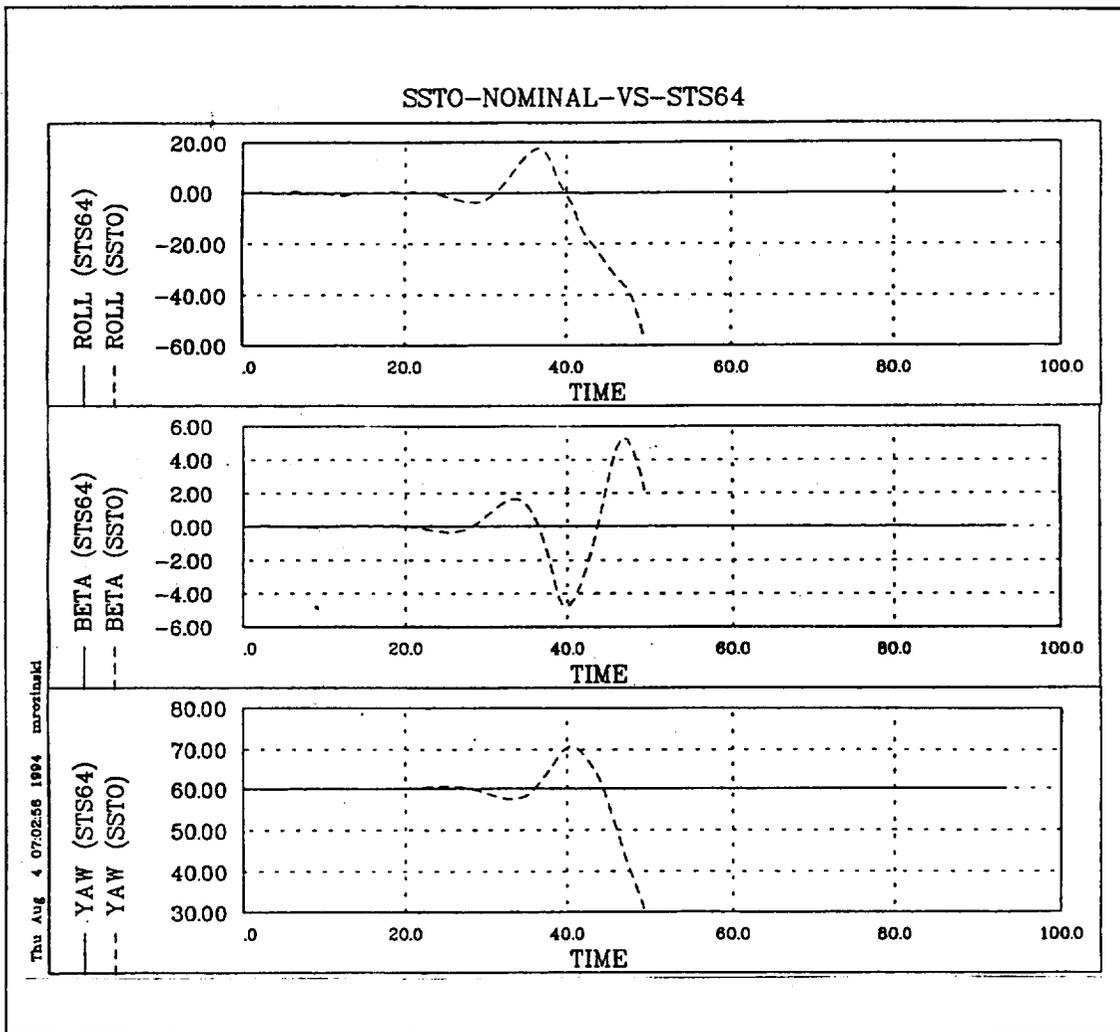


FIGURE 3.- SHUTTLE ENGINEERING SIMULATOR STRIP PLOTS COMPARING ROLL RATE [DEG/SEC], PITCH RATE [DEG/SEC] AND SIDESLIP ANGLE [DEG] VS. TIME [SECONDS] FOR THE SSTO TO ORBITER IN STS-64 CONFIGURATION

If we neglect sidewash dependence on β ($\partial\sigma/\partial\beta$) and assume the dynamic pressure ratio (η_v) is 1, equation 6 reduces to:

$$C_{n_{\beta_tip-fin}} = C_{L_{\alpha_tip-fin}} \frac{S_f}{S} \frac{X_f}{b} \quad (7)$$

$C_{L_{\alpha_tip-fin}}$ was calculated as 2.174/radian using methods outlined in the United States Air Force Stability and Control Data Compendium

Further, orbiter design vertical velocities at touchdown are approximately 2.0 feet/second. Using this and the assumed touchdown velocity of around 200 knots results in a flight path angle of approximately zero (0.34°) for substitution into the lateral equation set.

Attempts to obtain a trim solution to the lateral equations of motion proved impossible, indicating the vehicle incapable of landing in the design cross-wind. This, coupled with $C_{n\beta}$ static instabilities, suggested an analysis to improve the SSTO's inherent lateral stability.

Tip-Fin Sizing

Early landing simulation of the SSTO indicated lateral divergence from design trajectories in both roll and yaw with small β (Figure 3). Further attempts to land the vehicle indicated increasing lateral force and moment effects resulted in more favorable performance. Discussions with engineers at LaRC indicated a pending effort to reevaluate tip-fin size.

Sizing lateral control surfaces to produce a desired $C_{n\beta}$ derivative value is common practice in the preliminary design of aircraft. Considering vehicle weathercock stability ($C_{n\beta}$) as a superposition of effects from the wing-body and lateral control surfaces, in this case, the tip-fins, yields:

$$C_{n\beta} = C_{n\beta_wing-body} + C_{n\beta_tip-fin} \quad (5)$$

The tip-fin contribution is further expressed as:

$$C_{n\beta_tip-fin} = C_{L\alpha_tip-fin} \left(1 - \frac{\partial\sigma}{\partial\beta}\right) \eta_{tf} \frac{S_{tf}}{S} \frac{X_{tf}}{b} \quad (6)$$

where,	$C_{L\alpha_tip-fin}$	= Tip-Fin Lift-Curve Slope
	$\frac{\partial\sigma}{\partial\beta}$	= Change in Sidewash with Sideslip Angle
	η_{tf}	= Ratio of Dynamic Pressure at the Tip-Fin to the Wing
	S_{tf}	= Tip-Fin Surface Area [ft ²]
	S	= Wing Area [ft ²]
	X_{tf}	= Distance from Center-of-Gravity to Tip-Fin Aerodynamic Center [ft]
	b	= Wing Span [ft]

deflection trend smaller at higher Mach number, while flight path angle increases to values in excess of 20 degrees at 10,000 feet. These data bound possible trim solutions and this effort does not design a final descent trajectory.

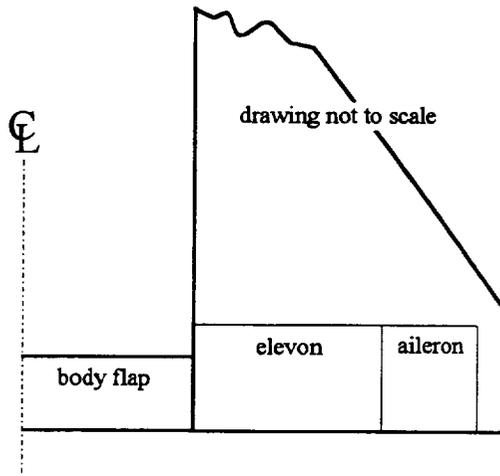


FIGURE 1.- WB001 CONTROL ORGANIZATION SCHEMATIC

Lateral-Directional Trim Results

Lateral-directional trim was attempted in a similar manner. Rather than look at conditions throughout the final descent, trim was attempted for a design cross-wind condition at landing of 25 knots. Assuming touchdown velocities comparable to shuttle (≈ 200 knots), this results in a design sideslip angle of approximately 7 degrees.

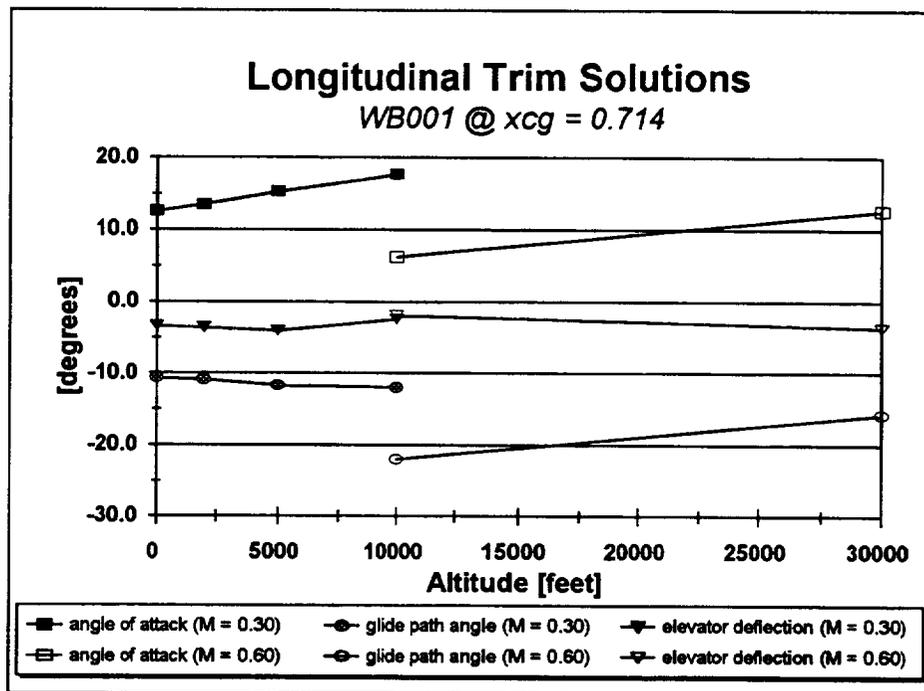


FIGURE 2.- LONGITUDINAL TRIM SOLUTIONS (α , γ , and δ_e) FOR TWO MACH NUMBERS VS. ALTITUDE

Lateral-Directional Equations of Motion

$$\begin{aligned} -mg \cos \gamma &= (C_{Y\beta} \beta + C_{Y\delta_{ail}} \delta_{ail} + C_{Y\delta_{tf}} \delta_{tf}) \bar{q} S \\ 0 &= C_{l\beta} \beta + C_{l\delta_{ail}} \delta_{ail} + C_{l\delta_{tf}} \delta_{tf} \\ 0 &= C_{n\beta} \beta + C_{n\delta_{ail}} \delta_{ail} + C_{n\delta_{tf}} \delta_{tf} \end{aligned} \tag{4}$$

In these equations, flight path angle (γ), sideslip angle (β), aileron deflection (δ_{ail}), and tip-fin deflection (δ_{tf}) are not known. Either the longitudinal flight path angle solution, or another flight path angle of interest, is substituted here, leaving three unknowns. It is important to differentiate tip-fin use in the longitudinal and lateral-directional cases. Tip-fins are employed for speed brake effect longitudinally and rudder effect laterally.

A trim solution does not guarantee feasibility. Maximum control deflections possible with the vehicle must exceed the solution obtained from these equation sets. Further, aerodynamic angles obtained must not exceed stall values in either α or β .

Longitudinal Trim Results

A note concerning trim solutions for elevator deflection is warranted. WB001's control configuration is depicted in the schematic below (Figure 1). For these calculations, it was assumed both inboard and outboard panels were deflected to trim pitching moment. Obvious advantages in this are smaller control deflections (more surface area moving to counter the moment), resulting in decreased drag due to control movement and a greater percentage of available deflection to address dynamic stability concerns. Control derivatives for the elevator in equation sets 2 and 3 above are, therefore, the addition of drag, lift, and pitching moment derivatives with respect to both aileron and elevon movement. Longitudinal trim results are shown in Figure 2.

Solutions at Mach 0.30 yield trim flight path angles on the order of 12 - 13 degrees (orbiter is $\approx 20.0^\circ$) and angles of attack varying from around 17.8 degrees at 10,000 feet to 12.6 degrees at sea-level. These calculations presume a standard atmosphere. Trim angles of attack and elevator

Moment Trim

The six degrees of freedom available to an aircraft are represented by six equations of motion (conservation of linear and angular momentum). Modeling aerodynamic forces as a first-order Taylor Series expansion and writing the equations for rectilinear flight yields:

$$\begin{aligned}
 mg \sin \gamma &= -(C_{D_o} + C_{D_\alpha} \alpha + C_{D_{\delta_{bf}}} \delta_{bf} + C_{D_{\delta_e}} \delta_e + C_{D_{\delta_{tf}}} \delta_{tf}) \bar{q} S \\
 -mg \sin \phi \cos \gamma &= (C_{Y_o} + C_{Y_\beta} \beta + C_{Y_{\delta_{ail}}} \delta_{ail} + C_{Y_{\delta_{tf}}} \delta_{tf}) \bar{q} S \\
 mg \cos \phi \cos \gamma &= (C_{L_o} + C_{L_\alpha} \alpha + C_{L_{\delta_{bf}}} \delta_{bf} + C_{L_{\delta_e}} \delta_e + C_{L_{\delta_{tf}}} \delta_{tf}) \bar{q} S \\
 0 &= C_{l_\beta} \beta + C_{l_{\delta_{ail}}} \delta_{ail} + C_{l_{\delta_{tf}}} \delta_{tf} \\
 0 &= C_{m_o} + C_{m_\alpha} \alpha + C_{m_{\delta_{bf}}} \delta_{bf} + C_{m_{\delta_e}} \delta_e + C_{m_{\delta_{tf}}} \delta_{tf} \\
 0 &= C_{n_\beta} \beta + C_{n_{\delta_{ail}}} \delta_{ail} + C_{n_{\delta_{tf}}} \delta_{tf}
 \end{aligned} \tag{2}$$

Further modification to these equations is possible by assuming zero bank angle, $\phi = 0$. This substitution decouples the longitudinal set of equations from the lateral and is representative of actual mission profiles flown from 10,000 feet to the runway.

Longitudinal Equations of Motion

$$\begin{aligned}
 mg \sin \gamma &= -(C_{D_o} + C_{D_\alpha} \alpha + C_{D_{\delta_{bf}}} \delta_{bf} + C_{D_{\delta_e}} \delta_e + C_{D_{\delta_{tf}}} \delta_{tf}) \bar{q} S \\
 -mg \cos \gamma &= (C_{L_o} + C_{L_\alpha} \alpha + C_{L_{\delta_{bf}}} \delta_{bf} + C_{L_{\delta_e}} \delta_e + C_{L_{\delta_{tf}}} \delta_{tf}) \bar{q} S \\
 0 &= C_{m_o} + C_{m_\alpha} \alpha + C_{m_{\delta_{bf}}} \delta_{bf} + C_{m_{\delta_e}} \delta_e + C_{m_{\delta_{tf}}} \delta_{tf}
 \end{aligned} \tag{3}$$

In these equations, flight path angle (γ), angle of attack (α), body-flap deflection (δ_{bf}), elevator deflection (δ_e), and tip-fin deflection (δ_{tf}) are unknowns. Reducing the five unknowns to three is possible by first setting the body-flap deflection to zero, typically done for shuttle in the final stages of descent to ensure ground clearance at touchdown. Eliminating angle of attack dependency is possible by solving the set with LaRC aerodynamic characteristics for WB001 at each angle of attack from -10.0 to 20.0 degrees (subsonic range of useful α). The solution set: γ , δ_e , and δ_{tf} , represents a longitudinal trim condition.

As current work concentrates on the descent-to-land mission phase (the aircraft is in unpowered, gliding flight), thrust forces of Table 1 are not considered. The only derivative contrary to our rules is the pitching moment response to a linear velocity, C_{mu} . Inclusion is a function of its physical significance.

Static Stability Results

Where possible, stability derivatives were determined directly from the LaRC aerodynamic database by applying a forward differencing method between available data points. Curve-fits were applied to data which was highly non-linear with the perturbed variable of interest (drag with angle of attack and longitudinal forces and moments with u-velocity (Mach number)). Results of this analysis are presented in Table 2.

TABLE 2.- WB001 STATIC STABILITY

Forces and Moments	Perturbed Variable							
	u	v	w	$\beta = v/U_1$	$\alpha = w/U_1$	p	q	r
Drag	STABLE							
Sideforce		STABLE						
Lift			STABLE					
Roll Moment				NOTE 2		STABLE		
Pitch Moment	NOTE 1				STABLE		STABLE	
Yaw Moment				NOTE 3				STABLE

Note 1: This SSTO variant exhibits a tendency to tuck its nose with increasing velocity ($C_{mu} < 0$) for angles of attack greater than 3.25 degrees (Mach = 0.30) and at α s greater than -8.04° at Mach = 0.60.

Note 2: Dihedral effect ($C_{l\beta}$) is unstable (positive) at Mach 0.30 and angles of attack less than 12.0 degrees. For Mach = 0.60, the vehicle is statically unstable at less than 10.0 degrees α .

Note 3: WB001 is statically unstable for all flight conditions studied.

DISCUSSION

Static Stability

Rather than look at every force and moment response to every input motion variable, an arbitrary set of static stability criterion is established:

- Velocity disturbances are initially opposed only by forces.
- Rotational velocity disturbances are initially opposed only by moments.
- Angle of attack and sideslip disturbances obtained by interpreting the velocity disturbances v and w as $\beta = v/U_1$ and $\alpha = w/U_1$ are initially opposed only by moments.

Static stability is indicated when the *initial* response of the vehicle is a return toward the equilibrium point.

The change in side-force coefficient, C_Y , with change in sideslip angle, β , serves to illustrate the shorthand notation used with stability derivatives:

$$\frac{\partial C_Y}{\partial \beta} = C_{Y\beta} \tag{1}$$

The criteria for static stability are outlined in Table 1.

TABLE 1.- AIRPLANE STATIC STABILITY CRITERIA

Forces and Moments	Perturbed Variable							
	u	v	w	$\beta = v/U_1$	$\alpha = w/U_1$	p	q	r
Drag + X-Thrust	$C_{Du} > 0$							
Sideforce + Y-Thrust		$C_{Y\beta} < 0$						
Lift + Z-Thrust			$C_{L\alpha} > 0$					
Roll Moment + Roll-Thrust				$C_{l\beta} < 0$		$C_{lp} < 0$		
Pitch Moment + Pitch-Thrust	$C_{mu} > 0$				$C_{m\alpha} < 0$		$C_{mq} < 0$	
Yaw Moment + Yaw-Thrust				$C_{n\beta} > 0$				$C_{nr} < 0$

high-performance aircraft are purposely designed statically unstable, increasing their maneuverability. Instabilities are easily controlled using feedback: pilot-in-the-loop, automatic flight control systems, or both.

Dynamic stability is the time-response of an aircraft to a perturbation. Dynamic instability is also acceptable in an aircraft, presuming the divergence from an equilibrium condition is slow enough. These instabilities are also dealt with using human or automatic feedback.

Moment Trim

Separate from, yet related to, stability is moment trim. For an aircraft to achieve an equilibrium flight condition, moments about the center of gravity, both longitudinally and laterally, must balance. This balance is dependent on static stability derivatives and control effectiveness, measured in terms of control derivatives, useful aerodynamic angles of attack, and maximum allowed deflections. An altitude and airspeed where moment trim is not achievable is an infeasible flight condition, regardless of the level of static or dynamic stability.

INTRODUCTION

A winged-body, single-stage-to-orbit (SSTO) vehicle is under evaluation as a candidate to serve as the next generation space transportation system. Intended to service International Space Station Alpha, and with a payload capability roughly equivalent to the current shuttle system, SSTO configuration WB001 (winged-body 001) represents the best ideas from emerging technologies, experience with the Space Shuttle system, the National Aero-Space Plane studies, and the Single Stage Rocket Technology study.

Engineers at Langley Research Center (LaRC) have developed a preliminary aerodynamics database and weight estimates. Team members at Marshall Space Flight Center (MSFC) lead the effort and have contributed preliminary mission parameters and initial vehicle geometry. One of Johnson Space Center's (JSC) tasks involved analyzing descent-to-landing trajectories using the Shuttle Engineering Simulator (SES). This effort currently focuses on the last eighty plus seconds of the flight (10,000 feet to the runway).

Review of LaRC's aerodynamic information indicates static lateral instabilities in the yaw and roll moment derivatives, and suggests a more comprehensive static stability analysis. Next logical steps include determination of possible longitudinal and lateral-directional trim points. This should assist in trajectory design by providing an initial 'point of departure' from currently-modeled shuttle orbiter mission profiles. A study was also accomplished to improve vehicle lateral stability (increase $C_{n\beta}$), concentrating on increasing size of the tip-fin control surface.

Static vs. Dynamic Stability

Aircraft static stability is determined by analyzing forces and moments generated by vehicles in response to a perturbation. This change in force or moment coefficient to a given input ($\partial\text{Force}/\text{Moment Coefficient}/\partial\text{Motion}$) is called a stability derivative. A set of generally recognized rules specifies the sign (positive or negative) the derivative's value must have for static stability. It is important to recognize static stability is not a requirement. Most modern

ABSTRACT

Three NASA centers: Marshall Space Flight Center (MSFC), Langley Research Center (LaRC), and Johnson Space Center (JSC) are currently involved in studying a family of single-stage- and two-stage-to-orbit (SSTO/TSTO) vehicles to serve as the next generation space transportation system (STS). A rocketed winged-body is the current focus. The configuration (WB001) is a vertically-launched, horizontally-landing system with circular cross-section. Preliminary aerodynamic data was generated by LaRC and is a combination of wind-tunnel data, empirical methods, and Aerodynamic Preliminary Analysis System- (APAS) generated values.

JSC's efforts involve descent trajectory design, stability analysis, and flight control system synthesis. Analysis of WB001's static stability indicates instability in 'tuck' ($C_{mu} < 0$: Mach = 0.30, $\alpha > 3.25^\circ$; Mach = 0.60, $\alpha > 8.04^\circ$), an unstable dihedral effect ($C_{l\beta} > 0$: Mach = 0.30, $\alpha < 12.00^\circ$; Mach = 0.60, $\alpha < 10.00^\circ$), and, most significantly, an unstable weathercock stability derivative, $C_{n\beta}$, at all angles of attack and subsonic Mach numbers.

Longitudinal trim solutions for Mach = 0.30 and 0.60 indicate flight path angle possibilities ranging from around 12 ($M = 0.30$) to slightly over 20 degrees at Mach = 0.60. Trim angles of attack increase from 6.24° at Mach 0.60 and 10,000 feet to 17.7° at Mach 0.30, sea-level. Lateral trim was attempted for a design cross-wind of 25.0 knots. The current vehicle aerodynamic and geometric characteristics will only yield a lateral trim solution at impractical tip-fin deflections ($\approx 43^\circ$) and bank angles (21°).

A study of the lateral control surfaces, tip-fin controllers for WB001, indicate increased surface area would help address these instabilities, particularly the deficiency in $C_{n\beta}$, but obviously at the expense of increased vehicle weight. Growth factors of approximately 7 were determined using a design $C_{n\beta}$ of 0.100/radian (approximate subsonic values for the orbiter).

ABSTRACT

The writing of code for capture, in a uniform format, of bit maps of words and characters from scanner PICT files. The coding of Dynamic Pattern Matcher for the identification of the characters, words and sentences in preparation for translation.

INTRODUCTION

There are many commercial software language to language translators available, however there does not appear (as of June,1994) to be any that translate from Russian to English. A search revealed that there was no commercially available optical character reader (OCR) for the Cyrillic alphabet.

There are several difficulties encountered in attempting such translations. The very first is the capture and identification of characters of the Cyrillic alphabet, that has several unusual characters as well as a large character set. The task is not lessened by the appearance from place to place in the text of non Cyrillic characters as well as the occasional diagram. After having identified the characters, or at least some high percentage of them, the next task, which is considerably larger, is the translation of the words identified into English in a fashion that is unstilted as well as accurately representing the meaning of the original text. The need for such software is clear in view of the cooperative venture, space station.

DISCUSSION AND RESULTS

The work this summer produced software that will, from PICT files, identify characters and words. The pattern recognizer employed is the dynamic pattern matcher (DPM) developed earlier by the author and his NASA colleague. The code for the translation of the found words has not been done, but it has been observed that scientific Russian lends itself to translation much more readily than, for example, articles on philosophy.

The overall plan is to have the documents scanned and stored. The user would then view the scans of the documents at a terminal and select, with the aid of a 'mouse', those portions that are to be translated. These portions maybe entire pages or parts thereof. After having selected the material that is to be translated, the user turns control back to the machine. The machine begins it work, usually during the overnight period when there are several machines (SUNS)

that are not in use. If the amount to be translated is small, say a few paragraphs, then it could be processed at that time.

Using parallel virtual machine (PVM), a software package we obtained from Oak Ridge National Laboratory (ORNL) where it was developed, unused machines (SUNS) are linked during the overnight period to form a large parallel machine. Mr. David Merritt has implemented PVM on a collection of our (SUNS). PVM should provide for rapid processing of the portions selected for translation. The use of a large array of machines obviously allows the use of large data sets for the identifying of characters and words as well as the processing power of the many processors.

The idea is to have character sets and a dictionary on line so that if a particular character is unrecognizable, then the remaining characters in conjunction with the dictionary will help determine the unresolved character. This will be then noted and during translation a decision will be made to determine, using the context in which the word in question is used, if the correct character, and hence word was used. This will not be seen by the user unless the matter in question cannot be resolved. In that instance the end user will be presented, in braces in order of preference, a list of possible alternatives for the questionable word or phrase. The result is that recognizing characters is closely tied to the translation to be made and not a separate process.

Text with embedded diagrams is dealt with by the determination of regions of non text items, using such statistics as average line height, line spacing and average character width. These regions are then isolated and the remaining portions are then processed. Diagrams are returned unprocessed. However, if a portion of a diagram has text and that text is selected by the user, it will be processed.

The machines will report back the next day with the results. The results reported will have the appearance of the original document as to position and font size, but in English. Should there be additional portions of the text that require translation or newly added text, the process maybe repeated.

CONCLUSIONS

It has been reported that several governmental agencies have been working on this problem for many years, which is an indication that

the general problem is quite difficult. Our hope is that the documents to be dealt with are all of a scientific nature, which we have already noticed are considerably easier to process. A second point is that even if the translations produced are not highly accurate, they should provide sufficient information to decide if the material deserves a precise, but costly, manual translation.

The search for such an OCR that can deal with the Cyrillic alphabet and a Russian to English translator has continued throughout the summer. During the first week of August a new software package has been released by a small group located in California for Russian to English translation as well as, at least claimed to be, a perfect Cyrillic OCR. During the second week of August we obtained a copy of each for evaluation and testing. The translator package 'Stylus' by Project MT Ltd., the OCR package 'Cuneiform' by Cognitive Technology, Cyrillic font set 'ParaWin+fonts' by ParaGraph International and finally a Russian dictionary 'Propis' by Agama. These software packages were supplied by Russian Business Services, Inc of Houston, Texas.

Section 17 not used.

**NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS
BY ARTIFICIAL NEURAL NETWORKS**

Final Report

NASA/ASEE Summer Faculty Fellowship Program-1994

Johnson Space Center

Prepared By:	Andrew J. Meade, Jr., Ph.D
Academic Rank:	Assistant Professor
University & Department	Rice University Department of Mechanical Engineering and Materials Science Houston, Texas 77251-1892
NASA/JSC	
Directorate:	Information Systems
Division:	Information Technology
Branch:	Software Technology
JSC Colleague:	Robert O. Shelton, Ph.D
Date Submitted:	August 12, 1994
Contract Number:	NGT-44-005-803

ABSTRACT

Conventionally programmed digital computers can process numbers with great speed and precision, but do not easily recognize patterns or imprecise or contradictory data. Instead of being programmed in the conventional sense, artificial neural networks (ANNs) are capable of self-learning through exposure to repeated examples. However, the training of an ANN can be a time consuming and unpredictable process.

A general method is being developed by the author to mate the adaptability of the ANN with the speed and precision of the digital computer. This method has been successful in building feedforward networks that can approximate functions and their partial derivatives from examples in a single iteration. The general method also allows the formation of feedforward networks that can approximate the solution to nonlinear ordinary and partial differential equations to desired accuracy without the need of examples. It is believed that continued research will produce artificial neural networks that can be used with confidence in practical scientific computing and engineering applications.

INTRODUCTION

Artificial neural networks (ANNs) have proven to be versatile tools for accomplishing what could be termed higher order tasks such as pattern recognition, classification, and visual processing. However, conventional wisdom has held that networks are unsuited for use in more purely computational tasks, such as mathematical modelling and physical analysis of engineering systems. Certainly the biological underpinnings of the neural network concept suggest that networks would perform best at tasks at which biological systems excel, and worse or not at all at other tasks.

Contrary to opinion the author believes that continued research into the approximation capabilities of networks will enable the neural network paradigm, with all of its advantages in behavior and adaptability, to be mated to the more purely computational paradigms of mathematically oriented scientific programming and analysis. Additionally, it is felt that the thorough investigation of network approximation capabilities will benefit the network field and connectionism in general.

In a field as conceptually difficult as the study of artificial neural networks, it is best to start investigation with supervised learning, test the established premises, and alter them to circumvent pitfalls in implementation.

FUNCTION APPROXIMATION

Learning as Function Approximation

Central to the author's research approach is the view that supervised learning in artificial neural networks is equivalent to the problem of approximating a multivariate function and that learning should be able to be explained by approximation theory. Approximation theory deals with the problem of approximating or interpolating a multivariate function. This approach has been considered by other researchers in the field of ANNs [1]-[4]. However, the author extends this assumption of function approximation by assuming that ANNs can model discontinuous multivariate functions and should be at least as accurate and numerically efficient as existing computational techniques used in science and engineering. Also, ANN behavior and adaptation difficulties, from supervised learning to machine vision, should be amenable to the standard error analysis techniques used in numerical analysis.

In a general sense experiments, analytical methods, and computational methods can be considered to be forms of function approximation. The governing equations derived from analytical methods are a compact representation of the functions that model some particular phenomena observed in experiments. Computational techniques are used to approximate the function or functions that satisfy the governing equations. The graphs and tables

made from experiments are representations of the functions that underlie observed physical phenomena.

PROGRAMMABLE ARTIFICIAL NEURAL NETWORKS

If we are to assume that ANNs are as valid as established computational techniques, then ANNs should be evaluated in the same manner as are computational techniques. In evaluating the capabilities of a new numerical method, it is prudent to first apply it to the solution of algebraic and ordinary and partial differential equations of known behavior. This same approach can be used for ANNs since the solution of algebraic and differential equations can be viewed as the approximation of a function that must satisfy the equation in question subjected to boundary and/or initial conditions.

Applying an ANN to the solution of an algebraic or differential equation effectively uncouples the influences of the quality of data samples, network architecture, and transfer functions from the network approximation performance. The solution of equations also allows us to study the influence of constraining the connection weights. The most immediate benefit in this approach would be the construction of networks that can approximate the solution to desired equations without the need for examples. This would be of value in engineering applications since considerable effort may be saved if the equations governing a physical process can be directly incorporated into the neural network architecture without the need of examples, thereby shortening or even eliminating the learning phase.

The author has previously reported the use of the piecewise linear hard limit transfer function [5] in solving model algebraic and linear ordinary differential equations by the feedforward architecture [7]. In this report the author will use the recurrent artificial neural network (RANN) architecture to model a chaotic system by solving a nonlinear ordinary differential equation known as Duffing's equation.

APPROACH

A simple RANN consists of two layers of processing elements as illustrated in Fig. 1 for two coupled time-dependent variables u and v , where I denotes the input layer, and II denotes the output layer. Generally speaking, the inputs u and v , fed into layer I at time t^n , are operated on by the processing elements (PEs) and then multiplied by the constant connection weights a_{iq} . The coefficient a_{iq} is the connection weight between the i th neuron of layer I and the q th neuron of layer II . As a result, the outputs of layer I , multiplied by the connection weights, are input to layer II PEs. The output of layer II neurons are the values of the dependent variables at time t^{n+1} . These values are in turn fed back into the input layer I for the next iteration.

The dynamic behavior of such a RANN is governed by the value of the connection weights, the type of processing elements in the RANN, and the transfer functions used. In this study we restrict ourselves to using the piecewise linear transfer function in all layers of the RANN, since it is one of the simplest functions to implement on both digital and analog forms [8].

The hard limit function $\Upsilon_q^K(u)$ (Fig. 2) operating on the dependent variable u for the q th neuron of the K th layer can be modelled by the following equations:

$$\Upsilon_q^K(u) = \begin{cases} -1.0 & \text{for } \xi_q < -1.0 \\ \xi_q & \text{for } -1.0 \leq \xi_q \leq 1.0 \\ +1.0 & \text{for } \xi_q > 1.0 \end{cases}$$

where

$$\begin{aligned} \xi_q &= \sum_i a_{iq} u_i \quad \text{for additive neurons} \\ \xi_q &= \prod_i a_{iq} u_i \quad \text{for multiplicative neurons} \end{aligned}$$

By correctly assigning different values to the RANN connection weights, one is able to integrate a system of coupled recurrent relations into the architecture without the need of training. We will now show how to obtain such recurrent relations from a system of ordinary differential equations.

Duffing's Equation

The inhomogeneous Duffing's equation describes the forced motion of a particle between two equilibrium states and can be written as the following nondimensional second-order ordinary differential equation

$$\frac{d^2x}{dt^2} + 2\mu \frac{dx}{dt} - \frac{1}{2}(x - x^3) = F_0 \cos(\omega t) \quad (1)$$

where t , x , μ , F_0 , and ω represent time, displacement, damping coefficient, force amplitude, and frequency of excitation respectively. The importance of this equation is that its chaotic and nonchaotic behavior has been extensively examined by theoretical [9], experimental [10], and numerical methods [6], [11] - [13].

In order that Duffing's equation can be more easily approached by the RANN programming method and so that we may obtain particle displacement, velocity, and acceleration, Eq.(1) is reduced to a system of two first-order equations by the following change of variables

$$s = \frac{x}{\kappa_1} \quad , \quad y = \frac{1}{\kappa_2} \frac{dx}{dt} = \frac{1}{\kappa_2} \dot{x} \quad \text{and} \quad w = \frac{1}{\kappa_3} \frac{d^2x}{dt^2} = \frac{\kappa_2}{\kappa_3} \frac{dy}{dt}$$

where κ_1 , κ_2 , and κ_3 are constants used to scale the values of the new variables. Therefore, Eq.(1) becomes

$$\frac{ds}{dt} = \kappa_4 y \quad (2)$$

$$\frac{dy}{dt} = -2\mu y + \frac{1}{2\kappa_4} (s - \kappa_1^2 s^3) + \frac{F_0}{\kappa_2} \cos(\omega t) = \frac{\kappa_3}{\kappa_2} w \quad (3)$$

$$\text{where } \kappa_4 = \frac{\kappa_2}{\kappa_1}.$$

It will now be demonstrated how the RANN architecture can be used to approximate the solution to Eq.(2) and Eq.(3) for arbitrary coefficients.

Integration of Duffing's Equation

For this problem the initial value problem of Eqs.(2) and (3) will be integrated by the explicit MacCormack method [14] which is a finite-difference, predictor-corrector scheme commonly used in the solution of time-dependent fluid dynamics equations. The application of the MacCormack technique to a general first order ordinary differential equation

$$\frac{du}{dt} - f(u, t) = 0 ,$$

where f is some arbitrary function of the dependent variable u and independent variable t , results in

$$\text{Predictor: } u^* = u^n + \Delta t f(u^n, t^n)$$

$$\text{Corrector: } u^{n+1} = \frac{1}{2} (u^n + u^* + \Delta t f(u^*, t^*)) = u^n + \frac{\Delta t}{2} (f(u^n, t^n) + f(u^*, t^*))$$

where the superscripts denote the time level and $t^* = t^n + \Delta t$.

Application of the MacCormack method to Eqs.(2) and (3) results in

$$\begin{aligned} s^* &= s^n + \Delta t f_1(s^n, y^n, t^n) \\ s^{n+1} &= s^n + \frac{\Delta t}{2} (f_1(s^n, y^n, t^n) + f_1(s^*, y^*, t^*)) \\ y^* &= y^n + \Delta t f_2(s^n, y^n, t^n) \\ y^{n+1} &= y^n + \frac{\Delta t}{2} (f_2(s^n, y^n, t^n) + f_2(s^*, y^*, t^*)) \end{aligned}$$

where

$$f_1(s, y, t) = \kappa_4 y \quad \text{and} \quad f_2(s, y, t) = -2\mu y + \frac{1}{2\kappa_4} (s - \kappa_1^2 s^3) + \frac{F_0}{\kappa_2} \cos(\omega t)$$

Substitution of the expressions for f_1 , f_2 , s^* , and y^* into the equations for s^{n+1} and y^{n+1} results in the following algebraic system,

$$p^n = (s^n)^3 \quad (4)$$

$$r^n = (s^*)^3 = (s^n + \kappa_4 \Delta t y^n)^3 \quad (5)$$

$$s^{n+1} = H_1 s^n + H_2 y^n + H_3 p^n + H_4 \cos(n\omega\Delta t) \quad (6)$$

$$y^{n+1} = H_5 s^n + H_6 y^n + H_7 p^n + H_8 r^n + H_9 \cos(n\omega\Delta t) + H_{10} \cos((n+1)\omega\Delta t) \quad (7)$$

The coefficients H_1 through H_{10} are

$$\begin{aligned} H_1 &= 1 + \frac{\Delta t^2}{4} \quad , \quad H_2 = \kappa_4 \Delta t (1 - \mu \Delta t) \quad , \quad H_3 = -\frac{\kappa_1^2 \Delta t^2}{4} \\ H_4 &= \frac{F_0 \Delta t^2}{2\kappa_1} \quad , \quad H_5 = \frac{\Delta t}{2\kappa_4} (1 - \mu \Delta t) \quad , \quad H_6 = 1 - 2\mu \Delta t + 2\mu^2 \Delta t^2 + \frac{\Delta t^2}{4} \\ H_7 &= -\frac{\kappa_1^3 \Delta t}{\kappa_2} (1 - 2\mu \Delta t) \quad , \quad H_8 = -\frac{\kappa_1^3 \Delta t}{\kappa_2} \frac{\Delta t}{4} \quad , \quad H_9 = \frac{F_0 \Delta t}{\kappa_2} \frac{\Delta t}{2} (1 - 2\mu \Delta t) \\ H_{10} &= \frac{F_0 \Delta t}{\kappa_2} \frac{\Delta t}{2} \quad , \quad (8) \end{aligned}$$

If Δt is required to be constant, then the time-dependent coefficients of Eq.(8) become constants for specific values of μ , F_0 , κ_1 , κ_2 , and κ_3 . Equations (4) through (7) are the linear and nonlinear algebraic equations that approximate Duffing's equation and must be modelled by the RANN.

Network Approximation of Duffing's Equation

To compare against previous numerical studies of Duffing's equation, we require only that the RANN generate values for particle displacement and velocity. Figures 3 - 5 show the connections required in a constrained RANN modelling Eqs.(4) through (7) respectively. It is understood that the network output is fed back to the input layer though the connections are not shown in the figures. An initial time $t^0 = 0$ is assumed.

Note that the magnitude of all inputs and outputs are scaled to be less than value 1 so that hard limits may be used in the processing elements. This constraint on the magnitude of the dependent variables s^n , y^n , p^n , and r^n requires that

$$\kappa_1 > |x|_{max} + \Delta t \left| \frac{dx}{dt} \right|_{max} \quad , \quad \kappa_2 > \left| \frac{dx}{dt} \right|_{max} \quad , \quad \kappa_3 > \left| \frac{d^2x}{dt^2} \right|_{max} \quad (9)$$

Equations (4) and (5) possess cubic time-dependent unknowns that are approximated by the RANN using the multiplicative neuron model also known as the Pi neuron (Fig. 3) [5]. In traditional ANNs, the Pi neurons are used to achieve greater processing power by using more complex connections.

The fully assembled RANN is shown in Fig. 6. Updated values of the cosine functions ($\cos(n\omega\Delta t)$ and $\cos((n+1)\omega\Delta t)$) must be input at each iteration in the RANN.

RESULTS

The RANN of Fig. 6 was constructed and run with a fixed damping value of $\mu = 0.084$ and fixed forcing amplitude of $F_0 = 0.178$. Previous results of the two-well potential system [6] indicate displacement ranging between ± 2.0 and velocity ranging between ± 1.0 . Assuming that $\Delta t < |x|_{max}$ Eq.(9) allows the use of $\kappa_1 = 4.0$, $\kappa_2 = 2.0$, and $\kappa_3 = 1.0$. Initial conditions were set for $x(0) = 1.0$ and $\dot{x}(0) = 0.0$, except for Fig. 8 where $x(0) = 0$ and $\dot{x}(0) = 1.0$ for the coexisting global attractor. To ensure that steady-state solutions were displayed, approximations were run for 900 periods, T , based on the forcing frequency ω ($T = 2\pi/\omega$). All computations were done in double precision and all test cases were run in less than 40 real time seconds on a Sparc 10 SX Model 512 workstation.

RANN Simulation Results

Figures 7 - 9 compare the constructed RANN output ($x = \kappa_1 s$ and $\dot{x} = \kappa_2 y$) with the numerical results of Masoud and Asfar [6], for the same parameter values and initial conditions. Here, as with reference [6] the damping and excitation amplitude are based on the work presented by Pezeshki and Dowell [13]. Throughout the simulation the time step used in the RANN was identical to that used in the Masoud and Asfar study, $\Delta t = T/100$. For clarity, plots of time histories have been excluded. Phase plane trajectories, as well as Poincaré maps are shown and compared in cases where they are shown by reference [6].

Figure 7 starts with $\omega = 1.020$ at which the period of the oscillator is equal to 1 (period 1 oscillation) and doubles as ω decreases. Thus we observe period 2, period 4, and period 8 oscillations, until the system evolves into a coexisting global period 5 motion (Fig. 8) at which the system oscillates between the two equilibrium states (\dot{x}, x) respectively $(0, -1)$ and $(0, 1)$. Figure 9 describes the first region of chaotic oscillations starting at $\omega = 0.865$ and ending at $\omega = 0.780$.

CONCLUSIONS

Using the piecewise linear transfer function and additive and multiplicative neurons the author was able to construct a recurrent artificial neural network that was capable of accurately solving a complex nonlinear problem. The solution of the inhomogenous Duffing's equation was shown as a numerical

example and compared well with the computational results of Masoud and Asfar. This was done without the need for data or conventional training. By establishing a direct link between a well known numerical method and the operation of a RANN we believe it may be possible to link common numerical methods from computational mechanics to the operation of the more popular ANN paradigms.

REFERENCES

- [1] Barron, A.R. and Barron, R.L., "Statistical learning networks: a unifying view," *Symposium on the Interface: Statistics and Computing Science*, Reston, Virginia, April 1988.
- [2] Omohundro, S., "Efficient algorithms with neural network behaviour," *Complex Systems*, **1**, 237, 1987.
- [3] Poggio, T. and Girosi, F., "A Theory for Networks for Approximation and Learning," *Artificial Intelligence Laboratory, Massachusetts Institute of Technology*, A.I. Memo No. 1140, July 1989.
- [4] Girosi, F. and Poggio, T., "Networks for learning: a view from the theory of approximation of functions," *Proceedings of the Genoa Summer School on neural networks and their applications*, Prentice Hall, 1989.
- [5] Maren, A., Harston, C. and Pap, R., *Handbook of Neural Computing Applications*, 48, Addison-Wesley Publishing, 1990.
- [6] Masoud, K. K. and Asfar, K. R., "Period Doublings, Bifurcations and Strange Attractors in a Two-Well Potential Oscillator," *European Journal of Mechanics, A/Solids*, Vol. 12, No. 3, 1993, pp. 417-428.
- [7] Meade, A. J., "Development of Programmable Artificial Neural Networks," *NASA Contractor Report*, Contract Number NGT-44-001-800, 1993, pp. 21-1 - 21-15.
- [8] Chua, L. O., Desoer, C. A. and Kuh, E. S., *Linear and Nonlinear Circuits*, McGraw Hill, New York, 1987. pp. 76-83, 171-212.
- [9] Holmes, P., "A Nonlinear Oscillator with a Strange Attractor," *Phil. Trans. The Roy. Soc. London*, Vol. 292, 1979, pp. 419-448.
- [10] Moon, F. C., "Experiments on Chaotic Motions of a Forced Nonlinear Oscillator: Strange Attractors," *ASME Journal of Applied Mechanics*, Vol. 47, 1980, pp. 638-644.

- [11] Ueda, Y., "Randomly Transitional Phenomena in the System Governed by Duffing's Equation," *Journal of Statistical Physics*, Vol. 20, No. 2, 1979, pp. 181-196.
- [12] Dowell, E. H. and Pezeshki, C., "On the Understanding of Chaos in Duffings Equation Including a Comparison With Experiment," *Journal of Applied Mechanics*, Vol. 53, 1986, pp. 5-9.
- [13] Pezeshki, C., and Dowell, E. H., "On Chaos and Fractal Behaviour in a Generalised Duffing's System," *Physica D*, Vol. 32, 1988, pp. 194-209.
- [14] Anderson, D. A., Tannehill, J. C. and Pletcher, R. H., *Computational Fluid Mechanics and Heat Transfer*, Hemisphere Publishing Corporation, New York, 1984, pp. 163-164.

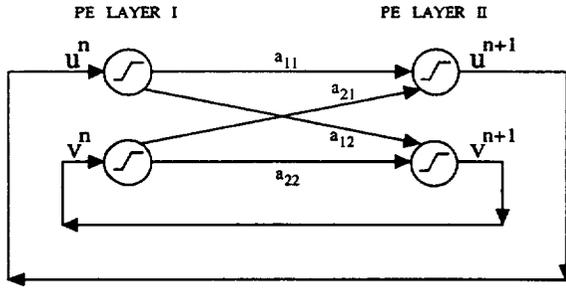


Figure 1: Basic RANN architecture.

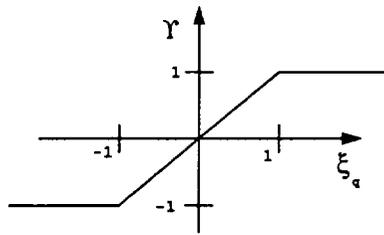


Figure 2: Hard limit transfer function.

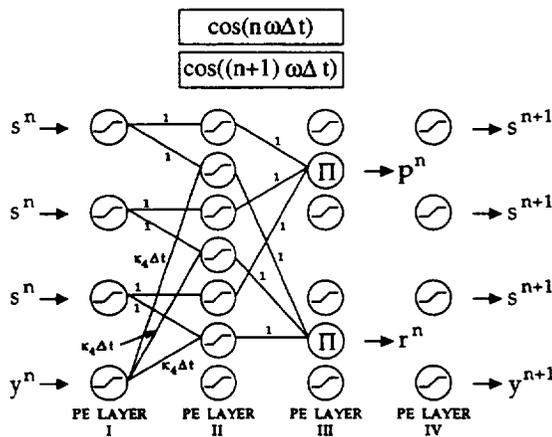


Figure 3: RANN representation of the nonlinear equations for p^n and r^n .

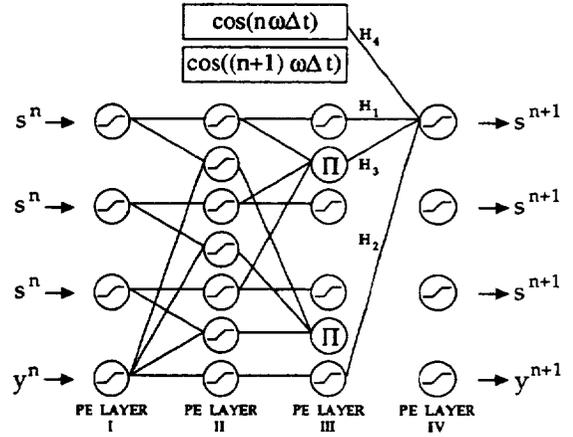


Figure 4: RANN representation of the recurrent equation for s^{n+1} .

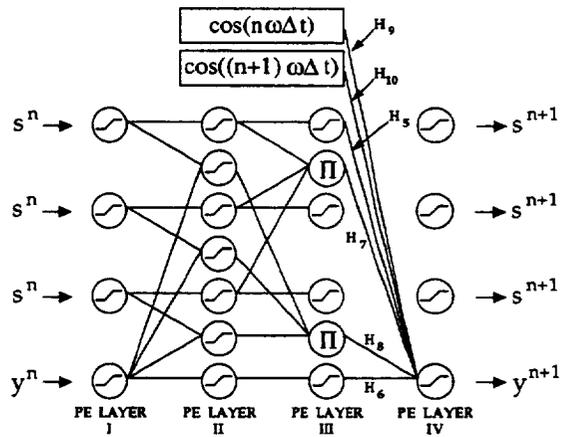


Figure 5: RANN representation of the recurrent equation for y^{n+1} .

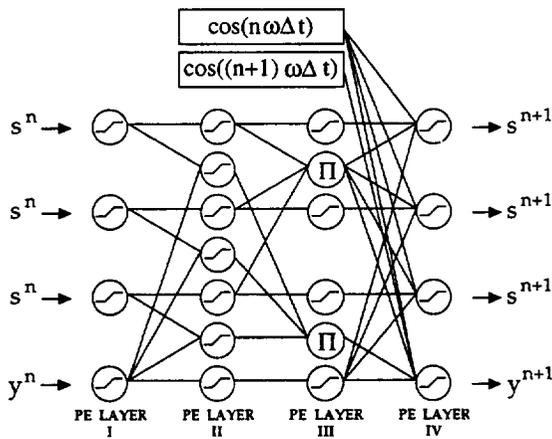


Figure 6: Full RANN assembly for the MacCormack scheme applied to Duffing's equation.

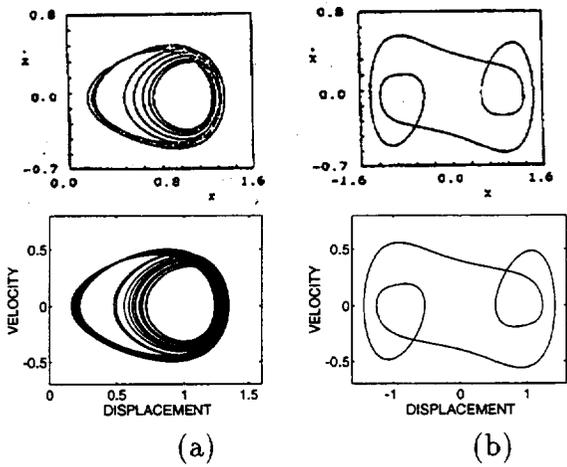


Figure 8: Phase space trajectories from Masoud and Asfar (top set) and the RANN (bottom set). From left to right: coexisting (a) $\omega = 0.957$ period 8, (b) $\omega = 0.957$ global period 5.

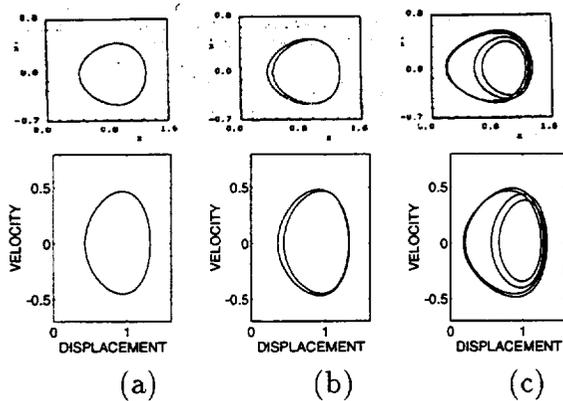


Figure 7: Comparison of phase space trajectories from Masoud and Asfar (top set) and the RANN (bottom set). From left to right: (a) $\omega = 1.020$ period 1, (b) $\omega = 1.000$ period 2, (c) $\omega = 0.963$ period 4.

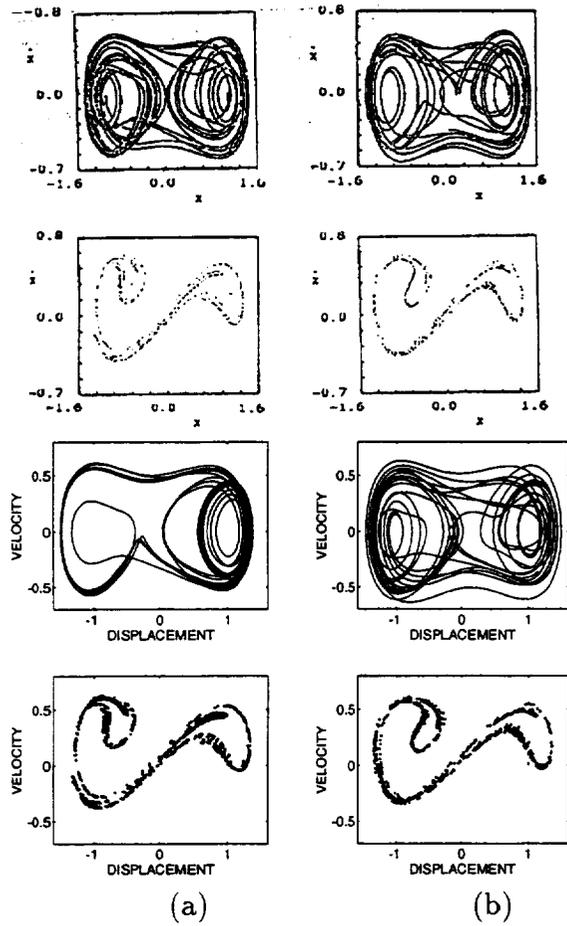


Figure 9: Comparison of phase space trajectories and Poincaré maps in the first chaotic region from Masoud and Asfar (top set) and the RANN (bottom set). From left to right: (a) $\omega = 0.865$ chaotic higher frequency boundary, (b) $\omega = 0.780$ chaotic lower frequency boundary.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE July 1995	3. REPORT TYPE AND DATES COVERED Contractor Report		
4. TITLE AND SUBTITLE National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program - 1994, Volume I			5. FUNDING NUMBERS	
6. AUTHOR(S) Richard Bannerot* and Donn G. Sickorez, Editors				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Lyndon B. Johnson Space Center Houston, Texas 77058			8. PERFORMING ORGANIZATION REPORT NUMBERS	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, D.C. 20546-0001			10. SPONSORING/MONITORING AGENCY REPORT NUMBER CR-188410	
11. SUPPLEMENTARY NOTES *University of Houston, Houston, Texas				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified/Unlimited Available from the NASA Center for AeroSpace Information (CASI) 800 Elkridge Landing Road Linthicum Heights, MD 21090-2934 (301) 621-0390			12b. DISTRIBUTION CODE Subject Category: 99	
13. ABSTRACT (Maximum 200 words) The JSC NASA/ASEE Summer Faculty Fellowship Program was conducted by Texas A&M University and JSC. The objectives of the program, which began nationally in 1964 and at JSC in 1965 are to (1) further the professional knowledge of qualified engineering and science faculty members, (2) stimulate an exchange of ideas between participants and NASA, (3) enrich and refresh the research and teaching activities of participants' institutions, and (4) contribute to the research objectives of the NASA centers. Each faculty fellow spent at least 10 weeks at JSC engaged in a research project in collaboration with a NASA JSC colleague. This document is a compilation of the final reports on the research projects completed by the faculty fellows during the summer of 1994.				
14. SUBJECT TERMS information transfer, research, research projects, urban research, engineering, science, universities, university program			15. NUMBER OF PAGES 202	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

