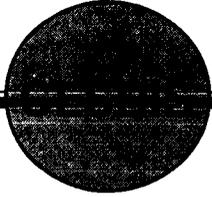




---



# *Center of Excellence in Space Data and Information Sciences*

## *Annual Report*

**Year 10  
July 1997 - June 1998  
Dr. Yelena Yesha, Director**

**CESDIS  
NASA Goddard Space Flight Center  
Code 930.5  
Greenbelt, MD 20771**

**phone: 301-286-4403  
fax: 301-286-1777  
e-mail: [cas@cesdis1.gsfc.nasa.gov](mailto:cas@cesdis1.gsfc.nasa.gov)  
<http://cesdis.gsfc.nasa.gov>**

**CESDIS**

***Center of Excellence  
in Space Data and  
Information Sciences***

***Annual Report***

**Year 10**

**July 1997 - June 1998**

**Dr. Yelena Yesha, Director**

**CESDIS  
NASA Goddard Space Flight Center  
Code 930.5  
Greenbelt, MD 20771**

**phone: 301-286-4403  
fax: 301-286-1777  
e-mail: [cas@cesdis1.gsfc.nasa.gov](mailto:cas@cesdis1.gsfc.nasa.gov)  
<http://cesdis.gsfc.nasa.gov>**

---

---

# ***FOREWORD***

---

This report summarizes the range of computer science-related activities undertaken by CESDIS for NASA in the twelve months from July 1, 1997 through June 30, 1998. These activities address issues related to accessing, processing, and analyzing data from space observing systems through collaborative efforts with university, industry, and NASA space and Earth scientists.

The sections of this report which follow, detail the activities undertaken by the members of each of the CESDIS branches. This includes contributions from university faculty members and graduate students as well as CESDIS employees. Phone numbers and e-mail addresses appear in Appendix E (CESDIS Personnel and Associates) to facilitate interactions and new collaborations.

---

---

# TABLE OF CONTENTS

---

<b>Foreword</b> .....	i
<b>A CESDIS Overview</b> .....	v
<b>The Director</b> .....	1
Activities (Task 1).....	2
Research .....	5
Selected Publications .....	8
<b>Consultants to the Director (Task 1)</b> .....	11
Ian Akyildiz, Georgia Institute of Technology .....	12
Burt Edelson, George Washington University .....	12
Richard Somerville, University of California, San Diego.....	14
Ouri Wolfson, University of Illinois .....	15
<b>Computational Sciences Branch</b> .....	19
Remote Sensing Group	
Jacqueline Le Moigne,	
<i>An Evaluation of Automatic Image Registration Methods (Task 28)</i> .....	20
Richard Lyon,	
<i>Imaging and Phase Retrieval-based Wavefront Sensing (Task 57)</i> .....	40
Timothy Murphy,	
<i>Image Processing and Phase Retrieval (Task 57)</i> .....	50
Nathan Netanyahu,	
<i>Image Registration for the Regional Application Center (Task 44)</i> .....	52
Scalable Systems Technology Group	
Phillip Merkey, Donald Becker, Daniel Ridge, Erik Hendriks	
<i>Beowulf Parallel Workstation (Tasks 38, 70)</i> .....	55
Terrence Pratt,	
<i>HPCC/ESS Evaluation Project (Task 31)</i> .....	59
Udaya Ranawake,	
<i>Highly-parallel Integrated Virtual Environment (HIVE) (Task 65)</i> .....	63
George Lake,	
<i>HPCC Earth and Space Science Project Scientist (Task 31)</i> .....	65
Adam Frank,	
<i>HPCC Earth and Space Science Project PR (Task 31)</i> .....	67
Tarek El-Ghazawi,	
<i>Wavelet-based Parallel Image Registration (Task 71)</i> .....	68
Jules Kouatchou,	
<i>Parallel Wavelet-based Image Registration on the</i>	
<i>Beowulf Architecture (Task 28)</i> .....	73

<b>Applied Information Technology Branch</b> .....	81
Digital Libraries Technology (Task 56) .....	82
Yair Amir,	
<i>Combining Satellite Communications in Commedia</i> .....	82
Susan Hoban,	
<i>NASA Digital Library Technology Project Support</i> .....	86
Global Legal Information Network (Task 69) .....	89
Nabil Adam,	
<i>Information Extraction Applications for GLIN</i> .....	89
Konstantinos Kalpakis,	
<i>Architectural Design for GLIN</i> .....	92
Executive Secretariat to the Data and Information Management Working Group (Task 45)	
Les Meredith .....	93
Executive Secretariat to the CENR (Task 62)	
Sushel Unninayar .....	95
Fran Stetina,	
<i>Pilot EOS Direct Readout Ground Systems Support (Task 61)</i> .....	98
Murray Felsher,	
<i>NASA Science and the Private Sector (Task 67)</i> .....	101
Miodrag (Misha) Rancic,	
<i>Support for NASA/NOAA Collaboration in Earth Science Modeling (Task 79)</i> .....	103
Dinshaw Balsara,	
<i>Development of an Implicit 2D Adaptive Mesh Refinement and De-refinement</i>	
<i>Magnetohydrodynamics Code (Task 68)</i> .....	107
Daniel Menascé, Mukesh Singhal,	
<i>Scalability Analysis of ECS' Data Server (Task 72)</i> .....	109
Santiago Egido Arteaga,	
<i>Numeric Simulation of a Volcanic Jet-plume (Task 73)</i> .....	126
David Ebert,	
<i>Volumetric Display of Earth and Space Science Data (Task 76)</i> .....	127
Lisa Singh,	
<i>Putting Log Data to Work (Task 80)</i> .....	128
<b>Administration Branch</b> .....	131
<b>Appendices</b>	
A. Workshop on Data Mining, Warehousing, and Large Data Recovery .....	135
B. Image Registration Workshop .....	141
C. Seminars .....	147
D. Technical Reports .....	163
E. Personnel and Associates .....	191

---

---

# OVERVIEW

---

CESDIS, the Center of Excellence in Space Data and Information Sciences, was developed jointly by the National Aeronautics and Space Administration (NASA), Universities Space Research Association (USRA), and the University of Maryland in 1988. It is operated by USRA, under a contract with NASA. The program office and a small, core staff are located on site at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

## USRA and the CESDIS Science Council

USRA is a nonprofit consortium of 80 colleges and universities, offering graduate programs in space sciences or related areas, which operates research centers and programs at several NASA centers. Most notable are the Lunar and Planetary Institute (LPI) at the Johnson Space Center in Houston, Texas, the Institute for Computer Applications in Science and Engineering (ICASE) at the Langley Research Center in Hampton, Virginia, the Research Institute for Advanced Computer Science (RIACS) at the Ames Research Center at Moffett Field, California, and the Stratospheric Observatory for Infrared Astronomy (SOFIA) in Waco, Texas.

Oversight of each USRA institute or program is provided by a science council which serves as a scientific board of directors. Science council members are appointed by the USRA Board of Trustees for three-year terms. Members of the CESDIS Science Council during 1997-1998 were:

- Dr. Rama Chellappa  
University of Maryland College Park
- Dr. Burt Edelson  
George Washington University
- Dr. Richard Muntz  
University of California, Los Angeles
- Dr. David Nicol  
Dartmouth College
- Dr. Jacob Schwartz  
New York University
- Dr. Harold Stone (Convener)  
NEC Research Institute
- Dr. Satish Tripathi  
University of California, Riverside
- Dr. Mark Weiser  
Xerox PARC

The CESDIS Science Council meets annually at Goddard to review ongoing CESDIS research programs and new initiatives.

## The CESDIS Mission

CESDIS was formed to focus on the design of advanced computing techniques and data systems to support NASA Earth and space science research programs. The primary CESDIS mission is to increase the connection between computer science and engineering research programs at colleges and universities and NASA groups working with computer applications in Earth and space science. Research areas of primary interest at CESDIS include:

- High performance computing, especially software design and performance evaluation for massively parallel machines,

- 
- 
- Parallel input/output and data storage systems for high performance parallel computers,
  - Parallel hardware and software systems,
  - Database and intelligent data management systems for parallel computers,
  - Image processing,
  - Information technology, and
  - Data compression.

CESDIS funds multiyear projects at U.S. universities and colleges. Proposals are accepted in response to calls for proposals and are selected on the basis of peer reviews. Funds are provided to support faculty and graduate students working at their home institutions. Project personnel visit Goddard during academic recess periods to attend workshops, present seminars and collaborate with NASA scientists on research projects. Additionally, CESDIS takes on specific tasks for computer science research and development requested by NASA Goddard scientists.

A small, core staff is housed on-site at NASA Goddard. (A CESDIS organizational chart is included at the end of this introductory section.) This staff includes USRA employees and university research personnel attached to CESDIS via subcontracts who work in one of three branches: Computational Sciences, Applied Information Technology, or Administration. The bulk of this report describes the work of each branch in detail.

### **CESDIS World Wide Web Homepage**

The CESDIS web site is fully indexed and can be located through:

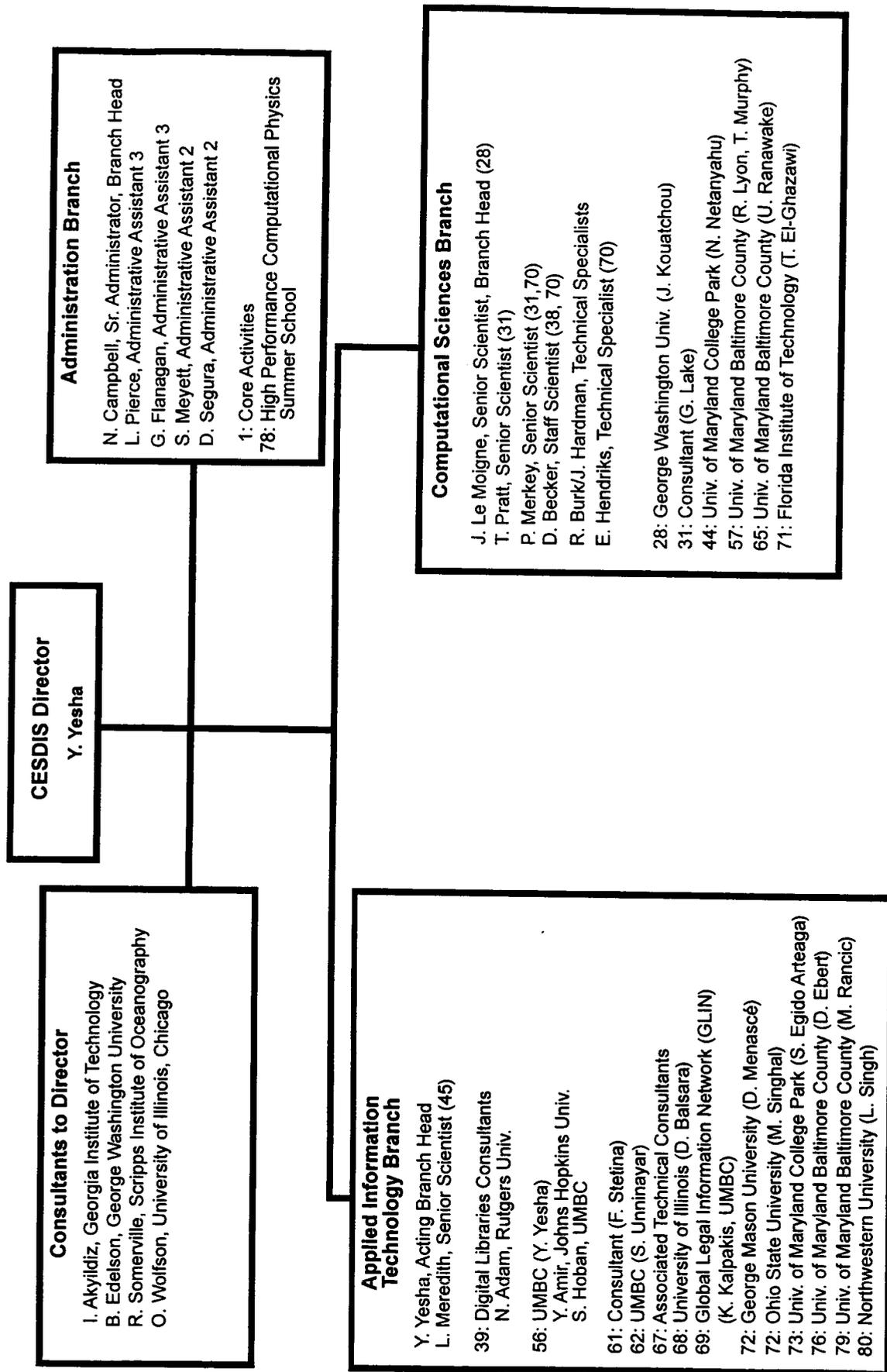
<http://cesdis.gsfc.nasa.gov/>

Contained in this web site are an overview of the CESDIS mission, special announcements, an explanation of the CESDIS organizational structure, and links to specific research projects and accomplishments.

The CESDIS home page is an active link to the heart of CESDIS activities. Feedback and comments are encouraged electronically to:

[cas@cesdis.gsfc.nasa.gov](mailto:cas@cesdis.gsfc.nasa.gov)

# CESDIS ORGANIZATIONAL CHART



---

---

# ***DIRECTOR***

---

**Dr. Yelena Yesha  
(yelena@cesdis.edu)**

Dr. Yelena Yesha is a tenured full professor in the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County (UMBC), holds a joint appointment with the University of Maryland's Institute for Advanced Computer Studies (UMIACS) in College Park, and serves as the CESDIS Director through a memorandum of understanding between the University of Maryland and USRA.

Dr. Yesha received a Bachelor of Science degree in computer science from York University in Toronto, Canada in 1984, and a Master of Science and Ph.D. in computer and information science from Ohio State University in 1986 and 1989 respectively. She is a Senior Member of the IEEE Society, and a member of the ACM and New York Academy of Science. Her research interests include distributed databases, distributed systems, and performance modeling. She has authored numerous papers and edited six books in these areas.

Prior to joining CESDIS in December 1994, Dr. Yesha was on leave from the University to serve as the Director of the Center for Applied Information Technology at the National Institute of Standards and Technology. The Center's mission was to advance the goals of the National Information Infrastructure by identifying, developing, and demonstrating critical new technologies and their applications which could be successfully commercialized by U. S. industry.

## ACTIVITIES

- I held a number of meetings with CESDIS Senior Scientist Jacqueline Le Moigne to discuss the research at CESDIS and prepare for the visit of the CESDIS Science Council on August 12, 1997.
- I met with Mr. Jim Fischer (930) and discussed with him the possibility of hiring an Earth Scientist through CESDIS.
- I met with Dr. Miodrag Rancic of NOAA, a prospective CESDIS consultant or subcontractor. Dr. Rancic's expertise is in the area of Earth science which will complement existing CESDIS expertise.
- CESDIS hosted a workshop on data mining and data warehousing August 19-21, 1997. Top researchers from all over the country gave presentations in these areas. The workshop was concluded by a panel on strategic research directions for NASA in the area of data warehousing and data mining. More information is available in Appendix A.
- I made significant progress on completing the manuscript on electronic commerce to be published by MIT Press. I also completed a proposal to the National Science Foundation that CESDIS jointly submitted with researchers from the Johns Hopkins University.
- I spent time with member of code 931 to understand their requirements for building a data warehouse.
- Professor Scheuermann from Northwestern University visited CESDIS. He has been on leave from Northwestern, and is serving as Program Director of Computer Systems Software in the Division of Computer and Computation Research at the National Science Foundation. Prof. Scheuermann met with CESDIS staff, and we are exploring the potential for collaboration between CESDIS and Northwestern University.
- I held a meeting with the Beowulf team and a few professors from UMBC to discuss the possibility of building Beowulf at UMBC.
- I held a meeting with Karen Moe of EOSDIS (588) and Professor Alex Brodsky of George Mason University. Prof. Brodsky presented his research on object-oriented databases and will submit a research proposal to EOSDIS at Ms. Moe's request.
- Professor Burt Edelson, Professor Nabil Adam, and I visited USRA's Goddard Visiting Scientist Program office to discuss with Dr. Bill Howard and Mr. David Holdridge the possibility of responding to the NSSDC solicitation.
- I attended the International Conference on "Women in Technology", and presented a lecture on "Challenges in Global Electronic Commerce".
- Ms. Lisa Singh, a Ph.D candidate in computer science from Northwestern University, visited CESDIS and presented a seminar on data warehousing.
- I attended the IBM CASCON conference in Toronto, Canada. I served as a workshop chair for four workshops on electronic commerce. At the invitation of the CEO and President of IBM Canada, Mr. John Whetmore, I served on the panel of distinguished women in information technology.
- I met with Ms. Martha Szczur, Chief Information Officer of Goddard (580), and discussed the opportunity for CESDIS to participate in new initiatives at Goddard in the information technology arena. She invited me to give a presentation to her management council about CESDIS research projects on January 28, 1998.

- My paper "Analytical Performance Modeling of Hierarchical Mass Storage Systems" (co-authored by Pentakalos, O., Menasce, D., Halem, M.), appeared in *IEEE Transactions on Computers*, October 1997, Vol. 46, No. 10, pp.1103-1119.
- I held a conference call with Dr. Paul Coleman, President of USRA, to finalize CESDIS' involvement in the Sofia project.
- I held a meeting with Dr. Milt Halem (Code 930), and Dr. Susan Hoban (UMBC/CESDIS), to discuss progress with the organization of the International Conference on Advances in Digital Libraries, to be held in Santa Barbara, California. I also accepted an invitation to give a plenary speech at the Workshop on Data Warehousing, sponsored by NASA, that will be held in conjunction with the ADL98 Conference.
- I traveled to New York and attended meetings with Mr. Uzia Galil, CEO of Elron, and a member of the executive board of the U. S.-Israel Science and Technology Commission, to discuss CESDIS' involvement in the information technology program.
- Dr. Paul Coleman visited CESDIS, and had a long discussion with me regarding the new initiatives at CESDIS.
- I held a two hour meeting with the Data Warehousing group from Code 931 to discuss CESDIS' involvement in this project.
- Professor Ouri Wolfson (University of Illinois) visited CESDIS, and worked with me on new initiatives in the area of digital libraries.
- I reviewed papers for two major international conferences, ACM SIGMOD '98 and Electronic Commerce 98. I am serving on the program committees of these conferences.
- I held a number of meetings with the data warehousing group to firm up the data warehousing project. Ms. Lisa Singh (Northwestern University) joined the project, and is making a significant contribution to it.
- CESDIS held a staff meeting to discuss the future of the CESDIS contract. At this point it looks like the CESDIS contract will be extended for 2 years.
- Professor and Chairman Dr. John Pinkston (Computer Science and Electrical Engineering Department/UMBC), visited CESDIS and met with a number of CESDIS scientists and with Dr. Milton Halem. The purpose of his visit was to strengthen the relationship between CESDIS and the Computer Science and Electrical Engineering Department at the University of Maryland Baltimore County.
- I gave a presentation to the management council of the Information Systems Directorate on CESDIS research projects. My presentation was extremely well received and several follow-up meetings were already scheduled.
- I met with Dr. Horace Mitchell (Code 930), and Professor David Ebert (Computer Science and Electrical Engineering Department /UMBC). The topic of the discussion was the creation of a joint program in information visualization.
- My Paper "Towards the Theory of Cost Management for Digital Libraries" was submitted to *ACM Transactions on Database Systems*.

- Professor Nabil Adam (Rutgers University) and I visited the Distributed Systems Laboratory supervised by Professor Yair Amir (Johns Hopkins University), and heard a status report on the work in digital libraries which is sponsored by CESDIS. Prof. Amir has made significant progress on his research in the last several years, and was able to publish his results in prestigious conference proceedings and journals. He has also been able to supplement his research funding by obtaining grants from DARPA.
- Dr. Miodrag Rancic visited UMBC. I introduced him to several faculty members in the Department of Computer Science and Electrical Engineering, and in the Physics Department. Dr. Rancic will develop a research team at UMBC with the primary research focus in the area of Earth Science.
- I had a meeting with Prof. George Lake (University of Washington) to discuss the possibility of Dr. Lake spending a sabbatical leave at CESDIS.
- Dr. Milton Halem (Code 930), Dr. Jacqueline Le Moigne (CESDIS), and I visited Ms. Kristi Brown of the Systems Engineering Division, and listened to a presentation on a new flight program. During the visit potential collaboration between CESDIS and Ms. Brown's group was discussed.
- The Director of the Hackensack (New Jersey) Meadowlands Development Commission and his research staff visited CESDIS.
- A significant amount of my time was dedicated to identifying data warehousing research issues as they relate to NASA storage needs. I held a number of research meetings with code 930 scientists in order to understand their requirements and challenges.
- I gave an invited lecture at the Johns Hopkins University and met with several faculty members, including Dean Westgate and Computer Science Department chair Dr. Jerry Mason.
- At the invitation of the program manager of the Advanced Technology Program at the National Institute of Standards and Technology, I participated in the ATP workshop on Challenges for Electronic Commerce. At this workshop I served as a panel member on "Information Overload and Filtering". This workshop was attended by industrial, academic, and government leaders in the area of information technology.
- CESDIS hosted Professor Vianu from the University of California, San Diego. Professor Vianu's research interest is in database theory. Dr. Vianu spent a day at CESDIS meeting with CESDIS scientists.
- I attended a USRA Board of Directors meeting and gave a talk on the present status of CESDIS. On March 27, 1998 I attended the USRA annual Council of Institutions meeting. On March 27, I had numerous meetings and discussions with other USRA program Directors and with Dr. Paul Coleman, USRA President.
- Professor Richard Somerville from Scripps Institution of Oceanography at the University of California, San Diego joined CESDIS as a consultant with a charter to build an Earth Science program at CESDIS.
- Dr. Bill Arms, Vice President of the Corporation for National Research Initiatives, gave an invited lecture at CESDIS. His talk was on standards for digital libraries. After the lecture, Dr. Milton Halem, Dr. Nabil Adam, Dr. Arms, and I held a meeting to discuss the future of digital libraries conferences and considered the possibility of having a joint ACM/IEEE conference.
- I met with Jeanne Behnke of EOSDIS (586), and discussed the possibility of CESDIS getting involved in the EOSDIS Data Warehousing project.

- I traveled to Santa Barbara, CA to attend the International Conference on Advances in Digital Libraries (ADL98). I gave a talk there on data warehousing and data mining issues as they apply to electronic commerce.
- I made significant progress in my research on developing cost models for digital libraries.
- I spent a significant amount of time meeting with the UMBC administrative staff to discuss the future of the collaboration between UMBC and CESDIS. I held a number of meetings with Dr. Freeman Hrabowski (UMBC President), Dr. JoAnn Argersinger (UMBC Provost), Dr. Shlomo Carmi (Dean of Engineering), and Dr. Pinkston (Chairman of Computer Science and Electrical Engineering).
- I visited George Mason University and gave an invited lecture. I met with a number of computer science faculty and learned about their research projects.
- I held a meeting with Mr. Howard Kea, principal engineer of the Information Systems Division (581). The topic of the meeting was the collaboration between CESDIS and the Information Systems Division.
- Mr. Pat Gary and Dr. Nand Lal (Code 930) joined Prof. Nabil Adam (Rutgers University) and me on our visit to Johns Hopkins University. The purpose of the visit was to evaluate the work performed by Prof. Yair Amir, and to explore the potential for future collaboration between CESDIS and Johns Hopkins University.
- I made significant progress in my research on wireless networks.
- I attended a Conference on Trends in Electronic Commerce in Hamburg, Germany. I also presented a paper (Strategies for Maximizing Sellers Profits Under Buyers Utility Values) and participated on the panel on Future Directions in Electronic Commerce. I held numerous meetings with the faculty at different European universities to discuss potential collaboration in the areas of distributed databases, electronic commerce, and digital libraries.
- I announced the appointment of Dr. Susan Hoban as Acting Associate Director of CESDIS.

## RESEARCH

### **Towards Free Information Markets** (with Baruch Awerbuch and Konstantinos Kalpakis)

In a multi-user environment, e.g., time-shared systems, Internet, etc., resources are traditionally managed based on concepts of "fairness" and rigid "priority" structure. As the demands exceed supply, performance degrades uniformly. In economic terms, this is the essence of a "communist" economy. Such an economic system is obviously the best if supply exceeds demand. Otherwise, a competitive open-market environment performs better.

We are currently living in the "communism" era, where payment for electronic services is very uncommon. However, we can observe that demand for resources (e.g., number of users on the Internet) grows exponentially while supply of resources (e.g., total bandwidth available) exhibits a much slower growth. This calls for a revision of our approach to research allocation on the Internet (or Global Information Infrastructure). More specifically, we call for the "information-perestroyka", namely transition from rigid central planning to open markets of electronic resources (computation, space, communication) and services (software, information). We argue that this is the next step that is crucial to reap fruit of the "information revolution".

## **Footprint Handover Rerouting Protocol for Low Earth Orbit Satellite Networks** (with Huseyin Uzunalioglu, Ian F. Akyildiz, and Wei Yen)

Low Earth Orbit (LEO) satellite networks will be an integral part of telecommunications infrastructures. In a LEO satellite network, satellites and their individual coverage areas move relative to a fixed observer on Earth. To ensure that ongoing calls are not disrupted as a result of satellite movement, calls should be transferred or handed over to new satellites. Since two satellites are involved in a satellite handover, the connection route should be modified to include the new satellite into the connection route. The route change can be achieved by augmenting the existing route with the new satellite or by completely rerouting the connection. Route augmentation is simple to implement, however the resulting route is not optimal. Complete rerouting achieves optimal routes at the expense of signaling overhead. In this paper, we introduce a handover rerouting protocol that maintains the optimality of the initial route without performing a routing algorithm after intersatellite handovers. The FHRP makes use of the footprints of the satellites in the initial route as the reference for rerouting. More specifically, after an optimum route has been determined during the call establishment process, the FHRP ensures that the new route due to handover is also optimum. The FHRP demands easy processing, signaling, and storage costs. The performance results show that the FHRP performs similar to a network without any handovers in terms of call blocking probability.

## **Pythia and Pythia/WK: Tools for the Performance Analysis of Mass Storage Systems** (with Odysseas I. Pentakalos and Daniel A. Menasce')

The constant growth in the demands imposed on hierarchical mass storage systems creates a need for frequent reconfiguration and upgrading to ensure that the response times and other performance metrics are within the desired service levels. This paper describes the design and operation of two tools: Pythia and Pythia/WK, that assist system managers and integrators in making cost-effective procurement decisions. Pythia automatically builds and solves an analytic model of a mass storage system based on a graphical description of the architecture of the system, and on a description of the workload imposed on the system. The use of a modeling wizard to perform this conversion from a graphical description of a mass storage system to an analytic model makes Pythia unique among analytic performance tools. Pythia/WK uses clustering algorithms to characterize the workload from the log files of the mass storage system. The resulting workload characterization is used as input to Pythia.

## **Analytical Performance Modeling of Hierarchical Mass Storage Systems** (with Odysseas I. Pentakalos, Daniel A. Menasce', and Milton Halem)

Mass storage systems are finding greater use in scientific computing research environments for retrieving and archiving the large volumes of data generated and manipulated by scientific computations. This paper presents a queueing network model that can be used to carry out capacity planning studies of hierarchical mass storage systems. Measurements taken on a Unitree mass storage system and a detailed workload characterization provided by the workload intensity and resource demand parameters for the various types of read and write requests. The performance model developed here is based on approximations to multi-class Mean Value Analysis of queueing networks. The approximations were validated through the use of discrete event simulation and the complete model was validated and calibrated through measurements. The resulting model was used to analyze three different scenarios: effect of workload intensity increase, use of file compression at the server and client, and use of file abstractions.

## **Towards the theory of cost management for Digital Libraries** (with Ouri Wolfson)

One of the features that distinguishes digital libraries from traditional databases is new cost models for client-access to intellectual property. Clients will pay for accessing data items in digital libraries, and we believe that optimizing these costs will be as important as optimizing performance in traditional databases. In this paper we discuss cost models and protocols for accessing digital libraries, with the objective of determining the minimum cost protocol for each model.

We expect that in the future information appliances will come equipped with a cost optimizer, in the same way that today computers come with a built-in operating system. This paper makes the initial steps towards a theory and practice of intellectual property cost management.

## **Electronic Commerce: Current Limitations and Future Visions**

In this paper we examine the key technologies behind electronic commerce (EC). Our emphasis is on the current limitations that exist in EC and the potential which might exist should these barriers be overcome. We adopt here the "can do" attitude, perhaps better said the "must-do" attitude, which assumes that vendors and buyers along with the financial power and interests that they possess *will* drive technology steadily towards an "ideal" EC system; one that is: global, fully interoperable across all sites, industry-independent, useable, and efficient. We acknowledge, however, that even with a "can do" attitude, the dramatic technological advances in EC outlined in this paper will take significant amounts of time as well as financial and human investment.

## **Relational Transducers for Electronic Commerce**

Electronic commerce is emerging as one of the major Web-supported applications requiring database support. We introduce and study high-level declarative specifications of business models, using an approach in the spirit of active databases. More precisely, business models are specified as *relational transducers* that map sequences of input relations into sequences of output relations. The semantically meaningful trace of an input-output exchange is kept as a sequence of *log* relations. We consider problems motivated by electronic commerce applications, such as log validation, verifying temporal properties of transducers, and comparing two relational transducers. Positive results are obtained for a restricted class of relational transducers called *Spocus transducers* (for semi-positive outputs and cumulative state). We argue that despite the restrictions, these capture a wide range of practically significant business models.

## **Evolving Databases: An Application to Electronic Commerce**

Many complex and dynamic database applications (such as product modeling and negotiation monitoring) require a number of features that have been adopted in semantic models and databases such as active rules, constraints, inheritance, etc. Unfortunately, such features have mostly been considered in isolation. Furthermore, participants in a commercial negotiation, staking their financial well-being, will accept a system only if they can gain a precise behavioral understanding of it. In this paper, we propose a rich and extensible database model, *evolving databases*, with clear and precise semantics based on *evolving algebras*. We also briefly describe a prototype implementation of the model and a preliminary validation of the prototype with electronic commerce applications.

## **Media Access Control Protocols for Multimedia Traffic in Wireless Networks** (with Ian F. Akyildiz and Ramon Puigjaner)

This paper presents a survey on Media Access Control (MAC) protocols for multimedia traffic in wireless networks. The MAC protocols covered in this paper include classical as well as recently proposed schemes intended for use in a single hop, TDMA-based or CDMA-based wireless system. The operation of each protocol is explained and its advantages and disadvantages are presented. A comparison between the MAC protocols is given. The activities of Standard Committees are reviewed.

## **SELECTED PUBLICATIONS**

### **Books**

Adam, N.R., Dogramaci, O., Gangopadhyay, A., and Yesha, Y. (1998). *Electronic Commerce: Technical, Business, and Legal Issues*. Upper Saddle River, NJ: Prentice-Hall PTR.

### **Publications in referred journals**

Yesha, Y., Awerbuch, B., and Kalpakis, K. (1997). Towards Free Information Markets. Accepted by *Mathematical Modeling and Scientific Computing*.

Yesha, Y., Pentakalos, O., and Menascé, D. (1997). Pythia and Pythia/WK: Tools for the Performance Analysis of Mass Storage Systems. *Software Practice and Experience*, 27(9)

Yesha, Y., Pentakalos, O., Menascé, D. and Halem, M. (1997). Analytical Performance Modeling of Hierarchical Mass Storage Systems. *IEEE Transactions on Computers*, 46(12)

Yesha, Y., Uzunalioglu, H., Akyildiz, I.F., Yen, W. (1998). Footprint Handover Rerouting Protocol for Low Earth Orbit Satellite Networks. To appear in *ACM-Baltzer Journal of Wireless Networks*.

Yesha, Y., Akyildiz, I.F., and Puigjaner, R. (1998). Media Access Control Protocols for Multimedia Traffic in Wireless Networks. Submitted for publication.

Yesha, Y. Electronic Commerce: Current Limitations and Future Visions. To appear in *IEEE Transactions in Knowledge and Data Engineering*.

Yesha, Y. and Wolfson, O. Towards the theory of cost management for Digital Libraries. *ACM Transactions on Database Systems*.

### **Publications in proceedings**

Yesha, Y., Pentakalos, O., and Menascé, D. (1997). Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems. *PNPM/Tools Conference*. San Malo, France.

Yesha, Y. (1997). Evolving Databases: An Application to Electronic Commerce. *Proceedings of the International Database Engineering and Application Symposium*. Montreal, Canada.

Yesha, Y. (1998). EDI As a Distributed Information Systems. *Proceedings of the Hawaii International Conference on Systems Sciences*.

Yesha, Y. (1998). Strategies for Maximizing Sellers Profits Under Unknown Buyers Utility Values. *Proceedings of the International Conference on Trends in Electronic Commerce TREC98*. Hamburg, Germany.

Yesha, Y., Abiteboul, S., Vianu, V., and Fordham, B. (1998). Relational Transducers for Electronic Commerce. *Proceedings of the 17th ACM SIGACT-SIGMOD-SIGACT Symposium on Principle of Database Systems*, pp. 178-188, June 1-3, Seattle, Washington.

---

---

# ***CONSULTANTS TO THE DIRECTOR***

---

Task 1 on the CESDIS contract (the general administrative task) allows the Director to bring to CESDIS consultants who are not funded by specific task originators. CESDIS entered into agreements with the individuals reported upon in this section for the purpose of program development.

**Ian Akyildiz**

Georgia Institute of Technology,  
Broadband and Wireless Networking Laboratory

**Burt Edelson**

George Washington University,  
Department of Electrical Engineering and Computer Science

**Richard Somerville**

University of California, San Diego  
Scripps Institution of Oceanography

**Ouri Wolfson**

University of Illinois, Chicago  
Department of Electrical Engineering and Computer Science

**Ian F. Akyildiz**  
**Georgia Institute of Technology**  
**Broadband and Wireless Networking Laboratory**  
**(ian@ee.gatech.edu)**

Low Earth Orbit (LEO) satellite networks will be an integral part of telecommunications infrastructures. In an LEO satellite network, satellites and their individual coverage areas move relative to a fixed observer on Earth. To ensure that ongoing calls are not disrupted as a result of satellite movement, calls should be transferred or handed over to new satellites. Since two satellites are involved in a satellite handover, a connection route should be modified to include the new satellite into the connection route. The route change can be achieved by augmenting the existing route with the new satellite or by completely rerouting the connection. Route augmentation is simple to implement, however the resulting route is not optimal. Complete rerouting achieves optimal routes at the expense of signaling overhead.

We finished a report [1] where we introduced a handover rerouting protocol that maintains the optimality of the initial route without performing a routing algorithm after intersatellite handovers. The FHRP makes use of the footprints of the satellites in the initial route as the reference for rerouting. More specifically, after an optimum route has been determined during the call establishment process, the FHRP ensures that the new route due to handover is also optimum. The FHRP demands easy processing, signaling, and storage costs. The performance results show that the FHRP performs similar to a network without any handovers in terms of call blocking probability.

In [2] we developed a long survey paper presenting all Media Access Control (MAC) protocols for multimedia traffic in wireless networks. The MAC protocols covered in [2] include classical as well as recently proposed schemes intended for use in a single hop, TDMA-based or CDMA-based wireless system. The operation of each protocol is explained and its advantages and disadvantages are presented. A comparison between the MAC protocols is given. The activities of Standard Committees are reviewed.

## References

[1] Uzunalioglu, H., Akyildiz, I. F., Yesha, Y., and Yen, W. (1998). Footprint Handover Rerouting Protocol for Low Earth Orbit Satellite Networks. *ACM-Baltzer Journal of Wireless Networks*.

[2] Akyildiz, I. F., Puigjaner, R., and Yesha, Y. (1998). Media Access Control Protocols for Multimedia Traffic in Wireless Networks. Submitted for publication.

**Burton I. Edelson**  
**George Washington University**  
**Department of Electrical Engineering and Computer Science**  
**(edelson@seas.gwu.edu)**

## Goals

Provide expertise to CESDIS in satellite communications and high performance networking; and plan and organize CESDIS cooperative projects with NASA, other U. S. government agencies, U. S. industry, and, where appropriate, foreign research organizations.

## Activities

1. Provided technical expertise in satellite communications to NASA GSFC. Led effort to get NASA and CESDIS involved in satellite communications as part of the ACTS experiments and G-7 Information Society programs. Worked with Pat Gary (930) to arrange for and conduct several high data rate transmission tests of the ACTS terminal and Goddard linked to other ACTS terminals at LeRC, JPL, Hawaii, and the Magic Test Bed.
2. Planned and obtained FY97 and FY98 funding for the Testbed for Space and Terrestrial Interoperability (TSTI) to test the capability and develop procedures for satellite and optical fiber to be inter-connected in high data rate networks. This testbed utilizes ACTS and other satellites and the ATDNet to test and develop transmission and networking procedures, standards, protocols, and equipment necessary to interconnect networks at data rates of 45, 155, and 622 Mb/s. Worked with Pat Gary to develop and instrument the TSTI and conduct tests.
3. Worked with Pat Gary, Susan Hoban, and Neil Helm (GWU) on arranging a digital libraries experiment to connect U. S. data archives at the Library of Congress, National Library of Medicine, Department of Agriculture, and NASA GLOBE data center with corresponding data centers in Japan.
4. Continued work led by Milt Halem and supported by Yelena Yesha, Susan Hoban, and others from GSFC and CESDIS to develop and expand the Global Legal Information Network (GLIN) project with the Law Library of Congress. Wrote plan for developing GLIN Intranet. Worked with Pat Gary and Neil Helm to prepare specifications for procurement of two small satcom terminals; one to be installed at NASA Goddard and the other at a remote GLIN station, likely in South America.
5. Served on executive panel with group of experts from government, industry and universities to conduct a worldwide survey of satellite communications sponsored by NASA and NSF. Helped write, review, and edit extensive report on "Global Satellite Communications Technology and Systems". (To be published in summer 1998).
6. Worked with Sam Venneri and Ramon DePaula (NASA HQ) and visited NASA centers including LeRC, ARC, and JPL, to plan inter-center coordination and cooperation in ACTS experiments and high performance networking.

## Conferences and Workshops

Japan-U. S. Science Technology and Space Applications Program workshop - Hawaii, November 1997

AIAA International Communications Satellite Systems Conference, G-7 GII Quadrilateral Satellite Working Group meeting, and Satellite Communications for the GII workshop - all held in Yokohama, Japan, February 1998

Grand Challenges for Space workshop - USRA-NIAC, Columbia, MD May 1998

Satellite Networks: Architectures, Applications, and Technologies workshop - NASA LeRC, Cleveland, June 1998

## Publications

Hyde, G., and Edelson, B. I. (1997). Laser Satcom offers Radio Links in Space. *Aerospace America*. 26-29.

Bargellini, P. L., Hyde, G., and Edelson, B. I. (1997). The Future of Communications Satellite Systems. *Proceedings of the Third Ka-Band Utilization Conference*. Sorrento, Italy.

Bergman, L., Edelson, B. I. et al. (1997). Distributed HDV Post-Production over Trans-Pacific ATM Satellites. *Proceedings of the Third Ka-Band Utilization Conference*. Sorrento, Italy.

Edelson, B. I. and Helm, N. R. (1997). High Data Rate Satellite Communications: Interoperability Issues (IAF-97-M.1.09). *Proceedings of the 48th International Astronautical Congress*. Turin, Italy.

Helm, N. R. and Edelson, B. I. (1997). Space Technologies and Systems for Disaster Mitigation, (IAF-97.C.2.01). *Proceedings of the 48th International Astronautical Congress*. Turin, Italy.

**Richard Somerville**  
**University of California, San Diego**  
**Scripps Institution of Oceanography**  
**(rsomerville@ucsd.edu)**

During the period ending June 30, 1998, I have consulted on the following tasks at the request of Dr. Milton Halem, Chief, Earth and Space Data Computing Division, Code 930.

## **1. Four-dimensional Data Assimilation of Satellite Remote Sensing Data for Meteorological Modeling**

Together with a post-doctoral researcher, Dr. Halem and I have revisited the classical problem of determining the extent to which satellite observations can contribute to an optimal estimate of atmospheric initial conditions for numerical weather prediction. We have used a numerical model of the atmosphere with idealized data assimilation techniques to assess the potential of satellite remote sensing data in this application. This research revisits a classical problem first explored by Charney, Halem and Jastrow in the early days of four-dimensional data assimilation, namely the ability of time-dependent wind observations to substitute for a lack of temperature information. Because of differing dynamical relationships between the mass and motion fields in the tropics and the extra-tropics, there is a latitudinal dependence of the results. In our most recent work, I have devised diagnostic procedures for exploring the way in which the model ingests the wind information. In future work, I plan to help implement more modern techniques for assimilation, in lieu of the simple data replacement method used thus far.

## **2. Utilization of Full-disk Satellite Imagery From the Proposed Triana Mission for Operational Numerical Weather Prediction and for Research Purposes**

The Triana proposal is for a relatively fine time-resolution, but relatively coarse space-resolution, visible image of the sunlit side of the Earth. This image would be provided by a camera parked at the L1 point between Earth and Sun, which is approximately 4 times as far from the Earth as the orbit of the Moon. This suggestion, due to Vice President Gore, is primarily motivated by educational concerns, but it has given rise to an examination of the scientific utility of such a mission. I have considered the meteorological value of Triana data, focusing on the possibility of utilizing cloud-track winds at polar latitudes, which are invisible from geostationary orbit. My tentative conclusion is that such information would be of marginal or even negligible value to operational numerical weather prediction but might be useful for research purposes. If the Triana mission could be expanded beyond a simple visible imager, however, then its scientific

value could be greatly enhanced. In particular, if Earth radiation budget parameters could be monitored essentially continuously from the L1 vantagepoint at high time-resolution, then Triana could do much more to help unravel the role of clouds in modulating climate variability.

### **3. Planning for Digital Earth**

Digital Earth, as defined by Vice President Gore, refers to the concept of a virtual planet Earth, composed of a large number of independent data sets. These data sets are to be linked by a high-speed system incorporating multi-dimensional display capabilities and a rich set of query and browse functionality. The hardware and software challenges posed by such a concept are daunting. I was heavily involved with Dr. Halem in the planning of a multi-agency workshop on Digital Earth, which was successfully held at GSFC on June 23 and 24, 1998. With Dr. Halem, I helped plan the workshop agenda, and I drafted written materials, which were sent to the invitees. I attended the workshop and made a presentation there. Because the Digital Earth is conceived as a tool to be used in science museums and similar settings, I have urged that specialists in science education and in presenting science to lay audiences be involved early in the planning stages. I have also suggested that a concrete demonstration or feasibility project could be carried out this year. Such a project ought to explore the path of implementing Digital Earth incrementally and to test the look and feel of the visualization capability. Digital Earth has great potential as a research tool as well as a public outreach vehicle. I have recently been considering its applicability to short-term climate prediction via combining heterogeneous data sets, e.g., crop data and hydrologic information.

### **4. Planning for Future Atmospheric Science Research in CESDIS and the Earth and Space Data Computing Division**

In continuing discussions with Dr. Halem and Dr. Yesha, I am working toward the long-term goal of building an in-house research capability in atmospheric science, comparable to the existing ones in space science and computer science. This goal includes the development of a strong collaborative research program with university scientists, but it also will require hiring to increase the in-house expertise. Most recently, I have concentrated on helping to recruit a suitable Ph. D. atmospheric scientist to join the group at Goddard.

**Ouri Wolfson**  
**University of Illinois, Chicago**  
**Department of Electrical Engineering and Computer Science**  
**wolfson@ouri.eecs.uic.edu**

### **Statement of Work**

Dr. Wolfson was tasked with developing cost models and protocols for accessing digital libraries, with the objective of determining the minimum cost protocol.

### **Results**

In general there are two basic business models for information providers. One is advertiser paid and the other is customer paid. In existing media both models coexist, e.g., newspapers and cable tv. Based on this we predict that a similar coexistence will occur in the future digital libraries. In this project we addressed the issue of cost management/optimization in the customer-paid business model.

Some specific digital library applications of our work include various forms of electronic news services, such as stock trading and electronic mail services. In the case of stock trading, an object is information (including price) of a particular stock or a group of stocks. In other cases the object may be a more complex data structure such as a view or a queue. For example, the object may be a queue of news items that satisfy a particular filter, or it may be an electronic mail box.

Other applications of our work include data warehousing and cache management on the web. In data warehouses, which maintain views on data from various sources, the views need to be kept up to date by getting new data. A new version here is not a complete new copy, but an incremental change from the previous one. Subscription versus Demand correspond to the terminology of "push" versus "pull" used in this context. The results also apply for cache management on the Web.

We introduced complexity measures and analyzed retrieval protocols for two cost models: the request cost model and the time cost model. In these cost models we considered the Subscription, Demand, Demand with cache invalidation, and Sliding Window protocols. These protocols can be employed by a client to access an object at the digital library server. The first two protocols are static in the sense that an object is either cached or isn't; the last two protocols are dynamic in the sense that an object may be cached at some time, and not cached at another time. The protocols are different in the two cost models, and they also vary depending on whether or not each read of the object must be consistent, i.e., access the latest version of the object.

It is important to emphasize that the set of cost models and protocols considered in this project is far from being exhaustive. Many other scenarios are conceivable, and this project should be regarded as a demonstration of our proposed approach to the problem of cost management in accessing digital libraries. For the rest of this section we summarize the results of our analysis.

First consider an object accessed in the request model using consistent reads. Assume that at any point in time, the probability is  $q$  that the next relevant access of the object is a write at the server (thus the probability is  $1 - q$  that the next relevant access of the object is a read at the client). If  $q$  is fixed and known a priori, then the protocol that has the optimal expected cost depends on the costs of a read  $rc$ , a write  $wc$ , and an invalidation notification  $ic$ . These results are summarized in a theorem. If  $q$  is unknown or it varies over time, then the Demand and Subscription static protocols are suboptimal. For the dynamic protocols, the average expected cost results are summarized in a theorem. If the relevant requests are chaotic (i.e., do not follow a probabilistic pattern) and the objective is to reduce the worst-case cost, then again one of the dynamic protocols is optimal; the one with the lowest competitive ratio can be computed based on  $rc$ ,  $wc$  and  $ic$  using two theorems.

Now consider an object accessed in the time cost model using consistent reads. Here the problem is to select between Subscription and Demand (possibly with cache invalidation) for each time slot; in contrast to the request cost model, the switch between the two protocols cannot occur in the middle of a slot, only at time-slot boundaries. This gives rise to a totally new set of concerns. The first problem is to determine the protocol with minimum expected cost for each time slot, assuming that we are given the number of expected relevant requests in a time slot. We devise an efficient algorithm that determines the optimal policy for each time slot, such that the average cost per time slot is minimized. We also devise the Sliding Window algorithm for this model. Cache invalidation is combined with the Demand protocol in a straightforward manner. Again, the issues are totally different than in the request cost model.

Finally, we consider an object in the request cost model using tolerant (or inconsistent) reads, i.e., reads that can tolerate an out-of-date version of the object. It turns out that straightforward use of Subscription and Demand cannot take advantage of such reads in order to reduce cost. Therefore, for this environment we proposed a hybrid mechanism between Subscription and Demand. In the previous scenarios, at any point in time the client is either on Subscription (and pays for the writes) or Demand (and pays for the reads); it may switch between Subscription and Demand periodically. In contrast, in the request cost model using tolerant reads, at any point in time the client is on Subscription for some reads and on Demand for

other reads, depending on the tolerance of the read. The client pays for the Demand-reads, and for some of the writes. We called this the Static Divergence Caching (SDC). The first problem that we solved in this model is determining (for given probabilities of the relevant requests) the optimal refresh rate of SDC, i.e., optimal lower bound on the tolerance of the Subscription reads. For the case where the probabilities of the relevant requests are unknown or they vary over time, we devised the Sliding Window algorithm for this model, called the Dynamic Divergence Caching (DDC). We showed that for optimizing cost in the worst case, the DDC algorithm is better than SDC. Finally we analyzed the DDC and SDC algorithms by simulations. We showed that tolerant reads improve the cost compared with nontolerant reads by a factor of two. We also showed that when the relevant probabilities are fixed but unknown, the cost of the DDC algorithm is almost as good as that of SDC with the optimal refresh rate. On the other hand, when the relevant probabilities vary over time, the cost of the DDC algorithm is 70% of the cost of the SDC algorithm having the optimal refresh rate.

We believe that in the future information appliances will come equipped with a cost optimizer, in the same way that computers today come with a built-in operating system. Similarly, customer agents searching for information may be equipped with similar optimizers. This project makes the initial steps towards a theory and practice of cost management and optimization in accessing information. Such a theory and its implications may become critical for the information economies of the future.

## Publications

- [1] Wolfson, O., Jajodia, S., and Huang, Y. (1997). An Adaptive Data Replication Algorithm. *ACM Transactions on Database Systems (TODS)*, 22(2), 255-314.
- [2] Wolfson, O. (1997). Data Management in Mobile Computing. *ACM/Baltzer Journal on Special Topics in Mobile Networks and Applications (MONET)*, 2(2). (guest editor's introduction).
- [3] Wolfson, O. and Huang, Y. (1998). Competitive Analysis of Caching in Distributed Databases. *IEEE Transactions on Parallel and Distributed Systems*, 9(4), 391-409.
- [4] Sistla, P., Wolfson, O., and Huang, Y. (1998). Minimization of Communication Cost Through Caching in Mobile Environments. *IEEE Transactions on Parallel and Distributed Systems*, 9(4), 378-390.
- [5] Pitoura, E., Bhargava, B., and Wolfson, O. Data Consistency in Intermittently Connected Distributed Systems. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*.
- [6] Tayeb, J., Ulusoy, O., and Wolfson, O. A Quadtree Based Dynamic Attribute Indexing Method. *Computer Journal*.
- [7] Sistla, P., Wolfson, O., Yesha, Y., and Sloan, R. Towards a Theory of Cost Management for Digital Libraries. *ACM Transactions on Database Systems (TODS)*.
- [8] Wolfson, O., Chamberlain, S., Dao, S., and Jiang, L. (1997). Location Management in Moving Objects Databases. *Proceedings of The Second International Workshop on Satellite-Based Information Services (WOSBIS'97)*, pp. 7-14. Budapest, Hungary.
- [9] Wolfson, O. (1997). Location Management for Moving Objects Databases. *Proceedings of The First Intensive Workshop on Spatio-Temporal Database Systems*. Austria.
- [10] Wolfson, O., Chamberlain, S., Dao, S., Jiang, L. and Mendez, G. (1998). Cost and Imprecision in Modeling the Position of Moving Objects. *Proceedings of the Fourteenth International Conference on Data Engineering (ICDE14)*, pp. 588-596. Orlando, FL.

[11] Wolfson, O., Xu, B., Chamberlain, S., and Jiang, L. (1998). Challenges and Approaches in Motion Databases. *Proceedings of the 14th International Conference on Advanced Science and Technology (ICAST98)*, pp. 182-194. Naperville, IL.

[12] Wolfson, O., Xu, B., Chamberlain, S., and Jiang, L. (1998). Moving Objects Databases: Issues and Solutions. *Proceedings of the 10th International Conference on Scientific and Statistical Database Management (SSDBM98)*, Capri (Italy).

[13] Sistla, P., Wolfson, O., Chamberlain, S., and Dao, S. (1998). Querying the Uncertain Position of Moving Objects. In O. Etzion, S. Jajodia, and S. Sripada (eds.), *Temporal Databases: Research and Practice*. Springer-Verlag.

[14] Rishe, Naphtali, Naboulsi, Khaled, and Wolfson, Ouri. (1998). Report Generators. *Wiley Encyclopedia of Electrical and Electronics Engineering*, John Webster (ed.).

---

---

# ***COMPUTATIONAL SCIENCES BRANCH***

---

## *Remote Sensing Group*

**Jacqueline LeMoigne**, Senior Scientist - Branch Head  
**Richard Lyon**, University of Maryland Baltimore County  
**Timothy Murphy**, University of Maryland Baltimore County  
**Nathan Netanyahu**, University of Maryland College Park

**Tarek El-Ghazawi**, George Mason University  
**Jules Kouatchou**, George Washington University

## *Scalable Systems Technology Group*

**Phillip Merkey**, Senior Scientist  
**Terrence Pratt**, Senior Scientist  
**Donald Becker**, Staff Scientist  
**Erik Hendriks**, Technical Specialist

**Udaya Ranawake**, University of Maryland Baltimore County

HPCC Earth and Space Science Project Scientist: **George Lake** (University of Washington)  
**Adam Frank**, University of Rochester

## ***An Evaluation of Automatic Image Registration Methods***

**Jacqueline Le Moigne, Senior Scientist-Branch Head  
(lemoigne@nibbles.gsfc.nasa.gov)**

### **Profile**

Dr. Le Moigne holds three degrees from the University of Paris VI, Paris, France: a Bachelor of Science in theoretical mathematics, a Master of Science in pattern recognition, and a Ph.D. in computer vision (1983). From 1983-1987 she served as a research scientist in the University of Maryland College Park, Center for Automation Research, Computer Vision Laboratory. She directed new software development for the Autonomous Land Vehicle project and studied a range sensor utilizing the principle of structured light by projection of grids.

From 1988-1990, Dr. Le Moigne worked as a scientist with Martin Marietta Laboratories. In this capacity she conducted research on the fusion of regions and edges by relaxation methods and studied texture analysis methods for safe Mars landings.

After two years as a National Academy of Sciences-National Research Council Senior Resident Research Associate with the Goddard Space Data and Computing Division (Code 930), Dr. Le Moigne joined CESDIS in October 1992 as a staff scientist. She was appointed to the position of Computational Sciences Branch Head in January 1995 and was promoted to senior scientist in June 1995. Professional memberships include the IEEE Geoscience and Remote Sensing Society for which she has been Chairman and Vice Chairman of the Washington/Northern Virginia Chapter and to which she was elected as senior member in 1996.

Dr. Le Moigne's current work involves the multi-sensor registration, fusion, and analysis of remotely sensed data. This research is of interest in many Earth science applications, such as GOES data landmark registration, the assessment of forested areas utilizing AVHRR and Landsat-TM data, as well as the validation and calibration of new sensor data (such as Modis) with already known data such as Landsat-TM data. This research is also very important for automatic multi-sensor integration when data is gathered at far-remote sites, such as for Mars exploration. All of the techniques involved in this research, especially wavelet-based image registration, have been developed as parallel algorithms on the MasPar MP-2.

### **Report**

In Collaboration with: Wei Xia (Global Science Technology), James C. Tilton (NASA/GSFC, Code 935), Tarek El-Ghazawi and Prachya Chalermwat (George Mason University), Nathan Netanyahu and David Mount (University of Maryland College Park).

### **Abstract**

The study of global environmental changes involves the comparison, fusion, and integration of multiple types of remotely-sensed data at various temporal, radiometric, and spatial resolutions. Results of this integration may be utilized for global change analysis, as well as for the validation of new instruments or for new data analysis. Furthermore, smaller missions will include many different sensors carried on separate platforms, and the amount of remote sensing data to be combined will increase tremendously. For all of these applications, the first required step is fast and automatic image registration.

As the need for automating registration techniques is being recognized, it becomes necessary to survey all the registration methods which may be applicable to Earth and space science problems and to evaluate their performances on a large variety of existing remote sensing data as well as on simulated data of soon-to-be-flown instruments. In this report we present the first steps toward an exhaustive quantitative evaluation: four automatic image registration algorithms are described and results of their evaluation are presented on three different datasets. The four algorithms are based on gray levels, edge features or wavelet features and compute translation, similarity or rigid transformations. Results show that the four selected methods are within 2 pixel accuracy, and that a trade-off must be achieved between computation time and accuracy of the computed deformation.

## 1. Introduction

Automatic registration and resampling of remotely-sensed data will be an essential element of future Earth satellite observation systems. New remote sensing systems will generate enormous amounts of data representing multiple observations of the same features at different times and/or by different sensors. Also, with the new trend of smaller missions, these sensors will be spread over multiple platforms. Automatic registration and resampling methods will become indispensable for such tasks as data fusion, navigation, achieving super-resolution, or optimizing communication rates between spacecraft and ground systems. Although automatic image registration has been extensively studied in other areas of image processing, it is still a complex problem in the framework of remote sensing. When images are acquired either by the same sensor at different times or by two sensors at the same or different times, a number of distortions prevent the two images from being "perfectly registered" to each other or to a fixed coordinate system. It is very difficult to determine exact location within an image using only ancillary data. Distortions usually correspond to orbit and attitude anomalies, but some continuous nonlinear distortions are also due to altitude, velocity, yaw, pitch, and roll. To investigate the best ways of dealing with these issues, we feel that there is a need to survey all registration methods which may be applicable to Earth Science problems and to evaluate their performances on a large variety of existing remote sensing data as well as on simulated data of soon-to-be-flown instruments.

Data registration can be defined as the process which determines the best match of two or more images acquired at the same or different times by different or identical sensors. One set of data is taken as the *reference data*, and all other data, called *input data* (or *sensed data*), is matched relative to the reference data. The general process of image registration includes three main steps: (1) the extraction of features to be used in the matching process, (2) the feature matching strategy and metrics, and (3) the resampling of the data based on the correspondence computed from matched features. Currently, the most common approach to registration is to perform step (1) manually by interactive extraction of a few outstanding characteristics of the data, which are called *control points* (CP's), *tie-points*, or *reference points*. The CP's in both images (or image and map) are matched by pair and used to compute the parameters of a geometric transformation. But such a point selection represents a repetitive, labor- and time-intensive task which becomes prohibitive for large amounts of data, and often leads to large registration errors [1].

This report focuses on steps (1) and (2), and describes four different methods for the automatic image registration of satellite imagery and the first results of their quantitative intercomparison. The methods described below utilize gray levels, edge points, or wavelet features as the features used in the matching process. For feature matching, we have been looking at multi-resolution strategies, correlation measures, and statistically robust techniques. For ease of use, these four methods have been integrated within an operational toolbox (briefly described in section 3.3 and in more details in [2]). We also implemented a semi-manual registration method which is being used as reference.

The algorithm intercomparison is based on the criteria described in section 3.1; mainly accuracy and timing results are reported here. The four algorithms are tested on three datasets described in section 3.2. Results are given in section 3.4.

## 2. Automatic Image Registration Algorithms

As was described in section 1, the two main steps of a registration algorithm are:

1. the extraction of features to be used in the matching process, and
2. the feature matching strategy and metrics.

According to Brown [3], step (2) can be described further as the combination of three components:

- (2.1) a *search space*, i.e., the class of potential transformations (or deformation models) that establish the correspondence between input data and reference data (e.g., rigid, affine, or polynomial transformations). As a first approximation, we consider that the distortions corresponding to orbit and attitude anomalies correspond mainly to an affine transformation, and that a few other small continuous nonlinear distortions due to altitude, velocity, yaw, pitch, and roll will be handled by global or local polynomial or elastic transformations of a higher degree.
- (2.2) a *search strategy*, which is used to choose which transformations have to be computed and evaluated. Search strategies are usually chosen to reduce the amount of computations. Hierarchical or multi-resolution techniques as well as tree or graph matching are all examples of search strategies. Other examples can be found in [3]
- (2.3) a *similarity metric*, which evaluates the match between input and transformed reference data for a given transformation chosen in the search space. Correlation (or cross-correlation) measurement is the usual similarity metric, although it is computationally expensive and noise sensitive when used on original gray level data. Using a pre-processing technique such as edge detection or a multi-resolution search strategy enables large reductions in computing time and increases the robustness of the algorithms. Other similarity metrics include the sum of absolute differences [4] or Hausdorff distances [5,6].

More extensive surveys on automatic image registration methods can be found in Brown [3], Le Moigne [7], Fonseca [8], Rignot [9], and Lester [10].

This large choice of techniques which could be utilized for the registration of remote sensing images leads us into performing their quantitative evaluation in the framework of remote sensing imagery. Given the variety of sensors, data, and applications, we anticipate that no single registration technique will satisfy all different data and applications. Furthermore, when looking at this large variety of techniques, although automated registration has been developed for a few Earth science applications, there is no general scheme which would assist users in the selection of a registration tool. The goal of this project is to provide the potential user with some guidelines on the choice of the registration technique, which would depend on such parameters as the type of sensor, the desired registration accuracy, the computer availability, or the speed requirement.

In this study, we implemented four automatic algorithms along with a semi-manual tool which is used as reference. The four automatic methods are correlation-based methods, using either gray levels or features. We assume the transformation to be either a rigid or an affine transformation. Since both types of transformations include compositions of translations and rotations (or "similarity" transformations), our preliminary search space is composed of translations, similarities, or rigid transformations. The first two automatic tools, "Spatial Correlation" and "Phase Correlation," are well-known methods based on the correlation of gray level intensities or edge intensities of the full size image in the spatial or in the frequency domain respectively. For both of these methods, the transformation is assumed to be a translation. The other two automatic tools, "Iterative Edge Matching" and "Wavelet-Based Registration," are new feature-based methods developed by the authors, for which the transformation is assumed to be either a rigid or a similarity transformation respectively. Since features are more reliable than intensity or radiometric values,

feature-based methods are usually more accurate than intensity-based methods. In particular, edge or edge-like features such as wavelet features (both of which have been chosen for these two methods) are very useful to highlight regions of interest such as coastlines without being sensitive to local variations of intensity [11].

The detailed description of the algorithms which are evaluated in this study follows:

### **Semi-Manual Registration**

The semi-manual tool is similar to the method most commonly utilized for registration; a human operator interactively selects *Ground Control Points (GCP's)* in two images, and these points become the input to compute the deformation model between the two datasets, often chosen as a polynomial transformation. In our implementation, the user selects GCP's from the displayed reference image and the image to be registered respectively. Zoom capabilities are available to help users choose the GCP's more accurately. Then a choice of transformations is provided to the user: rotation, translation (e.g., shift), rigid, affine, and polynomial transformations. The GCP's are then used to calculate the parameters of the chosen transformation: either the rotation angle, or the translation shifts, or the transformation coefficients for rigid, affine, or polynomial transformations.

### **Spatial Correlation**

With spatial correlation, the input image is shifted over a search grid and multiplied times the reference image. The search grid location that produces a maximum from the image multiplication is taken as the best amount of shift for registration.

The sharpest correlation peaks are usually obtained when the input and reference images are edge images. If input and reference images are edge images, a refinement of the initially detected image shift is obtained by expanding the images by a factor of four, thinning the edges in the expanded images, and performing another search over a grid centered around the previously detected registration location. This refinement process allows for the detection of the image shift location to a quarter pixel resolution. We should note that while this refinement process can produce excellent results for edge images, it will generally produce unpredictable results for other types of images.

Any of the many available edge detectors can produce edge images suitable for use with the spatial correlation method. One promising edge detection method is the "difference recursive filter" edge detector by Shen and Castan [12,13]. This edge detector uses an Infinite Symmetric Exponential Filter (ISEF), an optimal low-pass filter, for smoothing images prior to edge detection. A symmetric exponential filter can be written as:

$$f(x) = a * b^{|x|}$$

where, for the discrete case,  $b = (1-a)/(1+a)$  and  $0 < a < 1$  (implying  $0 < b < 1$ ).

To detect edges after performing the ISEF filter, an adaptive gradient is calculated at the zero crossings of the second derivative of the ISEF smoothed image, and the edges are detected through thresholding.

Suitable edge images may also be created from the region boundaries detected by an image segmentation approach, such as the hierarchical image segmentation developed by Tilton [14]. This approach alternates between region growing and spectral clustering. The region growing process controls the segmentation process and sets the threshold for spectral clustering, which itself is not allowed to merge spatially adjacent regions. This approach also finds natural segmentation convergence points by detecting significant jumps in the ratio of the dissimilarity criterion from one iteration to the next.

## Phase Correlation

Phase correlation is a mathematical technique that was developed to register images in which the misregistration is only a translation. The technique can be described as follows (from [15]): given a reference image,  $g_R$  and a sensed image  $g_S$ , with 2-D Fourier transforms  $G_R$  and  $G_S$ , respectively, the cross-power spectrum of the two images is defined as  $G_R G_S^*$  and the phase of that spectrum as:

$$e^{j\phi} = \frac{G_R G_S^*}{|G G^*|}$$

The phase correlation function,  $d$ , is given by the inverse Fourier transform of that spectrum:

$$d = F^{-1}\{e^{j\phi}\}$$

It is easily proven that the spatial location of the peak value of the phase correlation function,  $d$ , corresponds to the translation misregistration between  $g_R$  and  $g_S$ . An innovation employed in our implementation of phase correlation is in the finding of the peak of the correlation function,  $d$ . Instead of looking for an interpolated peak of  $d$ , the center of mass of the peak of  $d$  is found. We have found that this gives a more robust result than searching explicitly for the peak.

## Iterative Edge Matching

This edge-based method performs the registration in an iterative manner, first estimating the parameters of the deformation transformation on the center of the images, and then iteratively refining these parameters in larger and larger portions of the images. An edge detection computes the gradient of the original gray levels and highlights the pixels of the images with higher contrast. Currently, Sobel edges [16] are extracted in both reference and input images. Any of the edge detection methods described in the above "Spatial Correlation" section could be substituted for the Sobel edge detector.

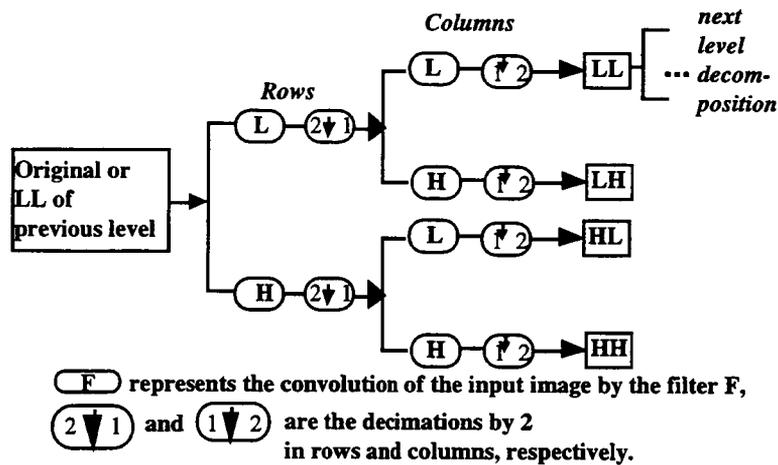
The Iterative Edge Matching method is based on the idea that registration parameters are usually well-known around the nadir or center point of the images but deteriorate for pixels considered away from the centers. For this implementation, we chose to model the transformation as a rigid transformation, e.g., as the combination of a scaling in both directions ( $ds_x, ds_y$ ), a rotation ( $d\theta$ ), and a shift or translation ( $dt_x, dt_y$ ) in both directions. This algorithm assumes that scaling parameters are small (within [0.9, 1.1]). At each iteration, the five parameters are retrieved by computing the cross-correlation measures for all successive values of the parameters taken at incremental steps. The algorithm can then be described as three successive iterations:

1. Using the 64x64 centers of both reference and input edge images, the best similarity function is first computed by maximizing the cross-correlation function for all successive values of rotation and translation. After transformation by this first approximation of  $\{\theta, tx, ty\}$ , the scaling factors in both directions are computed.
2. The same process is iterated on the portions of the images centered at the center of the full image, and of size  $(N+64)/2$  rows by  $(M+64)/2$  columns. But instead of searching the entire transformation space, only values of the parameters in small intervals around the previous values are considered and the previous values of the parameters are refined.
3. An identical search is performed using the full size images but searching a very small subspace of transformations centered around the approximations computed in 2.

The advantage of this iterative search is a reduction in computation time when compared to a complete search in the full size images. Potential problems might occur when the center of one or two of the images are covered by clouds or contain too much unreliable data; a potential improvement would include the detection of such conditions and the extraction of preferable "windows" located as close to the centers as possible.

## Wavelet Maxima Matching

Wavelet transforms provide a space-frequency representation of an image. In a wavelet representation, the original signal is filtered by the translations and the dilations of a basic function, called the "mother wavelet". In this wavelet-based registration, only discrete orthonormal bases of wavelets have been considered and are implemented by filtering the original image by a high-pass and a low-pass filter, thus in a multi-resolution fashion [17]. At each level of decomposition, four new images are computed; each of these images is a quarter the size of the previous original image and represents the low frequency or high frequency information of the image in the horizontal and/or the vertical directions; images LL (Low/Low), LH (Low/High), HL(High/Low), and HH (High/High). Starting again from the "compressed" image (or image representing the low-frequency information, "LL"), the process can be iterated, thus building a hierarchy of lower and lower resolution images. Figure 1 summarizes the multi-resolution decomposition.



**Figure 1: Multi-Resolution Wavelet Decomposition**

Our wavelet-based method represents a three-step approach to automatic registration of remote sensing imagery. The first step involves the wavelet decomposition of the reference and input images to be registered. In the second step, we extract at each level of decomposition domain independent features from both reference and input images. Finally, we utilize these features to compute the transformation function by following the multiresolution approach provided by the wavelet decomposition. Features are either chosen as the gray levels provided by the low-frequency LL compressed versions of the original image (for non-noisy images), or are based on the high-frequency information (e.g., maxima points of LH and HL images) extracted from the wavelet decomposition. In this second option, only those points whose intensities belong to the top x% of the histograms of these images are kept (x being a parameter of the program whose selection can be automatic, usually x=10%); we call these points "maxima of the wavelet coefficients." The related study reported in [18] shows that although high-pass coefficients are less sensitive to noise, the high-pass subbands are less robust to translation invariance than the low-pass subband.

The search is performed iteratively from the deepest level of decomposition (where the image size is the smallest), until the first top level of decomposition. At each level, the transformation is found with an accuracy D and is refined at the next level up with an accuracy D/2. More details on this algorithm can be found in [7, 11, 19, 20].

## Robust Point Pattern Matching

For both edge-based and wavelet-based registrations described previously, global cross-correlation of the feature points has been utilized. But more generally the fundamental problem of point matching is defined as follows: given two pairs of points, find the (affine) transformation that transforms one point set so that

the distance from the other point set is minimized. Because of measurement errors and the presence of outlying data points, it is important that the distance measure between the two point sets be robust to these effects. In this section, distances are measured using the partial Hausdorff distance [6].

Point matching can be a computationally intensive task, and there have been a number of theoretical and applied approaches proposed for solving this problem. In this study, we present the results of two algorithmic approaches to the point matching problem, in an attempt to reduce its computational complexity, while still providing guarantees on the quality of the final match. Our first method is an approximation algorithm, which is loosely based on a branch-and-bound approach due to Huttenlocher and Rucklidge [21,22]. We show that by varying the approximation error bounds, it is possible to achieve a trade-off between the quality of the match and the running time of the algorithm. Our second method involves a Monte Carlo method for accelerating the search process used in the first algorithm. This algorithm operates within the framework of a branch-and-bound procedure, but employs point-to-point alignments to accelerate the search. We show that this combination retains many of the strengths of the branch-and-bound search, but provides significantly faster search times by exploiting alignments. With high probability, this method succeeds in finding an approximately optimal match. For more details on this method, see [6].

### 3. Algorithm Intercomparison

Defining intercomparison criteria is a relatively difficult task, since each application might have different requirements and the importance of the criteria might vary from one application to the next. The main criterion on which our results are focusing is the "accuracy" measurement. Different methods can be thought of to quantify the accuracy of a given registration method. In this study, the "true" transformation is known for the first two datasets and manual registration is utilized as "relative" ground truth for the third dataset. The four automatic registration algorithms described previously are applied to the three datasets. Tables 1 and 2 show the results of the evaluation. Table 3 shows some partial results of the robust statistical method for sets of points obtained from wavelet processing of one image at four decomposition levels.

Computational requirement is another criterion for evaluation. As one measurement of this criterion, Table 4 includes the timings of each method when run on a SunUltra 1 Model 170E.

#### 3.1 Definition of the Criteria

Although, for the current evaluation, only accuracy and timing criteria were considered, this section presents a description of all the criteria which could be considered in such an evaluation.

- Accuracy

Several methods can be thought of to quantify the accuracy of a given registration method:

1. a first method consists of registering the same set of data manually and automatically. Then, considering the manual registration as our "ground truth", the error between manual and automatic registration characterizes the accuracy of the automatic registration. This method is reliable as long as a large number of well-distributed control points can be chosen throughout the images, which is not always possible.
2. another method requires a processing which corrects for the illumination variations from two scenes. This correction would be applied after transforming back the sensed image by the computed deformation model. Then, a Mean Square Error (MSE) would be computed between transformed sensed image and reference image.
3. if the registration is performed between a remotely-sensed image and a map, the MSE can be computed on selected ground features such as coastlines; similarly, if the registration is per-

formed between two images, image segmentations of reference image and transformed input image can be compared for computing an accuracy measurement.

4. another way to quantify the accuracy of an automatic method is described in [23]; it utilizes high-resolution data such as Landsat-TM or SPOT ("Satellite Pour l'Observation de la Terre") data which are degraded to lower spatial resolution. Then the lower resolution data are registered and accuracy can be measured at a subpixel level using the full high-resolution data.
5. finally, simulated data can be created where all navigation parameters as well as cloud and atmospheric conditions can be controlled with great accuracy. Although such data might not include all possible radiometric variations occurring in real data, such simulated data might be very useful in comparing several algorithms under similar conditions.

- **Computational Requirements**

The computational requirements of each method can be computed from two means:

1. the computational complexity of each algorithm is evaluated
2. each method is implemented and timed on various architectures.

- **Level of Automatization**

Given the future large amounts of data to process, automatic techniques should be as free of parameters to tune as possible. Whenever possible, thresholds or other such parameters should be computed adaptively from within the programs. If necessary, training on large numbers of data is performed and parameters are chosen from this training.

- **Applicability**

This last criterion intuitively corresponds to qualitative judgments, such as "if the scene includes a city grid, a corner-based method will work faster than a region-based method." A quantitative way to evaluate the "applicability" criterion might be statistical; a large amount of sensor data over a large variety of scenes is gathered and the results of the three previous criteria are combined to compute a probability of the applicability of an automatic registration technique given a particular dataset and particular scene contents.

### 3.2 Test Datasets

The four previous algorithms have been evaluated on three datasets. For the first two datasets, the true transformation parameters are known. For the third dataset, no ground truth is available but manual registration measurements have been gathered and results of this manual registration have been visually evaluated using a map of the coastlines.

- The Girl dataset represents a 512X512 image of a human face artificially translated and rotated. Figure 2 shows the original image as well as five transformed images, by rotation, translation, or a combination of the two.
- The TM dataset represents a 512x512 image extracted from Band 2 of a Landsat-Thematic Mapper (TM) scene over the Pacific Northwest, with artificial translations and rotations. Figure 3 shows the original image as well as seven transformed images.
- The AVHRR dataset represents a series of thirteen 512 rows by 1024 columns AVHRR/LAC images over South Africa. Raw AVHRR data are navigated and georeferenced to a geographic grid that extends from -30.20 S, 15.39 E (upper left) -34.79 S, 24.59 E (lower right). The navigation



Figure 2: First dataset - Original Image and Five Transformed Images

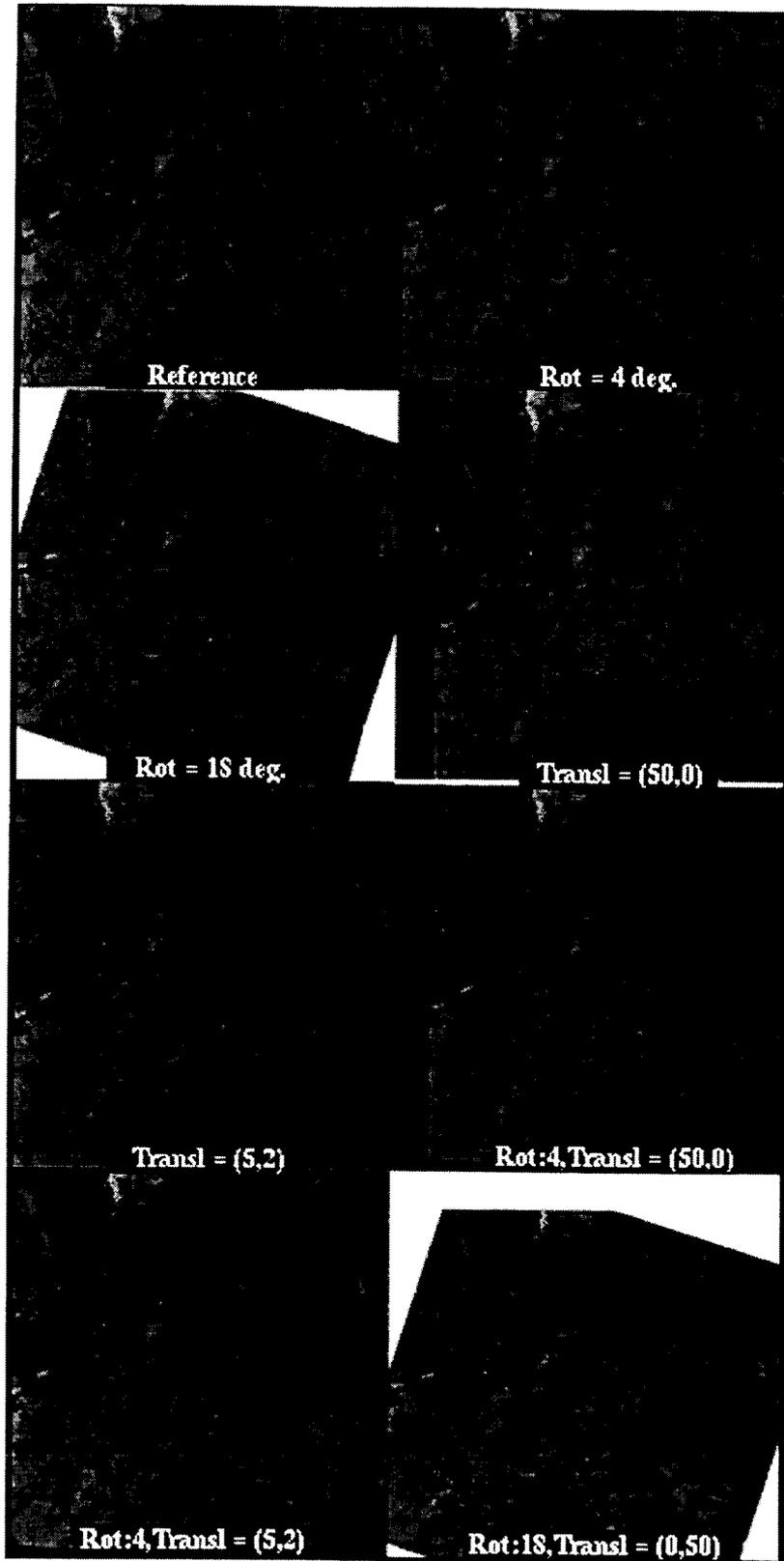


Figure 3: Second dataset - Original Image (Extracted from Band 2 of a Landsat/TM Scene over the Pacific Northwest) and Seven Transformed Images

process uses an orbital model developed at the University of Colorado [24] and assumes a mean attitude behavior (roll, pitch and yaw) derived using Ground Control Points [25]. A map of the coastline derived from the Digital Chart of the World (DCW) is generated for the same geographic grid. Figure 4 shows the map of the coastline as well as the thirteen multi-temporal AVHRR images. Figure 5 shows one image of this sequence ("avhrr\_sa1488") superimposed with the map of the coastlines. Note that in this case, there is a slight misregistration between map and image.

Eventually, all registration algorithms will be evaluated on simulated data as well as on a large variety of NASA datasets, which will represent at least three main types of applications:

- *Multi-temporal studies* with multi-temporal datasets of one sensor over the same areas collected at different times (various times of the day, various seasons, multiple years, etc.),
- *Multi-Instrument data fusion* with multi-sensor datasets representing multiple spatial, temporal, and radiometric resolutions,
- *Channel-to-channel co-registration* with multiple radiometric and spatial resolutions of the different channels of one given sensor; for example, a hyperspectral instrument.

### 3.3 Algorithm Implementation

We have chosen the Khoros environment as the framework for the implementation of these techniques. Khoros is an object-based data analysis, data visualization, and application development environment. In Khoros, a "toolbox" is a collection of programs and libraries that is handled as a single object. In that sense, our registration toolkit is also composed of the various registration routines, each of which can be handled as an object. Such a Khoros registration software is compatible with the software developed by the Applied Information Sciences Branch at NASA/Goddard Space Flight Center for the Regional Application Centers (RAC's) program [26]. The RAC's receive remote sensing data by direct readout from various satellites and their users utilize this software to process in real-time data needed for their regional applications (e.g., monitoring regional change, storm prediction, etc.). Such applications of our registration algorithms will provide us with feedback from the remote sensing community. From this feedback, future new algorithms may be developed which will be more adapted to specific applications. Figure 6 shows the top level of the Khoros graphic user interface of the current registration toolbox and Figure 7 shows the user interface when one of the registration methods is selected, in this case the iterative edge matching.

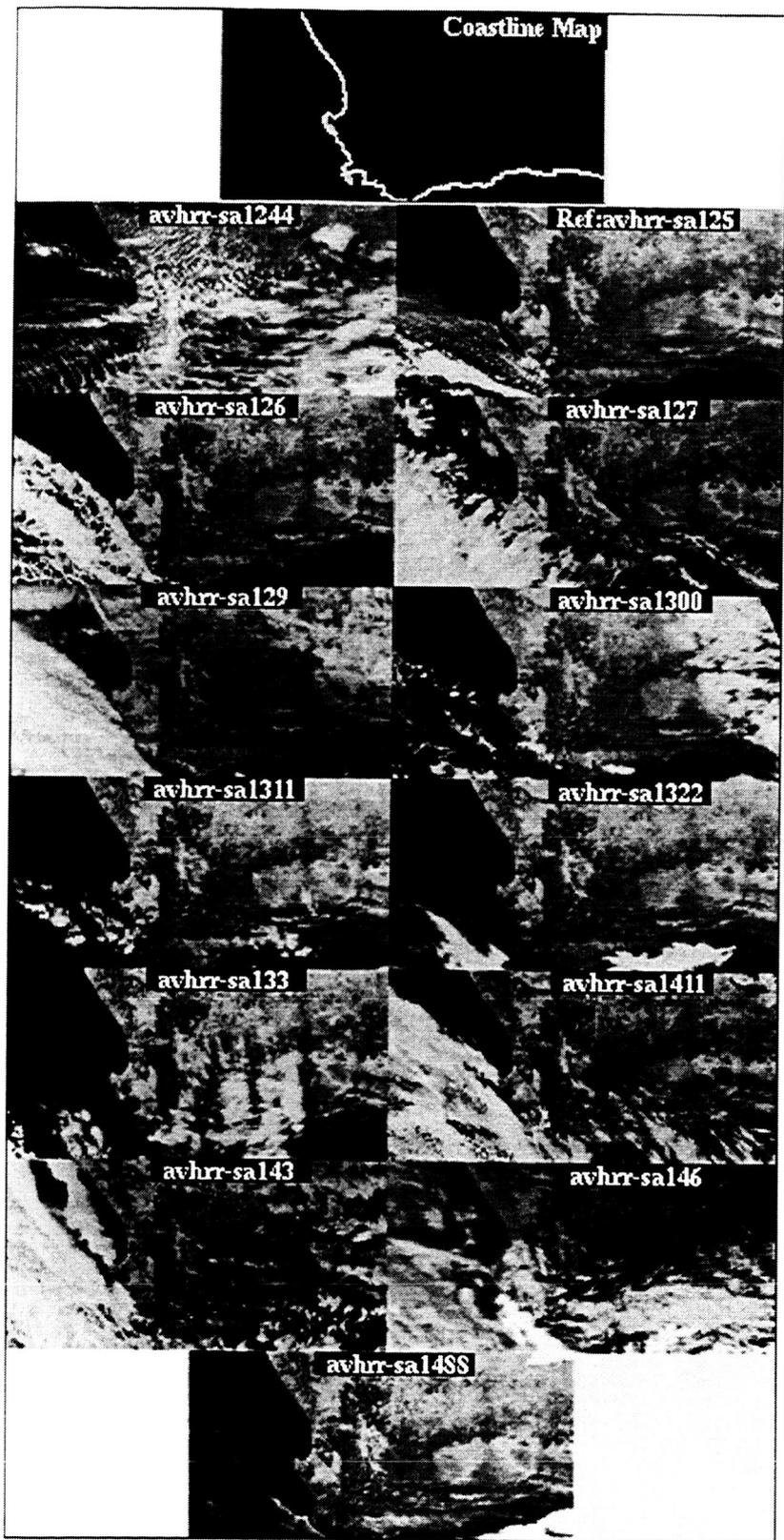


Figure 4: Third dataset - Coastline Map and Thirteen Images of a Multi-Temporal Series of AVHRR-LAC Band 2 Images over South Africa



Figure 5: Coastlines Superimposed on One of the AVHRR Images, "avhrr\_sa12488", with Zooming on Several Areas of the Coastline

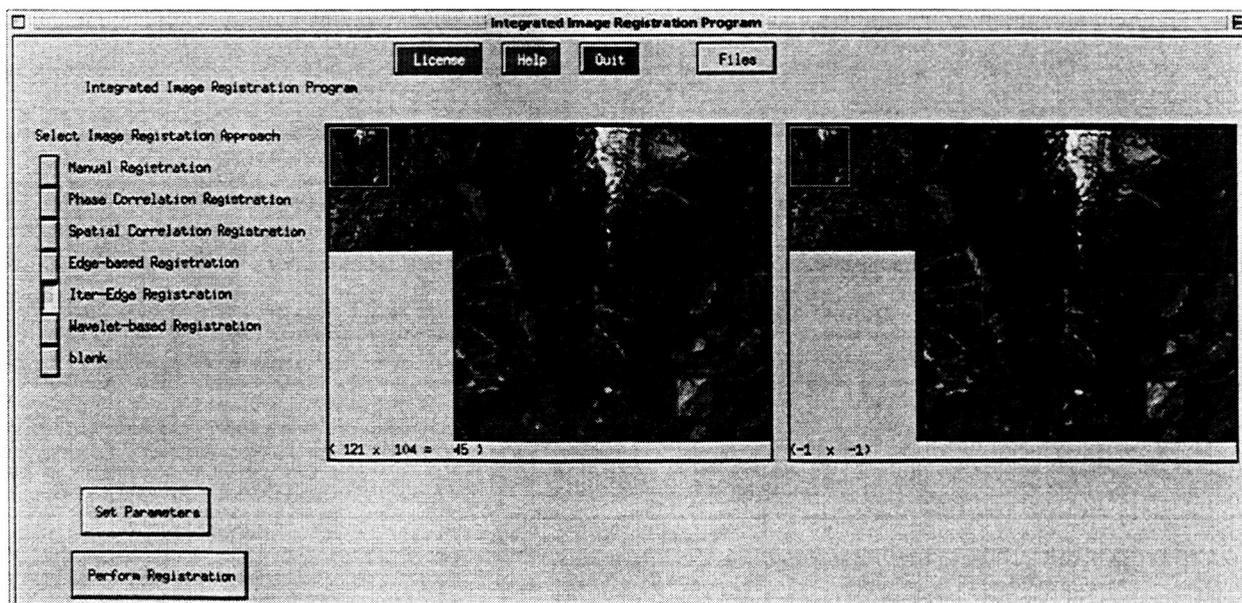


Figure 6: Khoros Graphic User Interface of the Image Registration Toolbox

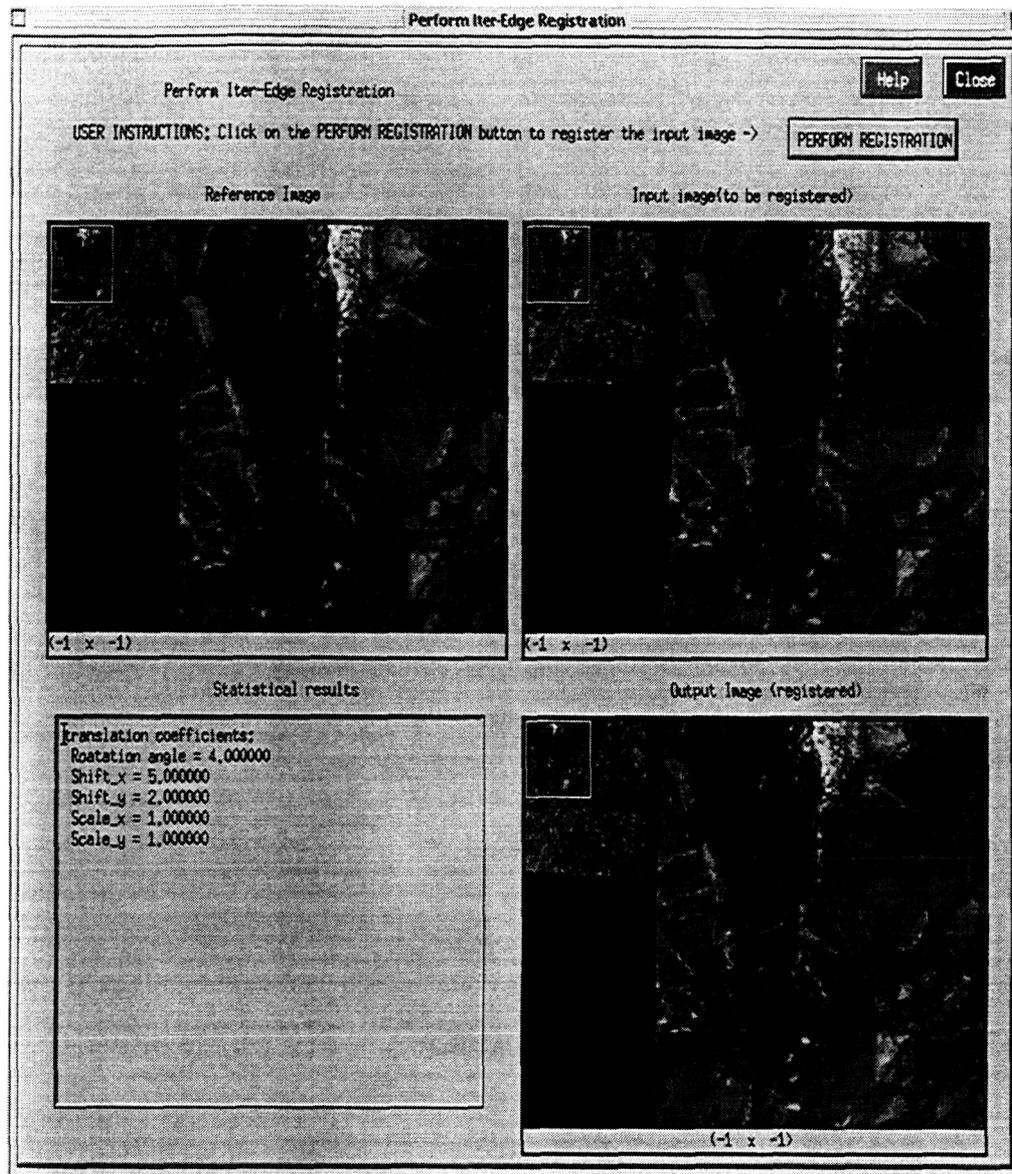


Figure 7: Khoros Graphic User Interface When Iterative Edge Matching is Selected

### 3.4 Results of the Evaluation

Table 1 shows the results of applying the four algorithms on the first two datasets, "Girl" and "TM," for which the true transformation is known. The Spatial and the Phase Correlation methods, since they are only computing translations, have been applied solely to the shifted images with no rotation. The Iterative Edge Matching computes a rotation angle (in degrees), translation components in x and y (in pixels), and scale components in x and y. The Wavelet Maxima Matching computes rotation angle and translation components. The two last methods have been applied to all transformed images. Wavelet matching can be computed using either the LL or the LH/HL coefficients. For these two first datasets, the reported results are obtained with LL coefficients since they are more accurate, especially for large translations. We notice also that the combination of large rotation angles with large translation components (e.g., "Girl.r0tx20ty60" and "TM.r18ty50") may result in a degraded accuracy. But generally, the edge- and wavelet-based methods provide a 100% accuracy for at least 8 of the 12 test data.

DATA		GIRL					TM						
Rotation	5	55	5	5	0	18	18	4	4	4	0	0	
Translation	(0,0)	(0,0)	(20,60)	(6,4)	(20,60)	(0,0)	(0,50)	(0,0)	(50,0)	(5,2)	(50,0)	(5,2)	
Scale	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	
<b>METHODS</b>													
<b>SPATIAL CORRELATION</b>						<b>SPATIAL CORRELATION</b>							
(Only) Translation (pixels)	-	-	-	-	(20,61)	-	-	-	-	-	(50,0)	(5,2)	
<b>PHASE CORRELATION</b>						<b>PHASE CORRELATION</b>							
(Only) Translation (pixels)	-	-	-	-	(20,60)	-	-	-	-	-	(50,0)	(5,2)	
<b>ITER EDGE MATCHING</b>						<b>ITER EDGE MATCHING</b>							
Rotation (degrees)	5	55	5	5	3	18	18	4	4	4	-1	0	
Translation (pixels)	(0,0)	(0,0)	(20,60)	(6,4)	(19,61)	(0,1)	(0,51)	(0,0)	(50,0)	(5,2)	(49,0)	(5,2)	
Scale	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	
<b>WAVELET COEFF (LL)</b>						<b>WAVELET COEFF (LL)</b>							
Rotation (degrees)	5	55	5	5	0	18	23	4	4	4	0	0	
Translation (pixels)	(0,0)	(0,0)	(20,60)	(6,4)	(20,60)	(0,0)	(0,50)	(0,0)	(50,0)	(4,2)	(50,0)	(6,2)	

Table 1: Results of the Four Algorithms on the First Two Datasets, "Girl" and "TM"

AVHRR DATA	sa1244	sa126	sa127	sa129	sa1300	sa1311	sa1322	sa133	sa1411	sa143	sa146	sa1488	
<b>MANUAL REGISTRATION</b>													
Rotation	0	0	0	0	0	0	0	0	0	0	0	0	Average Difference In Translation Autom. vs. Manual
Translation	(1,0)	(0,0)	(-1,-1)	(-3,-1)	(2,0)	(1,0)	(0,-1)	(0,-1)	(0,-1)	(-1,-3)	(-3,-5)	(2,3)	
Scale	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	
<b>METHODS</b>													
<b>SPATIAL CORRELATION</b>													
(Only) Translation (pixels)	(1,0)	(0,-1)	(-2,-1)	(-4,-1)	(2,0)	(1,-1)	(0,-1)	(0,-1)	(0,-2)	(-2,-3)	(-9,0)	(3,3)	1.54
<b>PHASE CORRELATION</b>													
(Only) Translation (pixels)	(1,0)	(0.5,-1.5)	(-2,-1.5)	(-6,0.5)	(2,-0.5)	(2.5,-1.5)	(1,-1)	(0,-1)	(0.5,-2.5)	(-2,3.5)	(-4,-5.5)	(3.5,2.5)	1.50
<b>ITER EDGE MATCHING</b>													
Rotation (degrees)	1	0	0	0	0	0	0	0	0	0	0	0	1.17
Translation (pixels)	(3,-4)	(0,0)	(-2,-1)	(-2,-1)	(2,0)	(2,-1)	(0,-1)	(0,-1)	(0,-2)	(-2,-3)	(-4,-5)	(3,3)	
Scale	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	
<b>WAVELET COEFF (LH/HL)</b>													
Rotation (degrees)	0	0	0	-1	0	0	0	0	0	0	-2	0	1.66
Translation (pixels)	(0,0)	(0,0)	(0,-2)	(-6,-2)	(4,0)	(2,0)	(0,-2)	(0,-2)	(0,-2)	(0,-4)	(-6,-6)	(2,4)	

Table 2: Results of the Four Algorithms on the Third Dataset, "AVHRR"

Table 2 shows the results of the four algorithms applied to the third dataset, "AVHRR." Since no registration ground truth is available for these images, all data have been manually registered, assuming only a translation transformation (rotation=0 degrees, scaling=1.0). Assuming only a small translation (between -5 and +5 pixels), all algorithms have been applied to all images in the "AVHRR" dataset. The results of the manual registration have been verified by superimposing the binary map of the coastlines onto the respectively shifted images. Some examples of these results are shown in Figure 8 by zooming in on a few areas for two different images. Most of the results obtained by manual registration are verified as accurate by the coastline map. But some of the data, especially for very cloudy images such as "AVHRR\_sa1244,\_sa129,\_sa143,\_sa146", manual results do not always match the coastline map and cannot really be considered as "ground truth." In the following, we will consider these manual results as "references" rather than "ground truth data." Figure 8 also shows some examples of the four automatic registrations superimposed with the coastlines. For this dataset, the wavelet matching is reported as applied with the LH/HL coefficients; since the data are more noisy, gray level correlation of the LL coefficients are not as accurate as the high frequency information. Also, the observed translations are small enough for the LH/HL coefficients to be reliable. Results reported in Table 2 generally show that for no or few clouds (i.e., high signal to noise ratio), all algorithms behave similarly. The differences occur when the level of clouds increases, for which Phase Correlation and Iterative Edge Matching seem to be more robust. In general, when comparing automatic and manual registrations, all results are within 2 pixels accuracy, with final results of 1.54 for Spatial Correlation, 1.50 for Phase Correlation, 1.17 for Iterative Edge Matching, and 1.66 for Wavelet Matching. The last result on wavelet-based registration is understood by the translation non-invariance property of the wavelets. This issue can be addressed by looking at other types of wavelets, by combining the information of the different subbands [18] and by applying a robust matching instead of an exhaustive search (see [6] and Table 3).

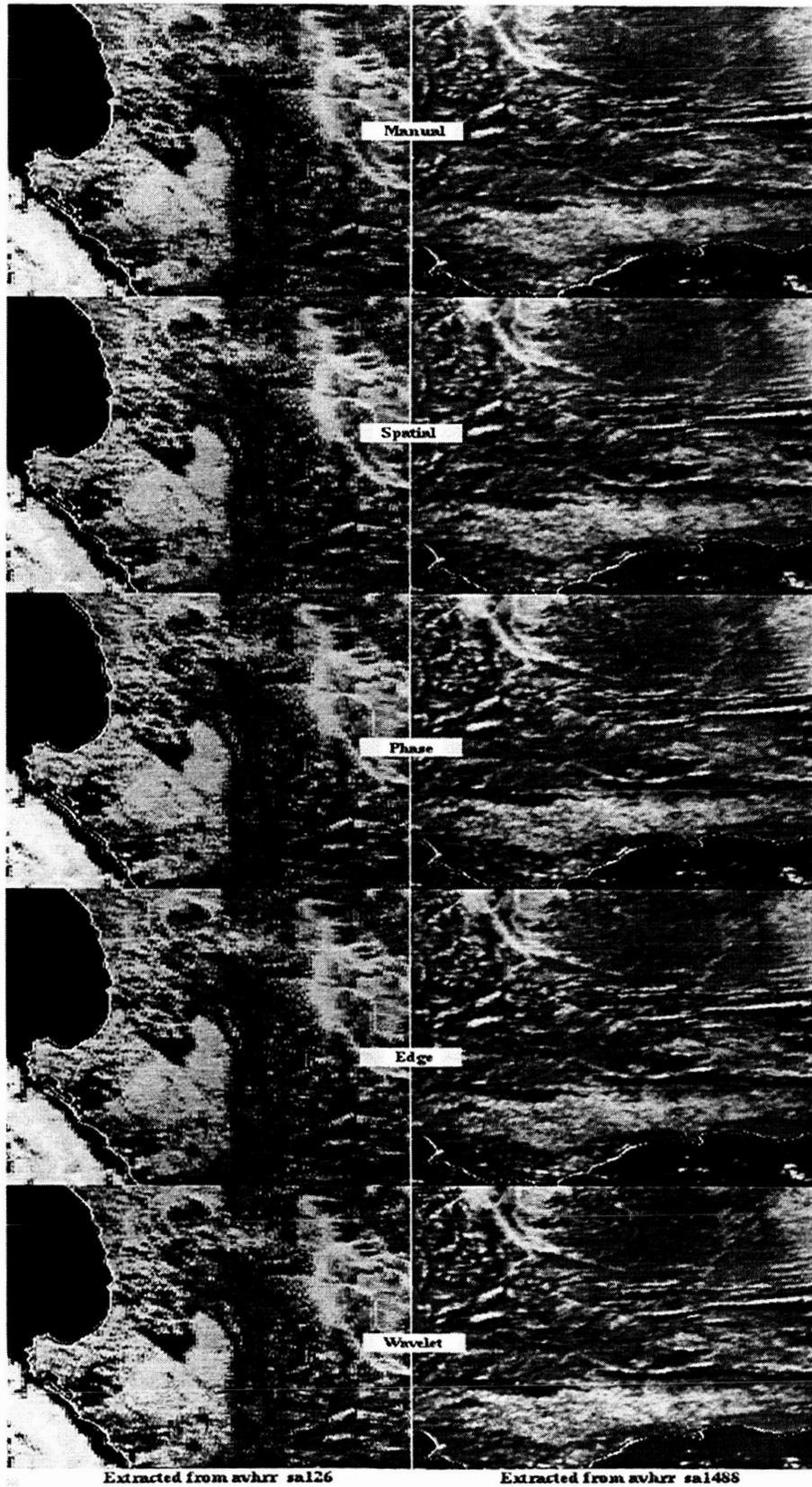


Figure 8: Zoom on Coastlines Transformed by Experimental Results and Superimposed for Two of the AVHRR Images, "avhrr\_sa126, avhrr\_sa1488"

GIRL.R4TK5TY2	Level 3 (28x28)	Level 2 (56x56)	Level 1 (112x112)	Level 0 (224x224)	Original (448x448)
Rotation (degrees)	4	4	4	4	4
Translation (pixels)	(0,0)	(0,0)	(1,0)	(2,1)	(5,2)
Scale	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)
<b>Branch &amp; Bound</b>					<b>Interpolation</b>
Rotation (degrees)	3.6	4.26	4.02	3.97	3.97
Translation (pixels)	(0.4,0.13)	(0.87,-0.05)	(1.19,0.34)	(2.33,0.62)	(4.66,1.24)
Timings (seconds)	16.7	1.47	15.2	63.6	Total= 96.97
<b>Bounded Alignment</b>					<b>Interpolation</b>
Rotation (degrees)	4.8	4.2	3.68	4.03	4.03
Translation (pixels)	(0.3,-0.12)	(0.42,0.33)	(1.17,0.17)	(2.16,0.91)	(4.32,1.82)
Timings (seconds)	2.4	0.06	1.48	2.57	Total= 6.51

**Table 3: Preliminary Results of Robust Matching Applied to the Registration of Wavelet Features Extracted from Images "TM" and "TM.r4tx5ty2" (cf. Figure 5)**

Table 3 represents some preliminary results of the Robust Point Pattern Matching applied to maxima feature points extracted from the LH wavelet subbands of the 4-level decomposition of a 448x448 image extracted from the second dataset "TM.r4tx5ty2" image (rotation=4 degrees, Translation in x=5 pixels, Translation in y=2 pixels). This matching is not yet integrated with the other automatic algorithms, but the results simulate the iterative multi-resolution search described by the wavelet-based method in section 2. Since the exhaustive search employed for wavelet- and iterative edge-matching is time-consuming and subject to getting trapped into local optima, it is replaced here by the use of the more robust partial Hausdorff distance. Table 3 shows very encouraging results for the two methods described in section 2, "Branch & Bound (BB)" and "Bounded Alignment (BA)." The two algorithms show comparable accuracies, but the second method, BA, shows faster computation times. An improvement of this method would be to simultaneously utilize all decomposition levels.

Table 4 shows timings for all four algorithms on a SunUltra 1 Model 170E. In the previous examples, because of border effects, both Spatial and Phase Correlation were only computed on a 256x256 window extracted at the center of the images. For homogeneity reasons, Spatial and Phase Correlations were also timed when run on larger images, and the wavelet-based method was timed when only computing a translation, using either LL or LH/HL coefficients. Since the wavelet-based method performs the correlation in an iterative fashion, it is the most computationally efficient, especially when using only the one LL sub-band. These results also show that the wavelet-based registration is faster for small images up to 512x512. But the computational requirements of this method grow more rapidly than those of the iterative edge matching. Therefore, for 1024x512 images, edge-matching becomes faster than the wavelet-LH/HL method.

TIMINGS (SECONDS)			
Image Size	256x256	512x512	1024x512
<b>Method</b>			
<b>Computing Only Translation</b>			
Spatial Corr.-Translation	14.14	60.57	129.62*
Phase Corr.-Translation	6.16	23.69	45.48*
Wavelet-Translation.LL	3.58	15.75	31.95
LH/HL	4.44	16.82	35.01
<b>Computing Similarity or Rigid</b>			
Wavelet-Similarity.LL	8.30	33.15	67.47
LH/HL	11.43	44.49	91.41
EdgeMatching-Rigid	17.88	48.08	88.97

**Table 4: Timing Results for the Four Algorithms Function of Image Size (in seconds) (\*These Numbers were Obtained by Linear Extrapolation)**

## 4. Conclusion and Future Work

Results of the intercomparison of four automatic image registration algorithms have been presented. Some concluding remarks of this evaluation are the following:

- The four automatic methods considered in this report provide an accuracy below 2 pixels.
- Accuracy is higher when transformation parameters are smaller.
- When a prior knowledge reduces the transformation search to a translation, Phase Correlation seems to combine the best accuracy for a low computational cost.
- When timing is the main concern, a wavelet-based method is the best choice.
- When a rigid transformation is needed and when computation time is not an issue, Iterative Edge Matching is the algorithm which is the most accurate and the most robust to noisy conditions.
- More generally, a trade-off must be achieved between computation time and accuracy of the computed deformation.
- Results also show that while the wavelet-based technique is computationally more efficient for smaller images, the edge-matching method is computationally less demanding for large images.

In future work, we will refine the four previous methods, and we will investigate a larger number of algorithms which will be implemented on several architectures. Detailed performance statistics will be gathered to evaluate accuracy and timing performance of each technique, by utilizing datasets representative of many current and future Earth science applications. The quantitative evaluation of these algorithms will also be extended to other criteria, using other types of datasets, such as space science data, medical imagery, or military applications data. Among the many methods which will be investigated, we will include point-to-point matching algorithms based on spatial relationships between features such as region-based and graph matching methods [14,27-29]. Pre-processing tools such as cloud masking [30] and image enhancement will also be integrated.

Future work will also include the study of computational issues and will focus on the computational aspects and speed of processing of the proposed techniques. Specifically, we will focus on the algorithm enhancement, the performance evaluations, and the parallel implementations of the proposed methods.

As an ultimate goal, we might consider (as was proposed by Rignot [9] and Fonseca [8]) to couple the registration toolbox with a planning-scheduling system [31,32] which would use the above criteria to decide which algorithm to use depending on the application, the type of data, the requested accuracy, and the time and computational constraints.

## Acknowledgments

The author wishes to acknowledge NASA Goddard Space Flight Center Applied Information Sciences Branch of the Earth and Space Data Computing Division for its support, as well as Compton J. Tucker, Eric Vermote, and Nazmi El-Saleous for their advice, their expertise, and for many experimental data. The author is also grateful for useful discussions and collaborations at earlier stages of the project with R.F. Crompt, B-T. Lerner, E. Kaymaz, J. Pierce, S. Raghavan, S. Chettri, and M. Mareboyana.

[1] Townshend, J. et al. (1992). The Impact of Misregistration on Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 30(5).

- [2] Le Moigne, J., Xia, W., Tilton, J. C., Lerner, B.-T., Kaymaz, E., Pierce, J., Raghavan, S., Chettri, S., El-Ghazawi, T., Manohar, M., Netanyahu, N., Campbell, W., and Crompt, R. F. (1997). Towards an Intercomparison of Automated Registration Algorithms for Multiple Source Remote Sensing Data. *Proceedings of the Image Registration Workshop*, NASA Goddard Space Flight Center. NASA Publication #CP-1998-206853.
- [3] Brown, L. (1992). A Survey of Image Registration Techniques. *ACM Computing Surveys*, 24(4).
- [4] Barnea, D. I. and Silverman, H. F. (1972). A Class of Algorithms for Fast Digital Registration. *IEEE Transactions on Computers*, C-21, 179-186.
- [5] Gualtieri, A. J., Le Moigne, J., and Packer, C. V. (1997). Computing Distances Between Images with a Parallel Hausdorff Distance Metric. *International Journal of Computers and Their Applications*, 4(1).
- [6] Mount, D. M., Netanyahu, N. S., and Le Moigne, J. (1998). Efficient Algorithms for Robust Feature Matching. To appear in Special Issue of *Pattern Recognition on Image Registration*.
- [7] Le Moigne, J., Campbell, W. J., and Crompt, R. F. An Automated Parallel Image Registration Technique of Multiple Source Remote Sensing Data. *IEEE Transactions in Geoscience and Remote Sensing*. (CESDIS TR-96-182)
- [8] Fonseca, L. M. G. and Manjunath, B. S. (1996). Registration Techniques for Multisensor Sensed Imagery. *Photogrammetric Engineering and Remote Sensing Journal*, 62(9), 1049-1056.
- [9] Rignot, E. J. et al. (1991). Automated Multisensor Registration: Requirements & Techniques. *Photogrammetric Engineering and Remote Sensing*, 57(8), 1029-1038.
- [10] Lester, H. (1998). A Survey of Hierarchical Non-Linear Medical Image Registration. Special Issue of *Pattern Recognition on Image Registration*.
- [11] LeMoigne, J., El-Saleous, N., and Vermote, E. (1997). Iterative Edge- and Wavelet-Based Image Registration of AVHRR and GOES Satellite Imagery. *Proceedings of the Image Registration Workshop*, NASA Goddard Space Flight Center. NASA Publication #CP-1998-206853.
- [12] Shen, J. and Castan, S. (1986). An Optimal Linear Operator for Edge Detection. *Proceedings of CVPR '86*, (pp. 109-114). Miami Beach, FL.
- [13] Shen, J. and Castan, S. (1992). An Optimal Linear Operator for Step Edge Detection. *Graphical Models and Image Processing*, 54(2), 112-133.
- [14] Tilton, J. C. (1998). Image Segmentation by Region Growing and Spectral Clustering with a Natural Convergence Criterion. *Proceedings of the 1998 International Geoscience and Remote Sensing Symposium*, Seattle, Washington.
- [15] Kuglin, C. D., Blumenthal, A. F., and Pearson, J. J. (1979). Map-matching techniques for terminal guidance using Fourier phase information. *Digital Processing of Aerial Images*, SPIE 186, 21-29.
- [16] Rosenfeld, A., and Kak, A. C. (1982). *Digital Picture Processing*. Academic Press (Second Edition).
- [17] Mallat, S. (1989). A Theory for Multiresolution Signal Decomposition. *IEEE Pattern Analysis and Machine Intelligence*, PAMI-11(7).
- [18] Stone, H. S., Le Moigne, J., and McGuire, M. (1997). Registration Using Wavelet Techniques. *26th Applied Imagery Pattern Recognition Workshop (AIPR'97)*, Washington DC.

- [19] LeMoigne, J. (1994). Parallel Registration of Multisensor Remotely Sensed Images Using Wavelet Coefficients. *Proceedings of the SPIE Aerosense Wavelet Conference*, Orlando, FL.
- [20] Le Moigne, J. (1995). Towards a Parallel Registration of Multiple Resolution Remote Sensing Data. *1995 International Geoscience and Remote Sensing Symposium, IGARSS'95*, Firenze, Italy.
- [21] Huttenlocher, D. P., and Rucklidge, W. J. (1992). *A Multi-Resolution Technique for Comparing Images Using the Hausdorff Distance* (Technical Report 1321). Dept. of Computer Science, Cornell University, Ithaca, NY.
- [22] Huttenlocher, D. P., and Rucklidge, W. J. (1993). A Multi-Resolution Technique for Comparing Images Using the Hausdorff Distance. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 705-706). New York.
- [23] Cracknell, A. P., and Paithoonwattanakij, K. (1989). Pixel and Sub-Pixel Accuracy in Geometrical Correction of AVHRR Imagery. *International Journal of Remote Sensing*, 10(4,5), 661-667.
- [24] Rosborough, G. W., Baldwin, D. G., and Emery, W. J. (1994). Precise AVHRR Image Navigation. *IEEE Transactions on Geoscience and Remote Sensing*, 32(3).
- [25] Baldwin, D., and Emery, W. (1995). Spacecraft Attitude Variations of NOAA-11 Inferred from AVHRR Imagery. *International Journal of Remote Sensing*, 16(3), 531-548.
- [26] Campbell, W. J., Short Jr., N. M., Coronado, P., and Crompton, R. F. (1994). Distributed Earth Science Validation Centers for Mission to Planet Earth. *8th International Symposium on Methodologies for Intelligent Systems, ISMIS'94*. Charlotte, NC.
- [27] Ton, J., and Jain, A. K. (1989). Registering Landsat Images by Point Matching. *IEEE Transactions in Geoscience and Remote Sensing*, 27(5).
- [28] Goshtasby, A., Stockman, G., and Page, C. (1986). A Region-Based Approach to Digital Image Registration with Subpixel Accuracy. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(3).
- [29] Goshtasby, A. (1988). Registration of Images with Geometric Distortions. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 60-64.
- [30] McGuire, M., and Stone, H. S. (1997). Techniques for Multiresolution Image Registration in the Presence of Occlusions. *Image Registration Workshop*, NASA Goddard Space Flight Center. NASA Publication #CP-1998-206853.
- [31] Boddy, M. et al. (1994) Planning for Image Processing. *NASA/Goddard Conference on Space Applications of Artificial Intelligence*.
- [32] Lansky, A. L. et al. (1995). The COLLAGE-KHOROS Link: Planning for Image Processing Tasks. *AAAI Spring Symposium on Integrated Planning Applications*. Stanford Univ.

## **Imaging and Phase Retrieval-Based Wavefront Sensing**

**Richard G. Lyon**  
**University of Maryland Baltimore County**  
**Department of Computer Science and Electrical Engineering**  
**(lyon@jansky.gsfc.nasa.gov)**

### **Profile**

Richard G. Lyon holds a Bachelor of Science in physics from the University of Massachusetts and a Master of Science in optics from the University of Rochester with work towards a Ph.D. in optics at the University of Rochester. He is member of the Optical Society of America (OSA) and the Society for Photo-Instrumentation Engineers (SPIE).

From 1987 to 1992 Mr. Lyon was employed by Hughes Danbury Optical Systems (now Raytheon Optical Systems Inc.) as an optical systems engineer in the Space Sciences directorate. In that capacity he served as principal investigator of Hubble Space Telescope phase retrieval efforts to determine the on-orbit telescope error. During this period he received a NASA Goddard Space Flight Center Certificate of Recognition for Contributions to the Hubble Space Telescope Program, a NASA Goddard Group Achievement Award for the Hubble Space Telescope Mission Operations Team, and a NASA Award of a Flag flown on STS-31 for contributions to the Hubble Space Telescope Program.

From January 1993 to June 1994 Mr. Lyon worked as a research analyst for Radex Incorporated where designed, developed and implemented automated celestial image processing algorithms for the Mid-Course Space Experiment (MSX), a U.S. Air Force radiometric satellite. In June 1994 he became a principal engineer with Hughes STX where he conducted research into the design and development of optical and image processing algorithms to operate in massively parallel computational environment, including image restoration and image deconvolution algorithms for the Hubble Space Telescope.

Currently, Mr. Lyon is a Research Scientist at the Center of Excellence in Space Data and Information Sciences and is Technical Director of the Optical Systems and Characterization Project (OSCAR) at NASA/Goddard Space Flight Center. OSCAR is currently funded by both NASA/GSFC and NASA/JPL to conduct research into computational and hardware methods of wavefront sensing and imaging for the Next Generation Space Telescope (NGST), the Deployable Cryogenic Active Telescope Testbed (DCATT) and Pathfinder III:Nexus. In addition, OSCAR is building its own wavefront sensing benchtop demonstration system. Mr. Lyon is also Co-I on the recently accepted pre-Phase A Integrated Instrument Science Module concept study for NGST to design a coronagraphic instrument for NGST.

### **Report**

#### **1. Hubble Space Telescope and the OSCAR Project**

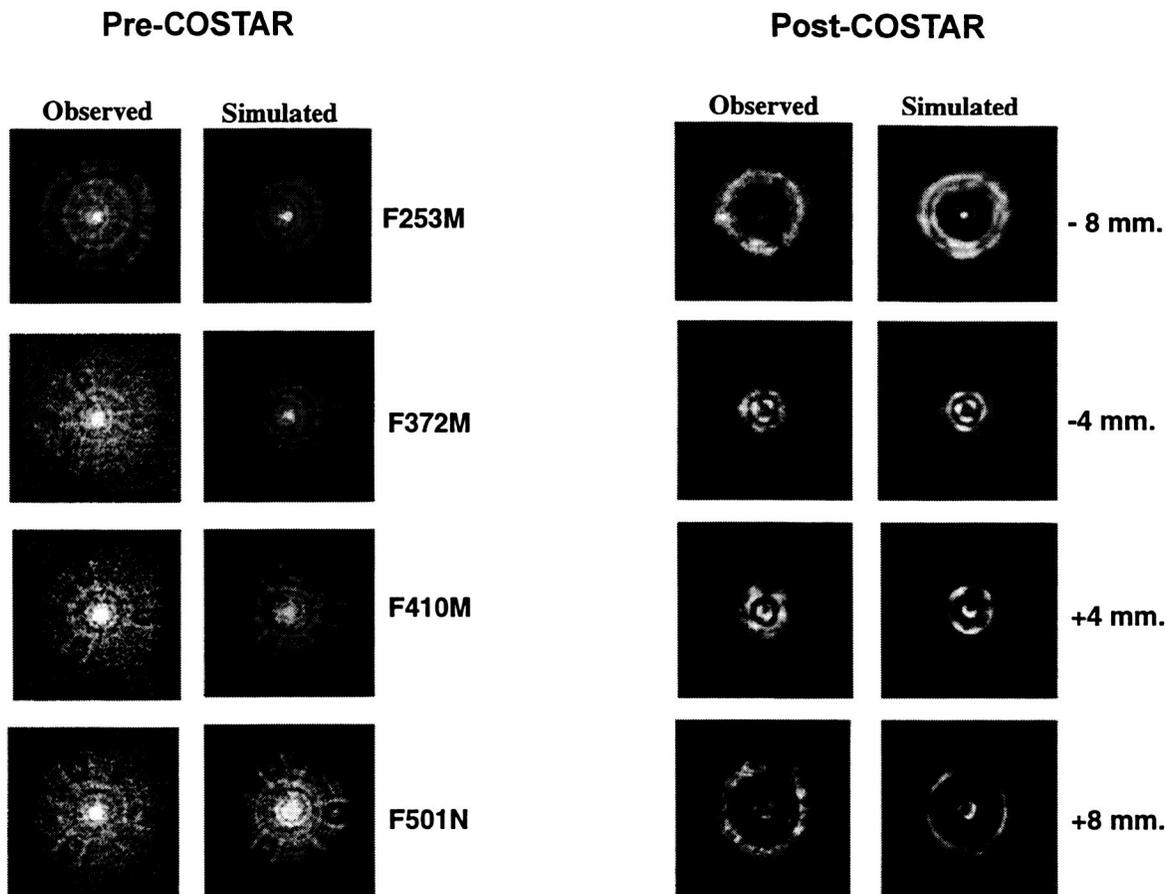
The Optical Systems Characterization and Analysis Research (OSCAR) Project at NASA/Goddard Space Flight Center's Earth and Space Data Computing Division (Code 930) conducts research into applying massively parallel computers and computational techniques to solve complex optical, imaging, and data analysis problems. The optical system problems with the original Hubble Space Telescope (HST) fostered the OSCAR project in 1990. Early on, it was recognized that massively parallel computers could efficiently and quickly calculate high fidelity optical models of the HST optical system to deduce the errors via phase retrieval techniques; this, in turn, led to calculated point spread functions (PSFs) which are necessary to perform optimal image deconvolution. These successes provided the rationale for phase retrieval to be adopted as the baseline wavefront sensing method to study the alignment and fine figure control of the Next Generation Space Telescope (NGST) [1][2][3].

Prior to HST launch in 1990, Grey and Lyon [4] proposed phase retrieval as a backup method in the event of failure of one or more of the three on-board wavefront sensors of the Hubble Space Telescope. They showed that modest amounts of focus, coma, and astigmatism could be determined by imaging unresolved stars through narrowband filters onto the focal planes of the Faint Object Camera (FOC) and the Wide Field Planetary Camera (WF/PC). Considering only misalignment dependent aberrations and by exploiting the different field locations of the cameras and different combinations of one or more of the on-board interferometers, the field dependence of the aberrations could be determined to align the HST's secondary mirror.

Phase Retrieval was resurrected following the discovery of the HST primary mirror conic constant error and its consequent spherical aberration of the telescope. The amount of spherical aberration was outside the dynamic range of the on-board wavefront interferometers. A number of different phase retrieval methods were used by different research groups [5][6][7][8] to determine the spherical aberration. Further refinement of phase retrieval also led to consistent signatures for misalignment dependent aberrations and to HST optical prescription predictions for the Corrective Optics Space Telescope Axial Replacement (COSTAR) mission [9]. Even finer refinement led to determination of the combined phase errors due to the residual polish marks on the HST primary and secondary mirrors [10][11]. It is the determination of these polish marks which eventually led to a unified consistent model for the telescope with enough accuracy to calculate pre-COSTAR PSFs with enough fidelity for reliable image deconvolution [12][13] and to subsequently predict post-COSTAR PSFs [8]. Figures 1 and 2 are graphical synopses of these results. Thus, one of the technological legacies of the HST is the adoption of phase retrieval methods as the baseline wavefront sensing method on NGST. Moreover, it may well prove that phase retrieval will be used to determine the initial on-board alignment of NGST optics as well as periodically maintaining fine figure control.

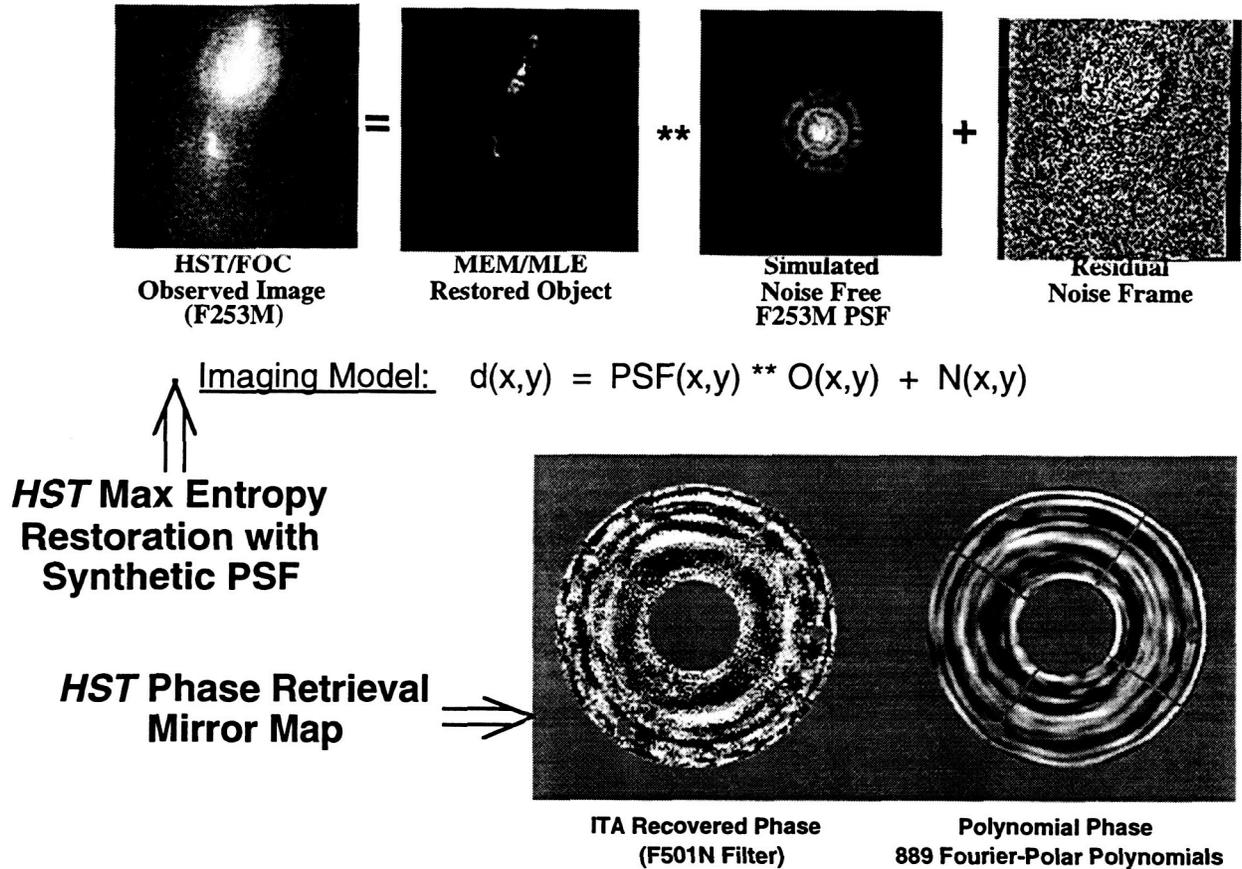
## 2. Optical Systems Modeling and Phase Retrieval

Phase retrieval is essentially a method of finding the wavefront error in an optical system from an ensemble of observed focal plane images. The wavefront error can result from a variety of causes including aberrations due to design residuals, fabrication errors, polish marks, alignment errors, and/or thermal and structural drift. In phase retrieval techniques, the inputs are observed images of a narrowband unresolved source such as the HST example in the left column of Figure 1. The output is the wavefront error in the optical systems exit pupil. An output example of two wavefront maps are shown in the lower right of Figure 2. The left map is a phase retrieval result utilizing only a single input PSF while the right map is from simultaneously phase retrieving nine PSFs with a diversity of both focus and wavelength. It is the mid- to high-spatial frequency wavefront which gives the fine detail in the Figure 1 PSFs. The phase retrieval method of wavefront sensing is akin to interferometry but with the advantage that, in principle, no additional hardware is required. The science camera is used to generate the input images and, therefore, inherently has the aberrations associated with the entire optical train. On the other hand, interferometry requires its own complex, flight-qualified optics which must be calibrated and, generally, the science camera cannot be used. Hence, this method does not "see" the entire optical train, which requires higher optical tolerances on the instrument. Compared to interferometry, phase retrieval trades hardware for a software solution. However, phase retrieval algorithms are non-linear and, therefore, computer runs are non-deterministic in time. Moreover, phase retrieval requires on the order of Tera-floating point operations and also requires a validated high fidelity computer model of the entire optical system. Further, solutions can yield potential problems with convergence and phase unwrapping. We are in the process of investigating methods of guaranteeing convergence while minimizing processing time. In addition, our NGST work will investigate a number of different algorithms and apply them in a Monte-Carlo fashion, to determine the best approach.



**Figure 1: Hubble Space Telescope Phase Retrieval Results  
Faint Object Camera Point Spread Functions**

- Leftmost column: Set of 4 pre-COSTAR observed Faint Object Camera (FOC) stellar images (PSFs), note the wide dynamic range and hence non-stationary signal to noise ratio.
- Second column from left: Set of LEO modeled PSFs utilizing phase retrieval to find the wavefront error. The numbers correspond to the filter number, e.g., F253M means the 253 nanometer filter and M means medium band, N means narrowband. Note the level of detail in the simulated PSFs which is partially "washed" out in the observed PSFs.
- Third column from left: Set of 4 post-COSTAR through focus observed PSFs.
- Rightmost column: Set of LEO modeled post-COSTAR PSF's. This set matches the observed set in nearly every detail. The numbers correspond to the position of the deployable optical bench. This verifies that a consistent unified wavelength independent model of both the pre- and post-COSTAR HST optical system exists. See reference [10] for more details.



**Figure 2: Hubble Space Telescope Maximum Entropy Image Deconvolution and Residual Wavefront**

(Top row) A Maximum Entropy restoration of a HST/FOC image with LEO calculated PSFs. The leftmost image is the raw FOC image at 253 nanometers. The second from the left is the MEM/MLE restoration (see reference [13]). The third from the left is the LEO modeled PSF. The far right image is the residual noise frame generated by convolving the restored image with the simulated PSF and subtracting it from the observed data and then weighting it by the noise standard deviation on a pixel by pixel basis. Ideally the residual noise frame should be entirely de-correlated, however, some residual structure is evident showing that the deconvolution process is less than perfect.

## 2.1 Forward Modeling

In order to test the various phase retrieval methods, the Lyon Electro-Optical (LEO) modeling and analysis package has been developed. LEO was used to previously simulate imagery for HST FOC and WF/PC and is currently being used to simulate imagery for the Deployable Cryogenic Active Telescope Testbed (DCATT) [14] and for NGST. LEO currently incorporates the following:

1. Multiple plane diffraction, Fresnel, Fraunhofer, and rigorous Angular Spectrum.
2. Segmented apertures and obscurations.
3. Full aperture Zernike polynomials.
4. Sub-aperture Zernike polynomials (i.e. each segment can have its own set with the center and normalization radius arbitrary).
5. Random power law surfaces with low and high cutoffs, integrated RMS power, and also power spectral density slope. This generates speckle in focal plane.
6. White noise, harmonic and low frequency jitter models.
7. Deformable mirror influence function models, quantization error, and range limits.
8. Detector modulation transfer function, charge transfer efficiency, pixelization effects, quantization error, and dynamic range effects.
9. Gaussian and Poisson noise models.
10. System radiometry, specify star color temperature, spectral filter functions, optics transmission, and quantum efficiency.
11. Some extended scene modeling capability as "seen" through the optical system.
12. Generic coronagraphic capability with assortment of masks and Lyot stops.

LEO output can take the form of any of the following:

1. Monochromatic or polychromatic Point spread functions.
2. Point response functions, with detector effects folded in.
3. Complex pupils functions including amplitude and phase (wavefront error).
4. Optical transfer function and modulation transfer function, both single wavelength and monochromatic.
5. Surface to surface raytrace on high density grids, e.g., 1024 x 1024 possibly higher.
6. Output "Scenes" as seen through the entire imaging system.

LEO is entirely written in MPL which is a massively parallel superset of "C". LEO runs extremely fast generally taking less than 1 second to execute for a 15 optical element system, and will perform 40 surface to surface diffraction calculations per second. It currently runs on Goddard Space Flight Center's MasPar MP2 computer which is a massively parallel compute engine consisting of 16,384 separate processors

with an associated communications grid. LEO can use adjustable array sizes, with the baseline being 512 x 512, thus it can raytrace grids of 512 x 512 rays and perform 512 x 512 FFT's. The simulated PSFs in Figure 1 for HST are output from the LEO package.

## 2.2 Inverse Modeling

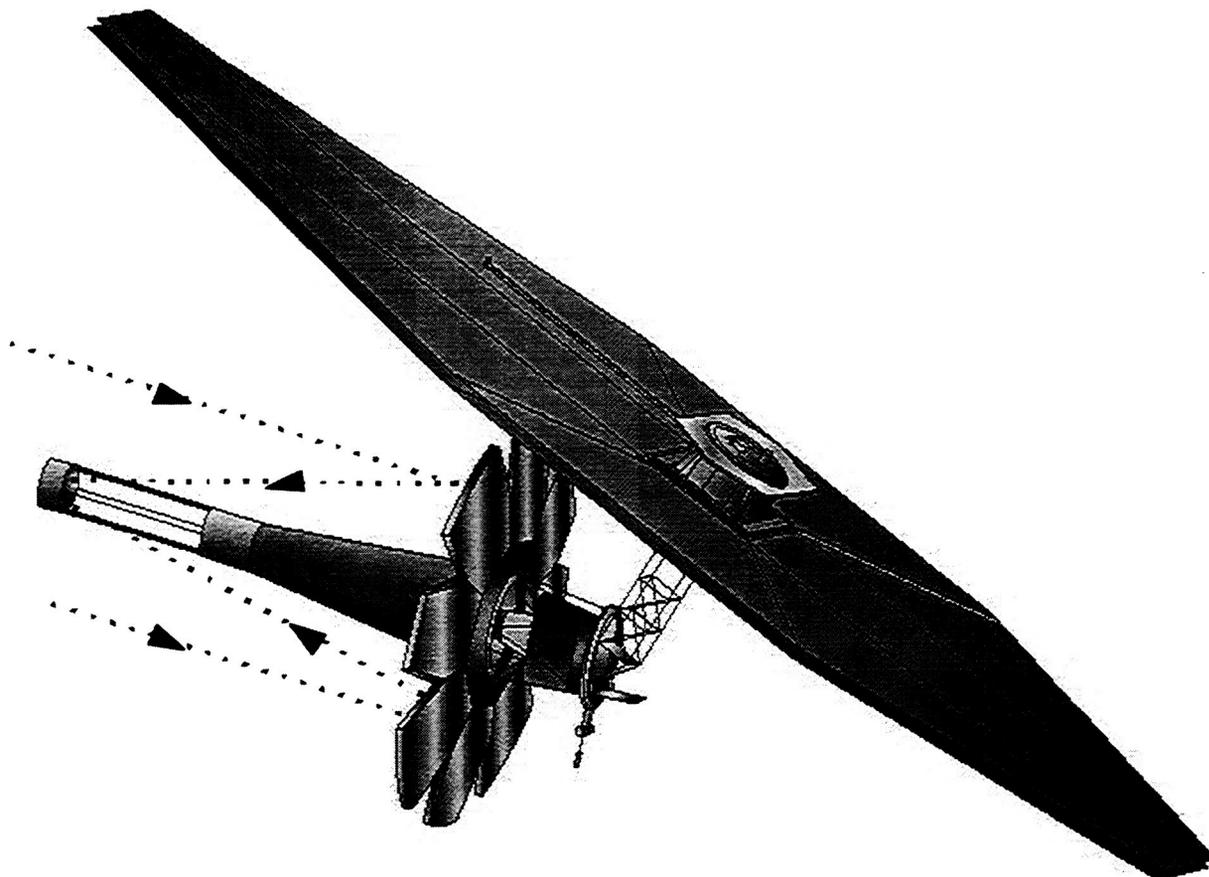
Currently a number of phase retrieval algorithms are coded and operational. We are in the process of quantitatively studying the accuracy and precision of each of the different algorithms. Some of the problems that need to be addressed are the effects of jitter, finite sampling, finite pixel size, finite spectral pass-band, convergence and stagnation issues, phase unwrapping, and some of the effects due to segmented optics and active optics. We are also studying whether any significant advantage can be gained by performing phase retrieval in an autonomous control loop on-board a spacecraft with minimal or no communications with the ground.

## 3. The Next Generation Space Telescope Wavefront Sensing and Optical Control System

The NGST will most likely be an 8 meter aperture telescope, operating at least over the wavelength band 1.0 to 5.0 microns. NGST is required to be diffraction limited ( $\lambda/14$ ) at 2 microns. Figure 3 shows a rendition of the GSFC/JPL design. The size and weight constraints for NGST dictate that the primary mirror (PM) be a lightweight multi-segmented mirror. One of the main technological challenges will be to initially phase the segments, to maintain the segment alignment, and, in general, maintain the alignment and figure control of the whole optical train. This can be quite a daunting task in orbit. Thus a phase retrieval-based optical control system will be studied, simulated and optimized, and tested on a ground testbed. This will first be done in a pure computing environment, then, a subset of the methods studied will be tested, in a hardware configuration, on the DCATT [14] and eventually on a technology demonstration flight mission known as Nexus[15]. The Nexus mission will be a segmented aperture telescope adopting the optimal phase retrieval-based optical control system tested on DCATT. This will be a validation flight for the final design of the wavefront sensing and optical control system for NGST. Figure 4 shows a schematic of one possible optical control loop for NGST. The telescope entrance pupil is imaged onto the deformable mirror (DM) via an off-axis parabola and the telescope's cassegrain focus is relayed to the wavefront sensor (WFS) camera and the fast steering mirror (FSM) camera. The FSM camera is essentially a quad cell motion detector. The control loop for the FSM feeds tip/tilt commands to the FSM to keep the image stationary due to system jitter. The WFS camera collects an image (or set of images) and passes them through the phase retrieval software system to recover the wavefront errors. The wavefront errors are then used to determine optimal actuator steps to minimize the wavefront error, and commands are sent to the PM/SM (secondary mirror) actuators and to the DM actuators.

We have currently modeled the NGST from an optical systems point of view, including the telescope baseline design, a generic science camera, wavefront sensing, and the optical control loop. This baseline model will be used to perform a number of parametric trade studies as well as a number of different phase retrieval based wavefront sensing options and a number of possible different actuator-based control loops. We are able to model the operational scenario and to investigate a number of different paths to minimize the RMS wavefront errors. Figure 5 shows the phase retrieval-based control loop. Two observed PSFs, one on each side of focus, are input to a phase retrieval method. The resultant wavefront is recovered, modulo  $2\pi$ , and input to a phase unwrapping algorithm. This resultant wavefront is fit to the DM actuator influence functions, and the DM surface is moved to compensate for the error to bring the wavefront error down to 0.05 wave. The resultant PSF and a "perfect" PSF is shown for comparison; both are logarithmically stretched to enhance low level detail. In this context, "perfect" refers to no wavefront error. Note the waffling in the DM corrected wavefront which is due to the underlying actuators.

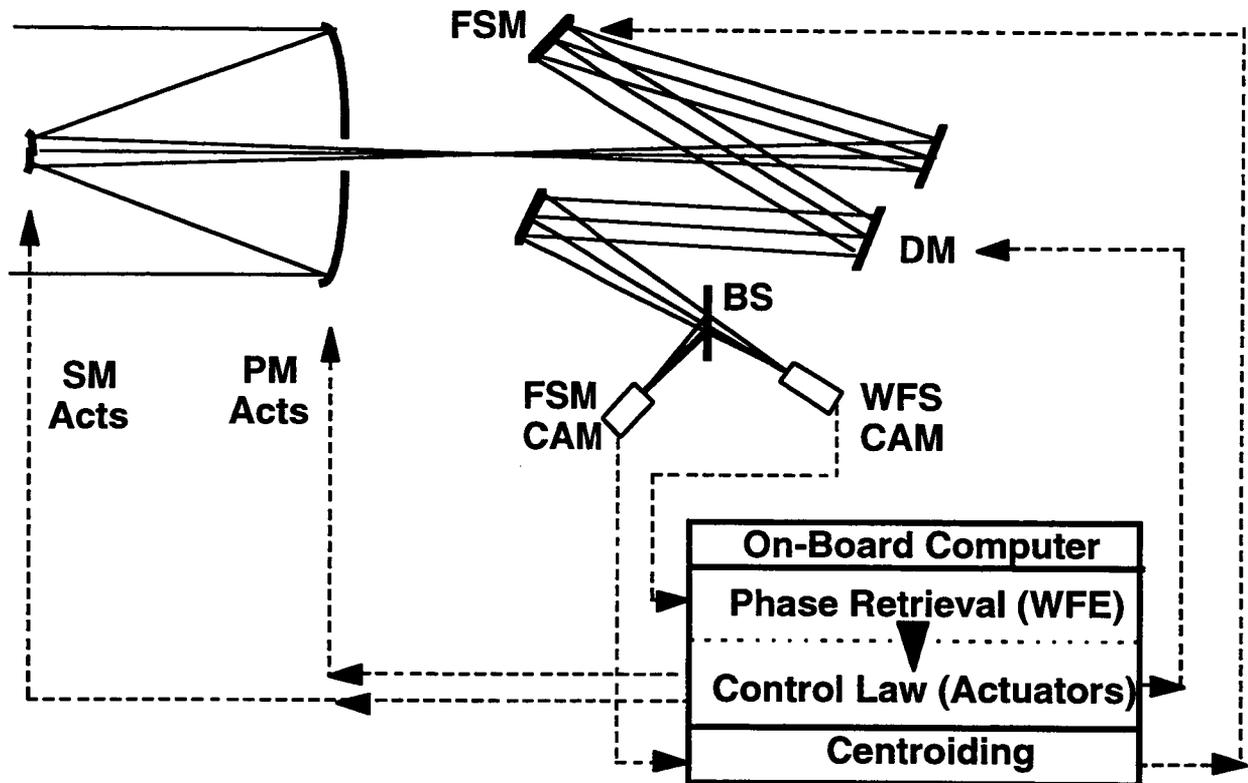
### NGST GSFC/JPL Design



**Figure 3: GSFC/JPL Design for NGST**

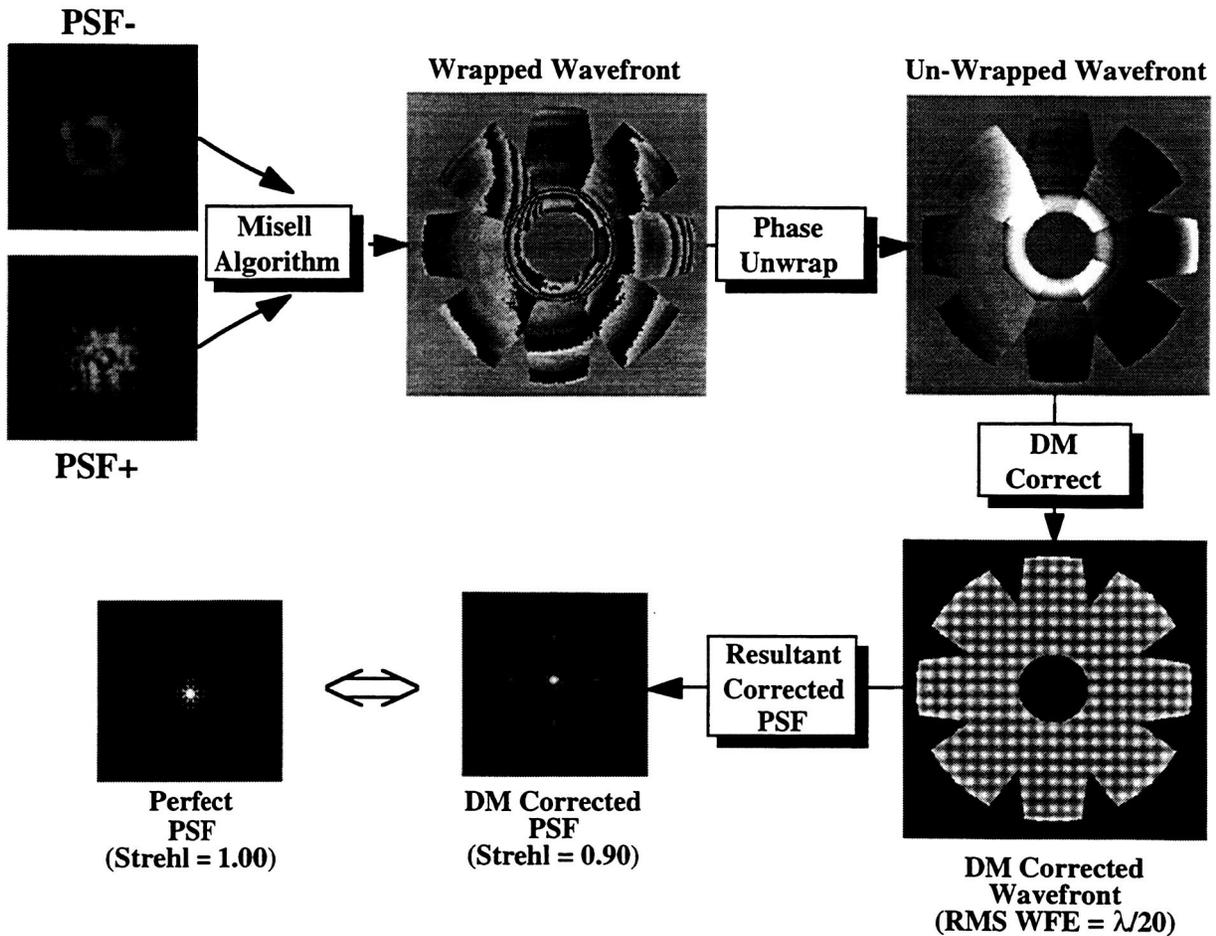
Light enters from the upper left, reflects off the segmented primary mirror to the monolithic secondary mirror then into the instrument aperture. This telescope design has a large sun shield (top) with the electronics packages in the center of the sun shield. There is no active thermal control system as on HST and thus the primary mirror temperatures could range from 30 to 70K with relatively strong thermal gradients.

## Conceptual NGST and DCATT Optical Control Loop



**Figure 4: NGST and DCATT Conceptual On-board Optical Control Loop**

Each of the primary mirrors (PM) will move in piston and tip/tilt, the secondary mirror (SM) will move as a rigid body in 6 degrees of freedom. The PM is re-imaged onto the deformable mirror (DM) and also onto a fast steering mirror (FSM). The prime focus of the telescope is relayed through the optical system to both an FSM camera and a WFS camera. The FSM camera is essentially a quad cell detector which centroids the PSF and feed forwards commands to the FSM. Thus the FSM tips and tilts to maintain the position of the PSF on the output camera detector array. This compensates for system jitter. The WFS camera, is essentially the science instrument camera, and measures a sequence of images at various foci. The resulting set of images is phase retrieved, and actuator commands are generated and fed to the DM, PM, and SM.



**Figure 5: NGST Phase Retrieval Based Optical Control Loop Simulation**

- Upper left: Two input PSFs, one on each side of focus. These are input to Misell phase retrieval algorithm.
- Upper middle: Output of phase retrieval is the entire optical train's wavefront error. This is returned modulo  $2\pi$  unless the wavefront error is less than 1 wave.
- Upper right: Unwrapped wavefront error. Note that each segment can have its own errors. This simulation also contains residual polish marks and surface microroughness.
- Lower right: The un-wrapped wavefront is fit to the actuator influence functions and the DM moved to correct the wavefront. Shown is the residual "quilt" pattern. The DM corrected wavefront is  $\lambda/20$  rms WFE.
- Lower middle: The resultant DM corrected PSF. Note that this is logarithmically stretched to bring up some of the residual background structure.
- Lower left: The perfect PSF, i.e., with no WFE in the entire system, is shown for comparison. This also has the same logarithmic stretch as the DM corrected PSF.

A number of the studies have been conducted, or are currently in progress, and will be briefly mentioned here. Different sequences of actuation on the PM/SM and DM combination are considered. For example, one trade-off study will address whether the PM segments should move only in piston and tip/tilt while the higher order aberrations are corrected by the DM, or, alternatively whether the PM should also correct higher order modes with and without a DM. There are other issues as well. For example: would it be better to thermally control the PM and possibly the SM? how can we continuously monitor image quality? what control scenario minimizes the "waffling" introduced due to the DM? how can we fold in trend data to the control loop? and, what can be gained, from a scientific point of view, by performing the WFS and optical control system (OCS) autonomously on-board the spacecraft without any operator/analyst in the loop? This latter aspect would reduce the telemetry bandwidth for the periodic alignment process and possibly allow alignment more often to maintain higher image quality throughout mission life. Also due to potentially long thermal settling times, it may be better to actively control the optics during and following a slew as opposed to letting the optics come to equilibrium then correcting them. Also how much better image quality can we obtain by using very high density deformable mirrors (~10,000 to 20,000 actuators) and/or a segmented aperture DM? We have also been modeling coronagraphic options and different methods of wavefront sensing through the coronagraph.

## References

- [1] Seery, B.D. (1998). Next Generation Space Telescope (NGST). *Proceedings of SPIE: Vol. 3356*. Kona Hawaii
- [2] Mather, J.C., and Stockman, P. (1997). NGST: Technology for Looking Back in Time. *BAAS.*, 191(54.02).
- [3] Lyon, R. G., Hollis, J. M., Dorband, J., and Murphy, T.P. (1998). Extrapolating HST Lessons to NGST. Invited Paper, *Optics and Photonics News*, 9(7).
- [4] Grey, L. G., and Lyon, R. G. (1989). Correction of Misalignment Dependent Aberrations of the Hubble Space Telescope via Phase Retrieval. *Proceedings of SPIE: Vol. 1168*.
- [5] Lyon, R. G., Miller, P. E., and Gruszczak, A. (1991). Hubble Space Telescope Phase Retrieval: A Parameter Estimation. *Proceedings of SPIE: Vol. 1567*.
- [6] Redding, D., Dumont, P., and Yu, J. (1993). Hubble Space Telescope Prescription Retrieval. *Applied Optics*, 32(10)
- [7] Burrows, C. J. (1991). Hubble Space Telescope Optics Status. *Proceedings of SPIE: Vol. 1567*.
- [8] Fienup, J. R. (1991). Phase Retrieval for the Hubble Space Telescope Using Iterative Propagation Algorithms. *Proceedings of SPIE: Vol. 1567*.
- [9] Wood, H. J. (1994). Hubble Space Telescope: Mission Update. *Optics and Photonics News*, 5(8).
- [10] Lyon, R. G., Dorband, J. E., and Hollis, J. M. (1997). Hubble Space Telescope Faint Object Camera Calculated Point Spread Functions. *Applied Optics*, 36(8).
- [11] Krist, J. E., and Burrows, C. J. (1995). Phase Retrieval Analysis of pre- and post-repair Hubble Space Telescope Images. *Applied Optics*, 34(22).
- [12] Hollis, J. M., Pedelty, J. A., and Lyon, R. G. (1997). Spatial Resolution of the R-Aquarii Binary System. *ApJ Letters, ApJ*, 488, L85-88.

- [13] Hollis, J. M., Lyon, R. G., Dorband, J.E., and Feibelman, W.A. (1997). Motion of the Ultraviolet R Aquarii Jet. *ApJ*, 475, 231-236.
- [14] Lebeouf, C., Davila, P. S., Redding, D. C., Coulter, D. R., and Pacini, L. (1998). Deployed Cryogenic Active Telescope Testbed for the Next Generation Space Telescope. *Proceedings of SPIE: Vol. 3356*. Kona Hawaii
- [15] Pacini, L., Geithner, P. H., Jacobson, D. N., and Redding, D. C. (1998). NEXUS: The Next Generation Space Telescope Integrated Flight Systems Demonstration. *Proceedings of SPIE: Vol. 3356*. Kona, Hawaii

**Timothy Murphy**  
**University of Maryland Baltimore County**  
**Department of Computer Science and Electrical Engineering**  
**(murphy@albert.gsfc.nasa.gov)**

## Profile

Mr. Murphy holds a B.S. and M.S. in Electrical Engineering from Columbia University. He was employed by the Perkin-Elmer Corporation from 1980 to 1991 where his responsibilities included analysis of space radiation effects, design of radiation shielding for electro-optical sensors, research into DSP applications for spectroscopy instruments and electro-optical system analysis. He was also employed by Dalsa, Inc., Waterloo, Ontario in 1991 and 1992 for design of a data acquisition and test station for the first 25 million-pixel CCD. From 1993 to 1998 he was employed by GN Nettest, Fiber Optics Division, Utica, N.Y., where he designed signal processing software for fiber optic equipment.

At CESDIS Mr. Murphy is working on image processing software for the Deployable Cryogenic Telescope Testbed (DCATT), phase retrieval software for the Next Generation Space Telescope (NGST) [1] and laboratory research in phase diverse imaging.

## Report

### 1. Phase Unwrapping of 2-D Phase Retrieval Outputs

Phase retrieval is an active area of optical systems research. One of the primary goals of this work is to determine the optical aberrations of a system from observations of its point spread function. The point spread function is submitted to an iterative algorithm which produces a complex 2-D map of the optical system response, which contains the aberrations of the system. The system phase response is then available as the arctangent of the ratio of imaginary over real parts. Because the range of arctangent is  $-\pi$  to  $\pi$ , the phase map cannot show responses outside this range. Thus, if, for example, the optical system has 2 waves, i.e.,  $4\pi$ , of spherical aberration, the phase map will show only values varying by  $2\pi$ . The phase map is said to be wrapped over a  $2\pi$  range. In places where the phase approaches  $+\pi$ , the phase map will be seen to jump to  $-\pi$ . Similar behavior is seen at  $-\pi$ , where the phase jumps to  $+\pi$ .

The algorithm by Servin, et al., [2] has been tried for unwrapping of phase retrieval results without a very good result. There are several reasons for this. This algorithm requires one to choose a starting point in a smooth region of the phase map, but in an autonomously operated machine, such as NGST, there is no intelligent operator to choose such a region. Also, this algorithm operates on a point-by-point basis. Thus,

it is difficult to get it to take advantage of parallel processors and, for the 512x512 images we have been using, it is slow. While testing this algorithm it became evident that if the starting point was chosen at a point of steep phase change, then it would produce incorrect results.

To try to get some headway with a simple phase unwrapping program, a modulo  $2\pi$  polynomial fit to the wrapped data was attempted. This was successful on some relatively simple phase maps. The polynomials used were not orthogonal and the fitting technique used was steepest descent. Since a steepest descent algorithm is a relatively naive approach which can require "hand tuning", I am currently incorporating a Polak-Ribière [3] manifestation of the conjugate gradient algorithm. Another improvement being incorporated is the use of Zernike polynomials as a basis set. These are an orthogonal set of 2-D polynomials used frequently in optical work.

The polynomial fit to the phase map does not need to be exact. If we find the unwrapped phase map to within  $\pm\pi$ , then we can use the wrapped phase map to remove any errors and find a corrected unwrapped phase.

$$\bar{\phi} = \hat{\phi} - \arctan(\tan(\hat{\phi} - \phi_w))$$

Here  $\bar{\phi}$  is the corrected unwrapped phase,  $\hat{\phi}$  is the uncorrected unwrapped phase, and  $\phi_w$  is the wrapped phase. Another approach I will be examining is to retain an unwrapped phase map at all stages of the phase retrieval iterations.

## 2. Image Processing for the DCATT Program

The DCATT program is intended to provide insights to the engineering problems associated with spaceborne segmented mirror telescopes such as NGST. The idea is uncover potential difficulties before the design of NGST is very far along. Phase retrieval algorithms will be tested on DCATT.

DCATT operates in the visible wavelengths while NGST is an infrared telescope. The detector array for NGST will be a hybrid compound detector layer (such as InSb or HgCdTe) coupled with a silicon readout array. The DCATT detector is a silicon CCD array. The effect of charge transfer inefficiency can distort the DCATT output by as much as a few percent compared to the NGST detectors which do not experience this effect. An image processing program was written in C to remove charge transfer inefficiency effects from the DCATT images. A program is underway to make a detailed characterization of the DCATT detector so that all aspects of this detector that may affect phase retrieval are quantified.

## References

- [1] Sttrockman, H. S. (1997). *The Next Generation Space Telescope*. Association of Universities for Research in Astronomy, Inc.
- [2] Servin, M., Marroquin, J. L., Malacara, D. and Cuevas, F. J. (1998). Phase unwrapping with a regularized phase-tracking system. *Applied Optics*, 37(10), 1917-1923.
- [3] Polak, E. (1971). *Computational Methods in Optimization*. Academic, New York.

## **Image Registration for the Regional Application Center**

**Nathan S. Netanyahu**  
**University of Maryland College Park**  
**Center for Automation Research (CFAR)**  
**(nathan@haifa.gsfc.nasa.gov)**

### **Profile**

Dr. Netanyahu holds a Bachelor of Science and Master of Science in electrical engineering from the Technion, Israel Institute of Technology and a diploma in computer science from Tel-Aviv University. He also holds a Master of Science and a Ph.D. in computer science from the University of Maryland at College Park. He is a Member of the IEEE.

From 1973 to 1978, Dr. Netanyahu served as a technical officer and project engineer in the Intelligence Unit of the Israel Defense Forces where he designed and developed electronic communication sub-systems. From 1978 to 1985 he served as a senior project engineer in the Electronic Research Department at the Israeli Ministry of Defense where he designed and developed electronic communication systems and computerized test modules for their automatic performance evaluation.

While working on his advanced degrees at the University of Maryland, Dr. Netanyahu was employed as a research assistant by the Center for Automation Research's Computer Vision Laboratory. From 1992 to May of 1994 as a National Research Council Associate attached to NASA Goddard's Space Data and Computing Division (Code 930), he worked on unsupervised methods for clustering air/spaceborne multi-spectral images, and derived computationally efficient algorithms for robust statistical estimation.

Dr. Netanyahu joined CESDIS through a subcontract with the University of Maryland College Park in May 1994. He has been working on image registration and supervised classification of remotely sensed images, and has continued to pursue unsupervised (robust estimation-based clustering of multispectral images and computationally efficient algorithms for robust estimation.

Current research interests include algorithm design and analysis, computational geometry, image processing, pattern recognition, remote sensing, and robust statistical estimation.

### **Report**

#### **1. Introduction**

To prepare for the challenge of efficiently handling the archiving and repeated querying of terabyte-sized scientific spatial databases, the Applied Information Sciences Branch (AISB), Code 935, NASA GSFC has developed over the years a number of data processing modules (e.g., the Intelligent Data Management (IDM) system, the Intelligent Information Fusion System (IIFS), etc.), the culmination of which is the end-to-end information system, known as the Regional Application Center (RAC).

The main objective of the RAC is to provide a user with the ability to directly receive and manipulate current, localized satellite data in a cost-effective manner and on a routine basis. To achieve that, the RAC is designed to efficiently perform a number of generic functions that play a key role in various remote sensing applications. Such functions include data browsing, data querying, and data characterization, i.e., automatic characterization/extraction of image content.

Since the analysis techniques that an RAC is expected to utilize in carrying out the above functions often involve the integration of multiple data sources (e.g., global coverage analysis of low-resolution data can be validated/refined by local, high resolution data), enabling a user to analyze large amounts of pertinent data more accurately and efficiently requires that the RAC contain a sound image registration scheme.

Various modules of such a scheme have been developed recently by a team of researchers at the AISB. The idea is to establish, essentially, a registration toolbox, i.e., a diverse set of tools for image registration which will eventually be incorporated into the RAC.

One of the fundamental building blocks in any control point-based registration scheme relies on matching features that are extracted from one image (the sensed image) to their counterparts in a second image (the reference image). The extracted features could be points, edge segments, corners, etc. Although feature-based methods tend to be relatively accurate (as features are more reliable than intensity or radiometric values), they could become computationally expensive. To alleviate this difficulty, we have developed an efficient algorithmic methodology for feature matching. The scheme derived was based largely on computational geometry techniques, and is expected to be incorporated into the RAC's registration toolbox.

## 2. Efficient Robust Feature Matching

One of the basic building blocks in any point-based registration scheme involves matching feature points that are extracted from a sensed image to their counterparts in a reference image. This leads to the fundamental problem of point matching: Given two sets of points, find the (affine) transformation that transforms one point set so that its distance from the other point set is minimized. Because of measurement errors and the presence of outlying data points, it is important that the distance measure between the two point sets be robust to these effects. We measure distances using the partial Hausdorff distance.

Point matching can be a computationally intensive task, and a number of theoretical and applied approaches have been proposed for solving this problem. In Mount, Netanyahu, and Le Moigne '97, '98 we presented two algorithmic approaches to the point matching problem, in an attempt to reduce its computational complexity, while still providing guarantees on the quality of the final match. Our first method is an approximation algorithm, which is loosely based on a branch-and-bound approach due to Huttenlocher and Rucklidge '92, '93. We show that by varying the approximation error bounds, it is possible to achieve a trade-off between the quality of the match and the running time of the algorithm. Our second method involves a Monte Carlo method for accelerating the search process used in the first algorithm. This algorithm operates within the framework of a branch-and-bound procedure, but employs point-to-point alignments to accelerate the search. We show that this combination retains many of the strengths of branch-and-bound search, but provides significantly faster search times by exploiting alignments. With high probability, this method succeeds in finding an approximately optimal match. We demonstrate the algorithms' performances on both synthetically generated data points and actual satellite images.

## 3. Committees

Served on the organizing committee of the CESDIS Image Registration Workshop held at NASA/GSFC, November 20-21, 1997.

## 4. Recent Relevant Publications

Netanyahu, N., Chettri, S., Garegnani, J., Robinson, J., Coronado, P., Crompton, R. F., and Campebell, W. J. (1997). Multiresolution Maximum Entropy Spectral Unmixing. *Proceedings of the International Symposium on Artificial Intelligence, Robotics, and Automation in Space* (pp. 347-352). Tokyo, Japan.

Netanyahu, N., Xia, W., Le Moigne, J., Tilton, J. C., Lerner, B-T., Kaymaz, E., Pierce, J., Raghavan, S., Chettri, S., El-Ghazawi, T., Manohar, M., Campbell, W. J., and Cromp, R. F. (1997). A Registration Toolbox for Multi-Source Remote Sensing Applications. *International Conference on Earth Observation and Environmental Information*. Alexandria, Egypt.

Netanyahu, N., Le Moigne, J., Xia, W., Chettri, S., El-Ghazawi, T., Kaymaz, E., Lerner, B-T., Manohar, M., Pierce, J., Raghavan, S., Tilton, J. C., Campbell, W. J., and Cromp, R. F. (1997). Towards an Intercomparison of Automated Registration Algorithms for Multiple Source Remote Sensing Data. *Proceedings of the CESDIS Image Registration Workshop*, NASA/GSFC, Greenbelt, MD (pp. 307-316), and in NASA Pub. CP-1998-206853.

Netanyahu, N., Mount, D. M., and Le Moigne, J. Efficient Algorithms for Robust Feature Matching. *Proceedings of the CESDIS Image Registration Workshop*, NASA Goddard Space Flight Center, Greenbelt, MD. (pp. 247-256), and in NASA Pub. CP-1998-206853; also, a full paper version is to appear in a special issue of *Pattern Recognition* on image registration.

Netanyahu, N., Mount, David M., Le Moigne, J. (1998). Improved Algorithms for Robust Point Pattern Matching and Applications to Image Registration. *Proceedings of the Fourteenth Annual ACM Symposium on Computational Geometry*. Minneapolis, Minnesota.

Netanyahu, N., Le Moigne, J., Xia, W., Chettri, S., El-Ghazawi, T., Pierce, J., Raghavan, S., Tilton, J. C., Campbell, W. J., and Cromp, R. F. (1998). An Image Registration Toolbox: First Evaluation of Automatic Image Registration Methods. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*. Seattle, Washington.

## **Beowulf Parallel Workstation**

**Phillip Merkey, Senior Staff Scientist (merk@cesdis.gsfc.nasa.gov)**  
**Donald Becker, Staff Scientist (becker@cesdis.gsfc.nasa.gov)**  
**Daniel Ridge, Technical Specialist (newt@cesdis.gsfc.nasa.gov)**  
**Erik Hendriks, Technical Specialist (hendriks@cesdis.gsfc.nasa.gov)**

### **Profiles**

#### Phillip Merkey

Dr. Merkey holds a Bachelor of Science degree in mathematics from Michigan Technological University, and took a Ph.D. in mathematics in the area of algebraic coding theory from the University of Illinois (1986). He is a member of the AMS and SIAM.

Prior to joining CESDIS in 1994, Dr. Merkey was employed as a research staff member by the IDA Supercomputing Research Center in a classified working environment. His experience includes application of high performance computers to grand challenge problems, investigation of instruction level parallelism using the VLIW parallel computer, benchmarking experiments on the Multiflow Trace computer, algorithmic design for empirical solutions to problems in applied discrete mathematics, and innovative parallel implementations of advanced algorithms.

Dr. Merkey is the technical lead on the Beowulf Bulk Data Server project. He is responsible for the overall design and progress on the project. He is also responsible for identifying and evaluating applications that will be suitable applications to demonstrate the machine capabilities and guide its development.

Dr. Merkey has also engaged in outside collaborations with the IDA Center for Computing Sciences, he has participated in Thomas Sterling's (Caltech/JPL) Petaflops workshops including studies of applications for the HTMT architecture, and has served as an instructor at the University of Maryland Baltimore County where he is developing a course on parallel and distributed computing.

#### Donald Becker

Mr. Becker holds a Bachelor of Science degree from the Massachusetts Institute of Technology in electrical engineering and has completed graduate computer science courses at the University of Maryland College Park. From 1987 to 1990 he was employed by Harris Corporation, Advanced Technology Department, Electronic Systems Sector as a senior engineer. He performed research and development work on the Concert multiprocessor, maintained and extended the Concert C compiler (based on PCC) and libraries, and wrote network software.

As a research staff member of the IDA Supercomputing Research Center from 1990 to 1994, Mr. Becker wrote a substantial portion of the low-level LINUX networking code, designed, implemented, and characterized an interfile optimization system for the GNU C compiler, implemented a peephole optimizer for a data-parallel compiler (DBC), and implemented several symbolic logic applications.

Since joining CESDIS in 1994, Mr. Becker has been the principal investigator for system software on the Beowulf Parallel Workstation project. He has established a world class reputation in the operating system community with his contributions in networking software. Mr. Becker continues to make CESDIS the center of the networking research community for Linux and Beowulf. He helped develop and has participated in several "How to build a Beowulf" tutorial sessions presented at leading conferences throughout the year.

He is a co-author of the "how to build a Beowulf" book that will be published by MIT Press. Mr. Becker is one of the primary researchers responsible for the Red Hat release of the "Extreme Linux" CD-ROM which defines the Beowulf Software distribution.

### Daniel Ridge

Mr. Ridge is working on his undergraduate degrees in computer science and aerospace engineering at the University of Maryland College Park. He began working with Donald Becker at CESDIS in 1995 and in 1996 took a leave of absence from Maryland to work as a Technical Specialist on the Beowulf project.

Mr. Ridge has shown himself to be among the most important contributors to CESDIS and the Beowulf project. In addition to developing system software for the Beowulf Workstation and the Beowulf Bulk Data Server, Mr. Ridge has also participated in the Beowulf tutorials and made significant contributions to the "Extreme Linux" CD-ROM.

Mr. Ridge left CESDIS early in 1998 for a position in the NASA Inspector General's office where he is engaged in the research and application of the Beowulf technology to address their specific requirements. Mr. Ridge has maintained close contact with CESDIS and the Beowulf community.

### Erik Hendriks

Mr. Hendriks received his Bachelor of Science degree in Computer Science from Johns Hopkins University in 1996. During his graduate studies, he worked for the physics department at Johns Hopkins writing parallel programs.

Mr. Hendriks' primary responsibility is the development of system software for the Beowulf Project. He worked with John Dorband and Udaya Ranawake to combine the HIVE and ecgtheow (the two large Beowulf clusters at GSFC) into the single cluster that first broke the 10 Gflop barrier for Beowulf-class computers. Mr. Hendriks has refined the installation procedure for Beowulf clusters, made significant contributions to the "Extreme Linux" CD-ROM, has conducted an extensive evaluation of the candidate disks for the Bulk Data Server and has developed system software that can access the hardware monitors on the motherboards used in the Bulk Data Server.

In addition to becoming an integral member of the Beowulf team, Mr. Hendriks has shown himself to be a valuable member of CESDIS as well. On numerous occasions he took over responsibilities of the CESDIS system administrator and repaired or installed systems that enabled CESDIS to meet its diverse obligations.

## **Report**

Don Becker is a member of a team awarded the 1997 Gordon Bell Prize for Price/Performance "in recognition of their superior effort in practical parallel-processing research." The award was announced and presented at Supercomputing97. The prize was given for a Beowulf cluster of Pentium Pro's assembled a year earlier at SC96 which achieved 2.1 Gflops/s on an n-body code, the equivalent of \$50,000 per Gflops/s. The code simulates gravitational attraction among particles, such as dark matter in cosmology models. Other award recipients are T.Sterling/JPL/Caltech, M.Warren, P.Goda/LANL, J.Salmon/Caltech, G.Winckelmans/Catholic University of Louvain, Belgium. The award represents a breakthrough in the community which has now come to recognize Beowulf-class systems as an important type of parallel computing. The Gordon Bell Prize winners presented talks on their award-winning work at SC97. The paper "Pentium Pro Inside: I. A Treecode at 430 Gflops/s on ASCII Red, II. Price/Performance of \$50/Mflop on Loki and Hyglac," is available at <http://scxy.tc.cornell.edu/sc97/proceedings/BELL/WARREN/INDEX.HTM>. The

Gordon Bell Prize was established to reward practical use of parallel processors by giving monetary awards for the best performance and best price/performance on an application, and for automatic compiler parallelization. The award is sponsored by the IEEE Computer Society and IEEE *Computer* magazine.

Becker collected and organized all of the software and documentation that is required to construct and operate a Beowulf cluster onto a CD-ROM mirror [see: <http://www.beowulf.org>]. This is significant because it provides a complete distribution of the Beowulf package. Moreover, it has been formatted so that one can boot and install a Beowulf cluster directly from this image, greatly improving the current method of augmenting and patching a Linux distribution. The "Extreme Linux CD," as it is called, is important to CESDIS and to the Beowulf because it is a concrete place to the Beowulf software effort. The Red Hat version of this material was prepared in the late spring and the Beowulf "Extreme Linux CD" had its debut at Linux Expo at Duke. The "Red Hat, NASA Team on Beowulf Tech CD-ROM Price Under \$30" was the number one requested article on HPCwire, 5/15/98.

The Beowulf Project continues to spread throughout the world, and CESDIS continues to maintain a leadership role in the development of Beowulf-class Cluster Computing. In addition to the CD, there is now a Website for the Project that is independent of any particular site or organization: <http://www.beowulf.org>. It is currently maintained by the CESDIS group. The CESDIS contributions to the project are now on the site: <http://beowulf.gsfc.nasa.gov>. These websites and associated mailing lists that are maintained by CESDIS continue to provide a focus for the activities of the Beowulf community.

Becker served on the program committee and as a session chair for the first NASA workshop on Beowulf-class Cluster Computing. This meeting helped develop a sense of unity and direction for the diverse groups across NASA and other agencies that make up the Beowulf community. This meeting also helped identify the areas of expertise. CESDIS will continue to be the center of activity in network research and through its web presence should continue to be the repository for the Beowulf software and Beowulf technology.

Consistent with this agenda, Becker continues to enhance Ethernet drivers for use in Beowulf cluster. We have also met with representatives from leading network vendors. For example, we met with HAL Computer Systems and negotiated an agreement to develop Linux drivers for their interconnection hardware and then evaluate that hardware on the Beowulf Bulk Data Server. In another instance, the Packet Engines' Gigabit Ethernet adapters have been installed in a pair of Alphas with 64-bit PCI slots. This provided the first opportunity to test the performance of Gigabit Ethernet cards and driver software at their full capability.

Erik Hendriks helped CESDIS contributed to the GSFC record of 10.2 Gflops on the Piecewise Parabolic Method code. The two GSFC Beowulf clusters, the HIVE and ecgtheow, were connected to form a single for the purposes of the experiment. Obtaining a rate above 10 Gflops is significant within the ESS community. The second phase of the ESS program is a milestone-driven program with the first milestone being 10 Gflops. In other words, the Beowulf-class cluster computers have reached a performance level that is considered high performance computing from the ESS perspective.

Becker was a co-instructor at numerous tutorials on the construction of Beowulf Clusters including an all day tutorial at SC'97 which received the distinction of being the most highly attended tutorial of the conference. This tutorial was given several times across the country including one-day workshops at Pasadena and at Florida Institute of Technology.

The Beowulf project was presented (by Donald Becker) as a keynote session at IEEE Aerospace '97 and a CESDIS/JPL/Caltech collaboration produced a tutorial for the Cluster Computing Conference in Atlanta.

Becker participated in the "Extreme Linux" workshop, a by-invitation-only workshop of the core Linux developers. One of the main topics on this year's agenda was Linux and clusters; for most in attendance clusters means Beowulf-class machines.

In addition to the presentations, tutorials, and published articles listed above, the Beowulf project has built and maintains a significant Web presence as its primary means of technical transfer.

Phil Merkey is developing a course on parallel and distributed computing based on the Beowulf technology. This course was given in the fall semester at UMBC. After discussing parallel computing from an academic point of view, the students were given accounts on the Beowulf cluster called hrothgar. This "lab" component of the course provides hands-on experience with parallel programming and debugging parallel programs and put the abstract analysis of parallel programs in a more tangible framework.

Erik Hendriks is responsible for numerous enhancements to the kernel and the system software that are critical to the construction of large clusters. For example:

- Modified the BIOS image on the Intel PR440 FX motherboards to allow netbooting: this allows the nodes in a cluster to be stateless at boot time.
- Developed a kernel performance counters package for the Pentium Pro: this hardware information proves very useful for debugging and performance tuning.
- Modified disk reads and writes to fully exploit three IDE disks on three separate channels: this bandwidth is required to meet the design specifications for the Bulk Data Server.
- Released a Linux driver for the LM78 hardware monitor this is on-board hardware monitor for the motherboards used in the Beowulf Bulk Data server: this driver provides a /proc interface that allows easy reading of current status and easy manipulation of limit registers. These hardware monitors will become more and more important as clusters get bigger and bigger.

The Beowulf Bulk Data Server has been upgraded to meet its phase two goals. The cluster currently has 100 Intel P6 processors running at 200 MHz and 7.6 Gbytes of memory. The cluster is connected in a fat tree network topology with Packet Engines Gigabit Ethernet at the root of the tree. Through sponsorship and collaboration with the team at Clemson University headed by Dr. Walter Ligon, CESDIS is meeting its milestones on the development and demonstration.

## **HPCC/ESS Evaluation Project**

**Terrence Pratt, Senior Scientist  
(pratt@cesdis.edu)**

### **Profile**

Dr. Pratt earned B.A., M.A., and Ph.D. degrees in mathematics and computer science at the University of Texas at Austin. He is a member of the ACM, the IEEE, and SIAM. In 1972-73 he served as an ACM National Lecturer, and in 1977-78 a SIAM Visiting Lecturer. His research interests include parallel computation, programming languages, and the theory of programming.

Prior to joining CESDIS, Dr. Pratt held teaching and research positions at Michigan State University in East Lansing, the University of Texas at Austin, and the University of Virginia. At the latter he was one of the founders of the Institute for Parallel Computation and served as its first director.

During the 1980s, Dr. Pratt worked with scientists at USRA's ICASE and NASA Langley on the development of languages and environments for parallel computers. He is the author of two books: *Programming Languages: Design and Implementation* (Prentice-Hall, second edition, 1984) and *Pascal: A New Introduction to Computer Science* (Prentice-Hall, 1990).

Dr. Pratt joined CESDIS as the Associate Director in October 1992 and was appointed Acting Director in October 1993 upon the retirement of Raymond Miller. He served in that capacity until November 1994 when he left CESDIS to pursue other interests, but maintained ties with CESDIS as a consultant on high performance Fortran. He rejoined CESDIS as a Senior Scientist early in 1996.

### **Report**

This research project is part of the NASA HPCC Earth and space science (ESS) project centered at Goddard. The ESS project funds nine "grand challenge" science teams at various universities and federal research laboratories. In addition, through a cooperative agreement with SGI/Cray, a 512 processor SGI/Cray T3E parallel computing system has been placed at Goddard to serve as a testbed system in support of the science team projects. During 1998, this system was upgraded by NASA to 1088 processors.

Each science team is responsible for developing large scale science simulation codes to run on the T3E and meet specified performance milestones (10 Gflop/sec in 1996, 50 in 1997, 100 in 1998). The codes are provided to an in-house science team at Goddard for performance verification, and ultimately the codes are submitted to the National HPCC Software Exchange for general distribution. For an overall view of the NASA HPCC/ESS project and its current status, visit the web page at <http://sdcd.gsfc.nasa.gov/ESS/>. For the current status of the project reported here, go to that web page and click on the "System Performance Evaluation" icon to get to the homepage for this project.

### **1. Research Goals**

The CESDIS System Performance Evaluation Project is concerned with the large scale science simulation codes produced by the nine Grand Challenge science teams, their behavior on the massively parallel testbed computer system, and to a lesser extent their behavior on other parallel systems such as the CESDIS and NASA Beowulf systems. We expect to work with about 10-15 different science codes in total.

Our interest is in understanding how these large science codes stress the parallel system and how the parallel system responds to these stresses. In particular, we wish to find ways to:

- Quantify the stresses produced by the science codes on the testbed hardware and software
- Quantify the performance responses produced by the system.
- Determine the causes of the observed responses in the codes and systems.
- Use the results to improve codes and systems.
- Develop new performance evaluation and prediction methods and tools as needed.

Ultimately the goal is publication of the results of this work in various journals and conference proceedings.

## 2. Approach

Our approach is to work directly with the science codes as they are submitted by the science teams to meet performance milestones. We use various measurement tools to understand the static structure of each code and its dynamic behavior when executed with a typical data set (also provided by the science team). Typically, a code is "instrumented" to collect the desired statistics and timings, and then run on the testbed system using various numbers of processing nodes. The results are analyzed, and if more data are required, the instrumentation is modified and the code rerun.

The insights gained from this research on a particular code often lead to understandings about how to improve the performance of the code. These insights are fed back to the science team to aid them in further development of the code. Results may also be useful to SGI/Cray in improving their hardware and software systems, so results are often forwarded to the in-house SGI/Cray team and the in-house science team.

## 3. Measurements of Interest

Part of the research effort is to determine what aspects of science code structure and behavior have the greatest effect on performance. To this end, we are measuring some of the following elements in each code:

- Flops counts and rates.
- Timings and execution counts of interesting code segments.
- Data flows between code segments.
- MPI/shmem/PVM message passing and synchronization profiles.
- I/O activity profiles.
- Cache use issues.
- Storage allocation sizes and use profiles.
- Scaling with problem size and number of processors.
- Load balance.

## 4. Tools Used

These studies use a variety of tools for instrumenting and measuring various characteristics of the science codes and their behavior. The primary tool to date has been a software system called Godiva (GODdard Instrumentation Visualizer and Analyzer) developed by this project. We also use the SGI/Cray Apprentice and PAT software tools on the T3E and are investigating other tools from universities and national laboratories that might prove of use, such as Pablo from the University of Illinois and AIMS from the NASA Ames Research Center.

## 5. Current Status and Results

The 1997-98 research year included several major developments. In 1997, working with NASA management, we revised the goals and milestones for this project to more accurately reflect the changes in direction of the activity that have taken place since 1996, due to changes in project leadership and resources. The new project plan includes three milestones, to be met at the end of each of FY 97-99.

The first milestone, which was successfully completed in September 1997, was the development of appropriate software tools to support the project. The Godiva software (discussed below) was completed to Version 3.4. This version has remained stable and in use since September. During the year, two research papers on this software work were accepted for conference presentation and one of the papers was presented.

The second milestone, targeted for September 1998, is the quantification of the "stress patterns" produced by three of the Grand Challenge science team codes and their use in comparing the performance of the HPCC testbed systems. We have made considerable progress toward this goal, including (1) the development and testing of methods for quantifying and displaying the stress patterns produced by science codes, and (2) an initial performance comparison of a code running on both the T3E and the Beowulf-class machine called "the Hive", situated at Goddard.

The third milestone, targeted for September 1999, is the quantification of the stress patterns in most of the GC codes developed by the project science teams, and a demonstration of the use of these stress pattern data in making performance comparisons among various large-scale parallel systems that serve (or might serve) as HPCC testbeds.

In addition to successfully completing the first milestone and making good progress toward completing the second, we have moved to increase the visibility and impact of this work in two ways: (1) publication of the results of the work (two papers mentioned above and discussed below) and (2) a complete revision of the web pages for the project, including regular updates to reflect the latest developments and web publication of postscript versions of useful documents, such as preprints of conference papers and a full version of the Godiva Users Manual. The project web pages are accessible both from the NASA HPCC/ESS project homepage and from the CESDIS homepage.

## 6. Godiva Software Instrumentation Tool

The Godiva software system, developed as part of this project, has proven to be a useful new tool for the study of large science codes. Using Godiva, a wide variety of aspects of a code may be instrumented so that the dynamic behavior may be observed as the program executes. Of particular importance to date have been the ability to study cache behavior on the T3E, computation (flop/sec) rates in selected code segments, parallel communication and synchronization profile using MPI, PVM, or shmem library calls, and load balance among processors.

Godiva has been developed as a personal research tool, not intended for general distribution, but it has been made available to other researchers within the NASA HPCC/ESS project. Because it is a personal research tool, it undergoes frequent change to meet the demands and new directions of the evaluation project.

The approach to code instrumentation used in Godiva is as follows. First, selected parts of the code are annotated to study whatever characteristics are of interest. These annotations use a syntax specified in the Godiva Users Manual. Annotations appear as comments to a Fortran or C compiler. The annotated code is fed through the Godiva preprocessor, which generates Fortran or C source code with calls to the Godiva run-time library inserted at appropriate points. The generated source program is then compiled and linked with the Godiva run-time library. Execution of the program generates a trace file on each processor. The trace file contains statistics collected on-the-fly during execution. After execution is complete, a Godiva postprocessor is used to generate tables, graphs, and histograms from the trace files produced by the processing nodes.

Currently Godiva supports about 30 different annotation types in the source program. These annotations may be used to generate about 20 different forms of output tables and graphs. Version 3.4 of Godiva has been stable and in use since September 1997. Version 4.0 is under development.

Two papers on the Godiva design have been accepted for conference presentation, see [1] and [2] below. The complete Godiva Users Manual [3] is available on the web.

## 7. Conclusion

The evaluation project is proceeding well. The Godiva software tool is proving useful, and good access to large scale science codes and to the T3E has been provided by the NASA HPCC/ESS Project. Collaborations with several members of the SGI/Cray in-house team, the Goddard ESS in-house team, and the members of the science teams have begun to develop. Useful small-scale results have been produced and disseminated. The outlines of more general insights and results are beginning to emerge. Publication of early work has begun.

## Publications

All publications are available on the web at [http://sdc.d.gsfc.nasa.gov/ESS/system\\_eval.html](http://sdc.d.gsfc.nasa.gov/ESS/system_eval.html)

[1] Pratt, T. (1998). Design of the GODIVA performance measurement system. *LCR98: Fourth Workshop on Languages, Compilers, and Run-time Systems for Scalable Computers*. Pittsburgh. (to appear in the Springer Lecture Notes in Computer Science series)

[2] Pratt, T. (1998). Using GODIVA for data flow analysis. *Second SIGMETRICS Symposium on Parallel and Distributed Tools*. Welches, Oregon.

[3] Pratt, T. (1997). Godiva Users Manual, Version 3.4. CESDIS.

## ***Highly-parallel Integrated Virtual Environment (HIVE)***

**Udaya A. Ranawake**  
**University of Maryland Baltimore County**  
**Department of Computer Science and Electrical Engineering**  
**(udaya@neumann.gsfc.nasa.gov)**

### **Profile**

Dr. Ranawake received a B.Sc. degree in Electrical Engineering from the University of Moratuwa, Sri Lanka in 1982, and an M.S. degree in Electrical Engineering and a Ph.D degree in Computer Engineering from Oregon State University in 1987 and 1992 respectively. Prior to joining CESDIS on a subcontract with the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County, he was a senior member of the technical staff at Hughes STX Corporation where he was the task leader for massively parallel research at NASA GSFC. His research interests are algorithms for scientific computation, parallel and distributed computing, computer architecture, and computer networks. Dr. Ranawake is a member of the IEEE.

### **Report**

#### **1. Introduction**

The rapid increase in performance of commodity microprocessors and networking hardware has provided the opportunity for exploring the potential of Pile-of-PCs (PoPC) as a low cost alternative to high end supercomputers in scientific computations. The PoPC model is used to describe a loose ensemble or cluster of PCs applied in concert to a single problem. It is similar to a network of workstations (NOW), but emphasizes the use of mass market commodity components, dedicated processors, and a private system area network (SAN).

In early 1994 the Beowulf project was initiated under the auspices of the NASA HPCC Earth and Space Sciences project to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA Earth and space science computational needs. The Beowulf project is based on the PoPC model and adds to it by emphasizing no custom components, easy replication from multiple vendors, a freely available software base, and a return of the design and improvements to the community.

The Beowulf class systems have emerged as a complementary computing medium to high end supercomputers. As they are based on the PoPC approach, these systems use hardware components that benefit from declining prices resulting from heavy competition and mass production. This approach also permits technology tracking allowing computing systems to be acquired with the best, most recent technology and at the lowest price. As the systems are not preconfigured by a vendor, Beowulf-class systems also permit the configuration of individual systems to suit user needs. Also, the free software base available for these systems is quite robust and as efficient as commercial grade software.

#### **2. Overview of the Hive**

The HIVE project's goal is to produce an inexpensive high performance parallel computer that is reliable and easy to use. This project is sponsored by the Mission to Planet Earth and NASA's Office of Space Science Advanced Technology. Therefore, the primary applications on the HIVE will be Earth science data manipulation, space data image restoration, ocean and atmosphere modeling, and other related applications.

The HIVE is a Beowulf class computer consisting of 64 nodes. Each node is a dual 200 Mhz Pentium Pro rack-mounted PC containing a total of 128 processors. Two additional PCs are used as hosts: a system host and a user host. The purpose of the system host is to maintain and monitor the HIVE. The user host is intended for application development and job submission to the HIVE. The HIVE is interconnected with five 16 port fast Ethernet switches for a maximum aggregate interprocessor communication bandwidth of 6.4 Gbits/second. It also contains 4 Gbytes of RAM and 160 Gbytes of disk storage distributed across the nodes.

### **3. Accomplishments**

#### **3.1 Building and Software Configuration of the HIVE**

As the co-investigator of the HIVE project, I played an active role in the design, building, and software configuration of the HIVE. The HIVE was built in July and August of 1997. Since then, it has supported the computational needs of a number of NASA users. The system has been highly reliable, and has experienced only a few node crashes. The HIVE software environment includes programming languages such as C, C++ and aCe, and interprocess communication software packages such as PVM, MPI, and BSP.

#### **3.2 The Piecewise Parabolic Method (PPM) Program**

The PPM program was implemented on the HIVE. It is a very high-resolution algorithm, which is particularly well suited for studying flows containing discontinuities. It can be used for simulations of various astrophysical systems such as supernova explosions, accretion, and supersonic jets. The PPM method is a finite volume technique in which each grid point uses the information at 4 nearest grid points along each spatial dimension to update the values of its variables.

The parallel implementation is based on the PROMETHEUS computer code, which solves Euler's equations for compressible gas dynamics on a logically rectangular grid. The algorithm was parallelized using domain decomposition where the grid is subdivided into rectangular tiles and one or more tiles are assigned to each processor. Each tile consists of a section of real zones surrounded by a frame of ghost points four zones wide. The boundary conditions are handled using the ghost zones.

Communication overhead is minimized by overlapping computations and communications. Each processor first updates its boundary tiles and sends the boundary values of these tiles to the neighbor processors. The interior tiles are updated next and their boundary values are copied to the ghost zones of the appropriate tiles. Finally, each processor reads the messages received from its neighbors and copies this data to the ghost zones of the boundary tiles. This program delivered 7.3 Gflops on 128 processors of the HIVE. The Mflop rate was obtained by using the operation count from the Cray C-90 hardware performance monitor and the wall clock execution time.

#### **3.3 The bview Software Tool**

The bview software tool was implemented on the HIVE. This program can be used to display the cpu and memory usage statistics of all the nodes of a Beowulf-class cluster of PCs. The information is displayed in the form of a bar chart with one entry for each node in the system. The delay between screen updates may be set at the time the software is configured. One may determine which bar belongs to which node by placing the cursor over that bar. This will cause a window to appear which will contain the name of the node. The status window also allows one to open a shell window on any node by clicking on its respective bar. Commands such as top may be executed within this window to obtain a more detailed view of the resource usage on a node.

The heart of the 'bview' software tool is a daemon called 'bstat' that runs on each node of the PC cluster to collect statistics on CPU and memory usage. The communication between the daemon processes is done via sockets using algorithms that employ a logarithmic number of communication steps. Preliminary studies on the 64 node HIVE computer have shown that the 'bstat' daemon incurs negligible overhead when collecting statistics at 1 second intervals. The user interface part of the 'bview' software tool is implemented using TCL/TK. This software is available as part of the HIVE software archive under <http://newton.gsfc.nasa.gov/thehive>.

### **HPCC/ESS Project Scientist**

**George Lake  
University of Washington  
Department of Astronomy  
(lake@astro.washington.edu)**

## **1. Introduction**

As project scientist, I set several goals:

1. Greater integration of the HPCC/ESS program into NASA missions and strategic planning.
2. Increased connections between the Science Teams and the Projects In-house Team.
3. Research in two areas that were neglected in the current program: extragalactic astronomy and the origin of planetary systems.

Significant progress has been made on all these fronts in the past year.

## **2. HPCC/ESS Programmatic Issues**

Over the last two years, we've worked with the Grand Challenge (GC) Team leaders to provide materials that better capture their work. The original Project Milestones (with an intense focus on Gigaflops) have proven useful to both the GC Teams and the Project Management, but they are ill-suited to capturing scientific successes. The Team leaders have now formulated metrics that also speak to how they are guiding the technology developments with tools that enable new scientific breakthroughs that will advance NASA Enterprise Strategic Plans. They have prepared concise slide sets that capture these goals and successes. In April 1998, the Project held a Science Colloquium at NASA HQ where this could be reported in greater depth.

With the In-house Team, attention has focused on means by which they can operate as a Team, work closely with the GC Teams, and work with the broader NASA community that has access to the HPCC/ESS Testbed (the T3E-512, recently upgraded to a T3E-1024).

Significant progress has occurred in indentifying Team projects that can have broad impact to NASA computational problems. The main areas where this has occurred are in the work developing a flexible adaptive mesh code and in the area of visualization.

Visits to the GC Teams and hosting of GC Team members at Field Centers has been strongly encouraged and now occurs at a rate greatly elevated from past years. Code 930 has been undergoing a transformation toward a computational science organization and away from a strictly service organization. In past years, the service orientation made the scientific charge of the In-house Team a bit hazy. To clarify their status as scientists, they were strongly encouraged to seek co-funding from other organizations. This had the additional goal of insuring that their unique expertise in high performance computing would be spread through the center and enhance the computational science community at GSFC. Like most changes, these were slightly threatening to morale at first, but have proven to be extremely effective. There is strong praise for the work of the In-house Team from both the GC Teams and the broader GSFC community. There is also a need to restaff the team based on the large funding base and slight attrition.

### 3. Scientific Work: Planet Formation

Planet formation in the inner Solar System is thought to proceed in four stages:

- Condensation and growth of grains into fluffy aggregates;
- Formation of km-sized planetesimals via pairwise accretion in the turbulent gas disk;
- Agglomeration of protoplanets by focused merging and runaway growth;
- Final incorporation into planets through slow perturbations.

To date most numerical studies have been restricted to the third stage and relied on statistical methods or small computational domains to make the problem tractable. Our group at UW has designed numerical methods capable of modeling both the third and fourth stage, that is, the transition from runaway growth to the regime of the infrequent large impacts leading to the final formation of the planets.

In order to start with planetesimals of an "interesting" size (about 100 km) and to model the entire inner Solar System disk, millions of planetesimals are required. This is many orders of magnitude more than any previous study has attempted. Furthermore, it takes as long as a million years to form protoplanets, and detecting collisions among the planetesimals requires timesteps of days.

We are approaching this goal of large computational domains, high spatial and temporal resolution needs, and long integration time. We have modified our stable cosmology code to search for particle collisions and to optimize the orbital integration for the central force field of the Sun. We have test results from a run on a Cray T3E using 128 nodes for about 24 wallclock hours. The test consisted of 1 million planetesimals in a thin cold disk around the Sun, and included the effects of the giant planets. In 100 years Jupiter has already begun to carve out resonance structure in the disk. Meanwhile, particles in the Earth-Mars region have started to agglomerate on the way to building planets.

A two year old survey of this field projected the kinds of simulations that might be possible over the next decade. In these projections, they estimated that a simulation like our modest test case could be done in the Year 2002 if an ambitious program of special purpose hardware was undertaken. Without such hardware, they estimated that general purpose computers would not be able to do this before the Year 2007. The algorithms developed by us during the last 18 months represent a leap that was not expected for a decade.

We are currently developing and testing a perturbative technique to increase the speed of our integrations by another factor of 10 to 100. However, to perform simulations of >10 million planetesimals for >10 million dynamical times, we need roughly a further hundred-fold improvement in the combined speed of computers and algorithms. We expect that this will be possible within a couple of years.

**Adam Frank**  
**University of Rochester**  
**Department of Physics and Astronomy**  
**(afrank@alethea.pas.rochester.edu)**

I joined the HPCC/ESS program in the Winter of 1997. I was asked by George Lake to help with the outreach effort. Since I am both a computational scientist and a science writer, he felt I might be able help place stories in national media outlets. My goal for the consulting agreement has been to get HPCC-related feature articles in magazines with national circulations. These stories satisfy the publishers' need for good stories about fundamental science as well as the HPCC need to show that the research it sponsors is exciting and useful.

I believe I have had some success in reaching these goals. My efforts have brought the work of the program and its researchers into the eyes of the public. The major achievement so far this year has been the placing of four stories focused on HPCC researchers in popular magazines with large reader bases.

My first work for the HPCC program was a piece on Chaos and the Solar System for *Astronomy* magazine. This article included interviews with George Lake and his team. I focused on the science of planetary stability as well as the need for high performance computing. The article was published in the May 98 issue.

My next work was a story for *Earth* magazine on Peter Olson and the geodynamo work of Glatzmeyer and Roberts. The article focused on the difficulty of knowing what processes govern the development of the inner core and geomagnetism. The story was published in the Feb 98 issue.

This spring *Astronomy* magazine accepted a story on Space Weather which focused on Gambosi's and Gardner's group. I was able to include numerous quotes by a number of HPCC scientists in the piece. The story should come out in the fall.

Currently I am working on a story for *Discover* magazine focusing again on the Geodynamo (Olson's group). I am including interviews with Peter Olson and Gary Glatzmeyer in the piece. *Discover* has a readership of more than 2 million.

For the next period I will continue to search for outlets for articles on HPCC researchers. I am hoping to place stories on the Curkendall, Malagoli, Carey, and Saylor groups. Possible outlets include *National Geographic* for a Curkendall story, *Discover* for a story on Carey's work, and *Sky and Telescope* for an article on Saylor.

## **Wavelet-Based Parallel Image Registration**

**Tarek El-Ghazawi**  
**George Washington University/Florida Institute of Technology**  
**now at George Mason University**  
**Institute of Computational Sciences and Informatics**  
**(tarek@science.gmu.edu)**

### **Statement of Work**

Dr. El-Ghazawi was tasked with supporting the development of high-performance implementations of wavelet-based processing of NASA Earth Science Imagery. This included:

- Developing sequential wavelet-based registration for integration into the NASA image registration toolbox, for regional application centers, and studying the performance of variations of the algorithm using a selected suite of NASA imagery;
- Developing efficient wavelet-based coarse-grain algorithms for massively parallel systems with focus on reducing communication latency and load imbalance.

### **Profile**

Tarek El-Ghazawi received the Ph.D. in electrical and computer engineering in 1988 from New Mexico State University. Recently he joined the faculty of George Mason University where he will hold a tenure track joint appointment as an Associate Professor of computational sciences and computer engineering in the Institute for Computational Sciences and Informatics and in the Department of Electrical and Computer Engineering. Previously Dr. El-Ghazawi was a member of the Department of Electrical Engineering and Computer Science at George Washington University and also taught at the Florida Institute of Technology and Johns Hopkins University.

Dr. El-Ghazawi's research interests include high performance computing, experimental computer architectures, high performance I/O systems, experimental performance evaluation, and computer vision. His research has been supported by NASA, the Army Corps of Engineers, and Computer Science Corporation, and he has had more than 50 refereed papers published. He served as the workshop chair for Frontiers '95 and as the program co-chair for the International Conference on Parallel and Distributed Computing and Systems in 1991. Dr. El-Ghazawi is a Senior Member of the IEEE and a member of the ACM and Phi Kappa Phi.

### **Report**

Image registration is a key operation in processing remote sensing data from NASA's Earth observing satellites. Fast and accurate image registration can be obtained using the wavelet transformation techniques. Image registration can benefit greatly from parallel processing on contemporary massively parallel architectures. In parallel image registration, however, the two main sources of overhead that can impede scalability are communications and load imbalance. This study focuses on the parallel algorithm trade-offs between communications and load imbalance overheads on selected NASA ESS testbeds.

## 1. Sequential Image Registrations for the Toolbox

Extensive work has been done by Jacqueline Le Moigne in the area of wavelet-based image registration. In this work we have adopted her basic algorithm and implemented it for integration into the NASA image registration toolbox, for regional application centers and in order to study the effectiveness of using different sub-bands for registering images from a number of NASA satellites and other data. Our experimental work have shown that the wavelet technique works effectively for both rotational and translation changes in remote sensing data and is quite fast as compared to other correlation-based techniques. In specific, it was shown that LL wavelet coefficients work best for photographs and for the Thematic Mapper images. In the cases of GOES and AVHRR, the combined LH-HL techniques with thresholding works best.

## 2. Parallel Image Registration

Figure 1 demonstrates the parallel work that can be done in a single iteration from the iterative refinement algorithm for image registration, in case of translation and rotation. Essentially, a lower-resolution input image (or its wavelet representation) is correlated with the equivalent reference image for every possible x-translation, y-translation, and rotation combination. The correlation maxima is used to determine the best match for this iteration. In the next iteration, this maxima is used as an initial guess and its neighborhood is searched in the same manner but with a higher resolution. The process continues until searching a very small neighborhood in the full resolution image is done. For such a sequential algorithm, we consider two different parallel mappings onto massively parallel systems, a MIMD style (or coarse grain) mapping in which each processor gets an integral number of image correlations to do and a SIMD style (or fine-grain) in which all processors collaborate on each image correlation operation. We note that while the MIMD style mapping creates load imbalances, due to the large data granularity, the SIMD mapping creates excessive communications. Based on that we have suggested a mixed-mode approach.

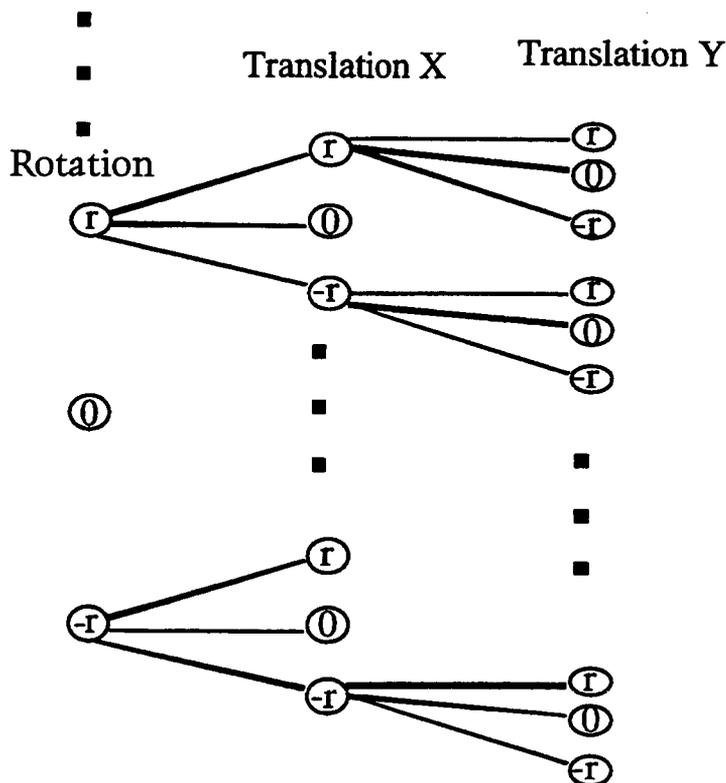


Figure 1: Execution Threads in the Image Registration Iterative Refinement Algorithm

The mixed-mode parallel implementation gives the greatest common integral number of full image correlation operations to each of the processors, and gets the processors to fully collaborate in each of the remaining full image correlations. Our experimental results have shown that the fine grain mapping on MIMD massively parallel computers, as expected, cannot scale as the communication requirements grow very rapidly with the growth in the number of processors. However, if the interprocessor networking infrastructure is quite fast as compared to the processor speed, then the mixed mode mapping works better than the MIMD (coarse-grain) mapping. In other words, when the communication system overhead is small, the benefit from the load balancing in the mixed mode becomes apparent. However, when the communications infrastructure is poor, the additional communication overhead from the mixed mode becomes dominant and overshadows the benefit from load balancing giving rise for the MIMD mapping to perform better. This can be seen from Figure 2 which compares the scalability of the MIXED mode approach to that of the MIMD mode on the Cray T3D, which is known to have an excellent communications network. Figure 3, however, provides the same comparison for the Hive prior to increasing the memory per node from 64M to 448M this summer. The performance of the Ethernet-switch(s) based interconnection makes the communication overhead clear, giving rise to the MIMD mapping to scale better than the mixed mode. After increasing the memory, Figure 4 surprisingly demonstrates a change in this trend at least initially, up to 32 processors. We believe that this could be due to the improved communication bandwidth as a result of the added memory. Additional measurements may be considered to verify this belief. Beyond 32 processors, it seems that the increase in communications overshadows the benefit from load balancing in the mixed mode case.

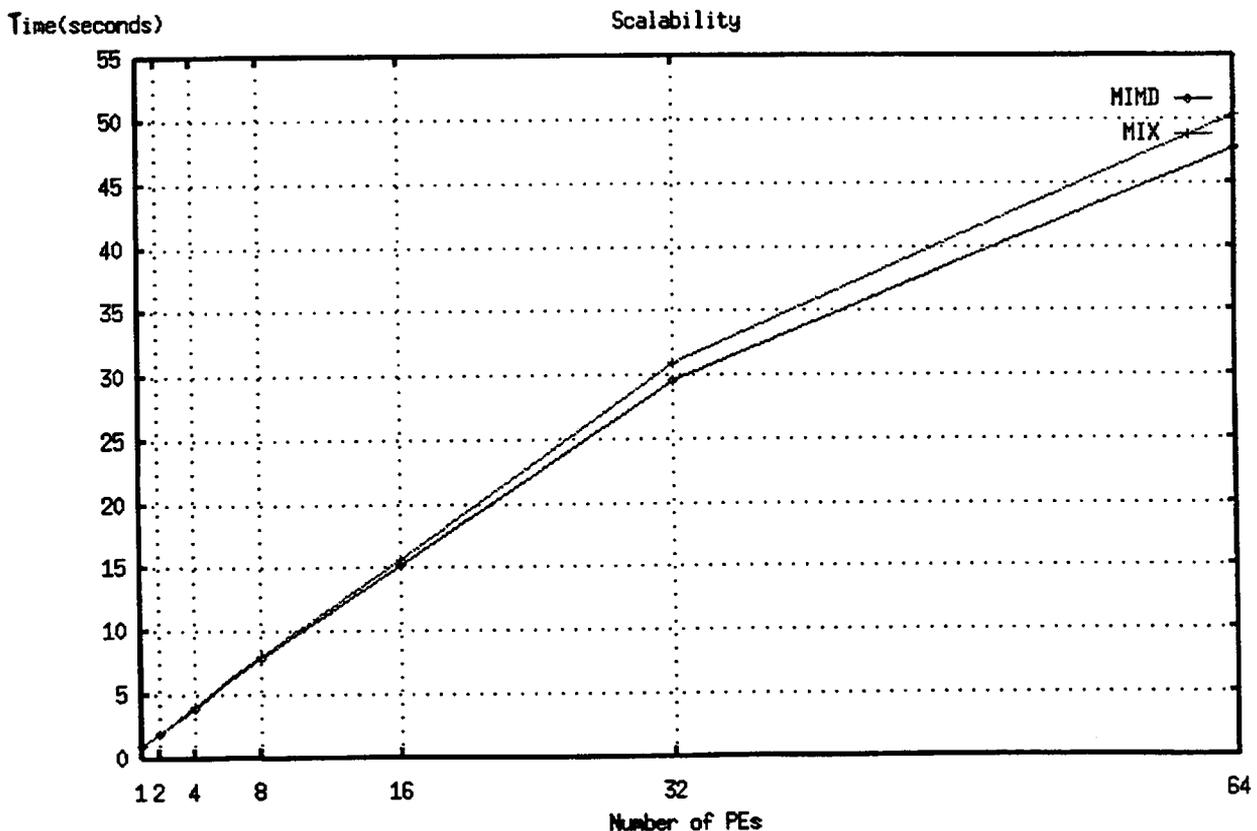


Figure 2: Scalability of Parallel Image Registration Over the Cray T3D

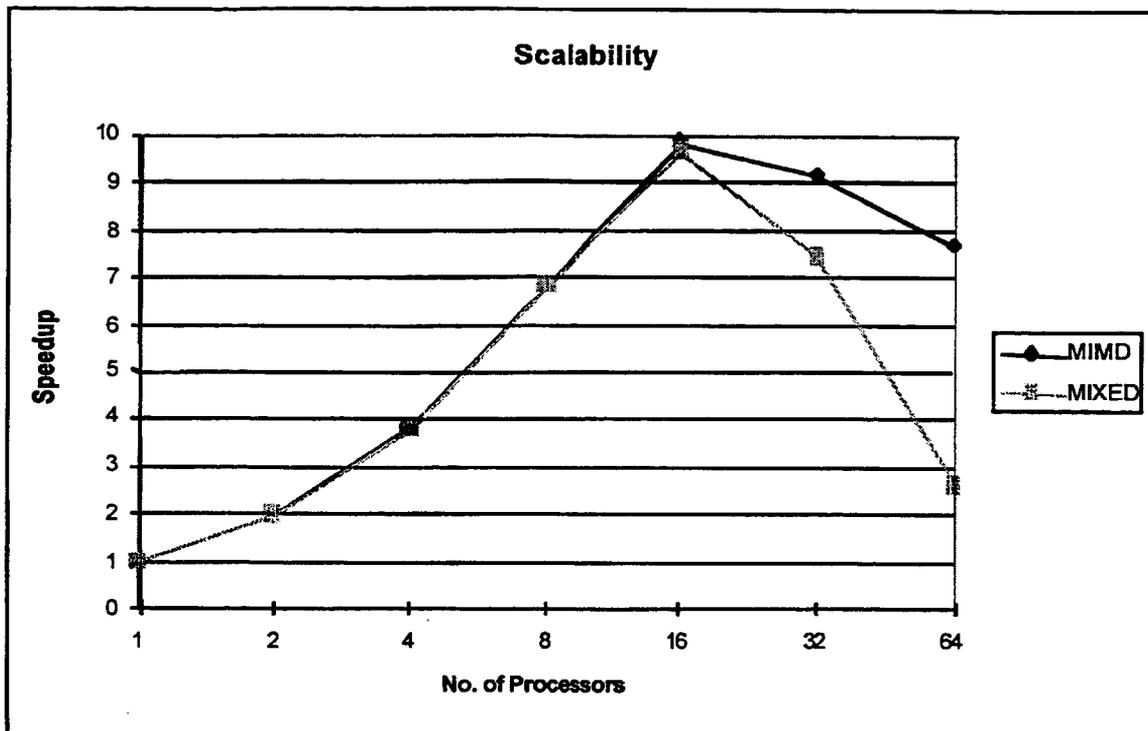


Figure 3: Scalability Over the Hive with 64MB/node

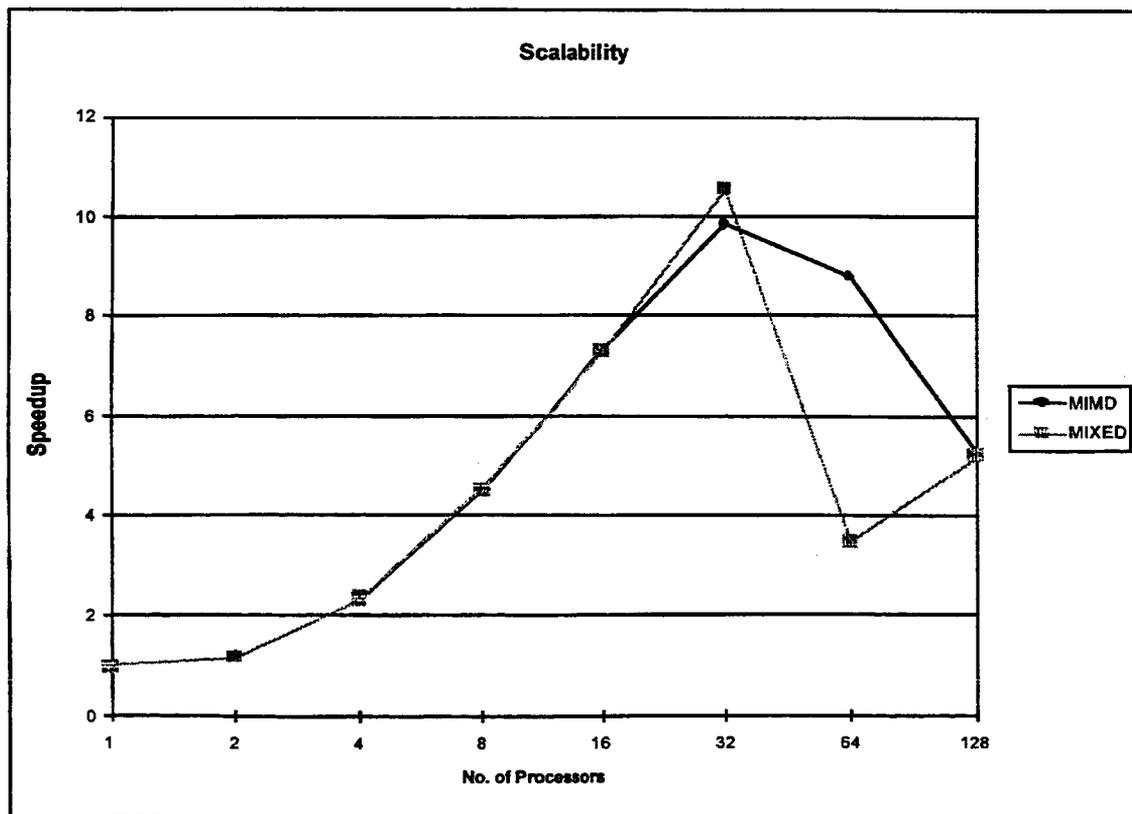


Figure 4: Scalability of Image Registration on the Hive with the Memory Upgrade

## Publications

Le Moigne, J., Xia, W., Chalermwat, P., El-Ghazawi, T., Mareboyan, M., Netanyahu, N., Tilton, J., Campbell, W., and Cromp, R. (1998). First Evaluation of Automatic Registration Methods. *IEEE IGARSS98*. Seattle, WA.

El-Ghazawi, T., Chalermwat, P., and LeMoigne, J. (1997). Wavelet-Based Image Registration on Parallel Computers. *Proceedings of Supercomputing'97*. San Jose, CA.

Chalermwat, P., El-Ghazawi, T., and Le Moigne, J. (1997). Image Registration by Parts. *Image Registration Workshop*. Greenbelt, MD.

Xia, W., Le Moigne, J., Tilton, J., Lerner, B.-T., Kaymaz, E., Pierce, J., Raghavan, S., Chettri, S., El-Ghazawi, T., Manohar, M., Netanyahu, N., Campbell, W., and Cromp, R. (1997). A Registration Toolbox for Multi-Source Remote Sensing Applications. *EOEI'97*. Alexandria, Egypt.

Xia, W., Le Moigne, J., Tilton, J., Lerner, B.-T., Kaymaz, E., Pierce, J., Raghavan, S., Chettri, S., El-Ghazawi, T., Manohar, M., Netanyahu, N., Campbell, W., and Cromp, R. Towards an Intercomparison of Automated Registration Algorithms for Multiple Source Remote Sensing Data. *Image Registration Workshop*. Greenbelt, MD.

El-Ghazawi, Tarek and Le Moigne, Jacqueline. (1996). Wavelet Decomposition on High Performance Computing Systems. *Proceedings of the 25th International Conference on Parallel Processing (ICPP'96)*. CRC Press. Bloomington, IL.

El-Ghazawi, T. and Le Moigne, J. (1994). Multiresolution Wavelet Decomposition on the MasPar Massively Parallel System. *Journal of Computers and Their Applications*, 1(1).

Berry, Mike and El-Ghazawi, Tarek. (1996). Parallel Input/Output Characteristics of NASA Science Applications. *Proceedings of the International Parallel Processing Symposium (IPPS'96)*. IEEE Computer Society Press. Honolulu, Hawaii.

## Parallel Wavelet-based Image Registration on the Beowulf Architecture

Jules Kouatchou  
 George Washington University  
 Department of Mathematics  
 (kouatcho@math.gwu.edu)

### 1. Performances of ScaLAPACK Routines on the CRAY T3E

The objective of this Task was to present the performance of some ScaLAPACK routines on the Cray T3E. More precisely, for different types of linear system of equations of various sizes, we tested ScaLAPACK routines to solve those systems and record the elapsed times required to obtain solutions when the number of processors varies. We utilized band matrices as well as dense matrices for the study.

The following routines (that are part of Cray Scientific Library) are employed:

- LU decomposition and solution of linear distributed systems of real linear equations (PSGETRF/PSGETRS)
- Cholesky factorization and solution of real symmetric distributed systems of linear equations (PSPOTRF/PSPOTRS)
- Inversion of distributed matrices (PSGETRI)
- Eigenvalue solver for real symmetric distributed matrices (PSSYEVX).

We are mainly interested in finding elapsed times, speedups, memory usage for different matrix sizes, number of processors, mapping of the processors, matrix decomposition, etc.

#### 1.1 Choice of Problems

In order to test ScaLAPACK routines, we want to use band and dense matrices. In this section, we present three linear systems of equations whose corresponding matrices are symmetric. The first two systems come from the Poisson equation and the last one is derived from a Toeplitz problem.

The Poisson equation serves as a test problem for linear solvers. Its various discretizations give rise to linear systems of equations with tridiagonal block matrices. We consider the two dimensional Poisson equation with Dirichlet boundary conditions.

$$u_{xx} + u_{yy} = f(x, y) \quad (x, y) \in \Omega$$

$$u = g(x, y) \quad (x, y) \in \partial\Omega$$

We discretize (1) with the use of a second order five point formula (FPF) and with a fourth order nine point formula (NPF). Matrices arising from linear systems obtained from these two discretizations have  $2n$  and  $2n + 2$  as band widths respectively (where  $n$  is the number of interior grid points in each direction).

The second problem is the Toeplitz linear system of equation

$$Tu = b$$

where  $T$  is a Toeplitz matrix whose entries  $t_{ij}$  are given by

$$t_{ij} = 1 + |i - j|$$

$T$  is a symmetric dense matrix that is not positive definite.

## 1.2 Performance Results

Four basic steps are required to call a ScaLAPACK routine [1]:

1. Initialize the process grid
2. Distribute the matrix on the process grid
3. Call ScaLAPACK routine
4. Release the process grid.

Step 1 is used to initialize a  $NPROW \times NPCOL$  process grid by using a row-major ordering of the processes, and to obtain a default system context. Context allows us to create arbitrary groups of processes, to create an indeterminate number of overlapping and/or disjointed process grids, and to isolate the process grids so that they do not interfere with each other.

Table 1 gives an example of mapping of the PEs that we employed for our computations.

	Number of PEs						
	1	2	4	8	16	32	64
nproW	1	1	2	2	4	4	8
npcol	1	2	2	4	4	8	8

**Table 1: Setup Process Grid**

All global matrices must be distributed on the process grid prior to the invocation of a ScaLAPACK routine. All ScaLAPACK routines assume that the data has been distributed on the process grid prior to the invocation of the routine. After the desired computation on a process grid has been completed, it is advisable to release the process grid. The choice of an appropriate data distribution heavily depends on the characteristics or flow of the computation in the algorithm. For dense matrix computations, ScaLAPACK assumes the data to be distributed according to the two-dimensional block-cyclic data layout scheme. Dense matrix computations feature a large amount of parallelism, so that a wide variety of distribution schemes have the potential for achieving high performance. The block-cyclic data layout has been selected for the dense algorithms implemented in ScaLAPACK principally because of its scalability, load balance, and communication properties. The block-partitioned computation proceeds in consecutive order just like a conventional serial algorithm. The basic idea is to distribute a  $N \times N$  matrix on a  $NPROW \times NPCOL$  processor grid using a  $r \times c$  block decomposition. Examples of distributions are given in References [1, 4].

### 1.2.1 LU Factorization

To record the elapsed times and speedups, we just consider the cases  $n = 32$  and  $n = 50$ . Table 2 gives the elapsed times as function of the number of processors elements (PEs) obtained by solving the Poisson equation and the Toeplitz problem with LU decomposition.

PEs	Poisson				Toeplitz	
	FPF		NPF			
	1024	2500	1024	2500	1024	2500
1	21.99	349.75	22.77	350.62	23.70	347.35
2	11.30	149.52	11.44	150.51	12.05	152.54
4	6.32	77.21	6.46	78.31	6.88	80.23
8	3.62	35.42	3.71	35.64	3.93	36.82
16	2.40	20.08	2.46	20.29	2.74	21.56
32	1.62	10.85	1.64	10.96	1.86	11.90
64	1.21	7.22	1.24	7.34	1.49	8.30

**Table 2: Elapsed Time as Function of the Number of PEs. Solution of Linear Systems by LU Decomposition When the Number of Unknowns is 1024, 2500.**

We note that as the number of PEs increases, the elapsed time decreases. In addition, all the three problems have the same timing results even though their corresponding matrices have different band width. During the initialization of the matrices, all the zero entries are included. We do not take advantage of the fact that matrices arising from the discretization of the Poisson equation are banded matrices. Another consequence of our implementation strategy is in the usage of memory. We found that for the same number of unknowns, the three problems use the same amount of memory i.e., 2.45Mwds for 1024 unknowns and 12.38Mwds for 2500 unknowns.

Another implementation strategy is to compress band matrices. Only the nonzero entries are stored. We do not apply this method here even though it is more likely to yield smaller elapsed times and lower memory usage (but perhaps at the expense of more computational complexity during the initialization phase). In its present form, ScaLAPACK does allow the user to perform any data compression as LAPACK does.

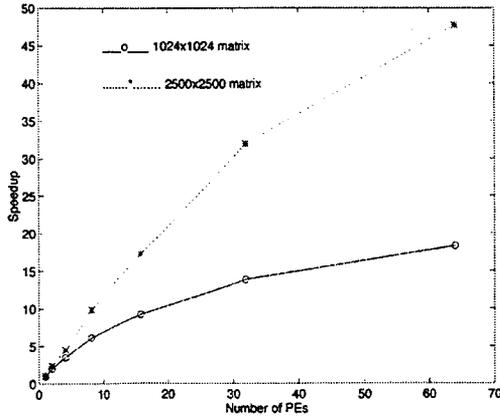


Figure 1: Speedup as Function of the Number of PEs for the *NPF* Poisson Problem with LU Decomposition.

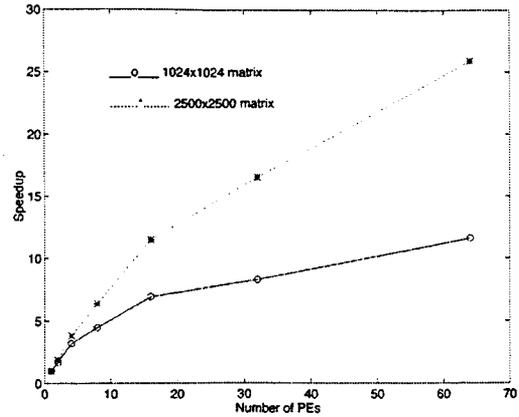


Figure 2: Speedup as Function of the Number of PEs for the *NPF* Poisson Problem with Cholesky Factorization.

We can summarize the timing outputs in Figure 1, where we show the speedup as a function of the number of PEs for the *NPF* Poisson problem. We observe that better performances are obtained when the number of unknowns is large.

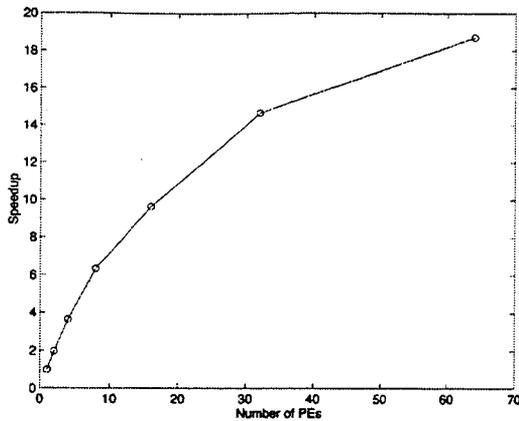
### 1.2.2 Cholesky Factorization

We carry out a similar analysis with the Cholesky factorization for *FPF* and *NPF* Poisson problems. In Figure 2, we plot the speedup as a function of the number of PEs. All the conclusions obtained with LU decomposition are also observed here.

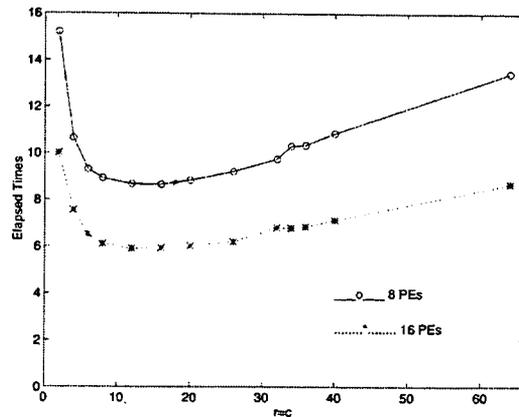
### 1.2.3 Eigenvalue Solver

For the Poisson equation with *FPF*, we find the eigenvalues and eigenvectors of the corresponding matrix for  $n = 32$ . We determine the speedup as function of the number of PEs when the block size for the matrix decomposition is  $r \times c = 2 \times 2$ . The results are summarized in Figure 3. We note that the increase of the number of PEs leads to a decrease of the elapsed time.

For a fixed number of PEs (8 and 16 respectively), we plot in Figure 4 the elapsed time as function of the block size  $r \times c$  where  $r = c$ . We observe that the graphs look like parabolas, and large and small block sizes give the largest elapsed times.



**Figure 3: Eigenvalue Solver: Speedup as Function of the Number of PEs for the calFPF Poisson Problem for a 1024 x 1024 Matrix.**



**Figure 4: Eigenvalue Solver: Elapsed Times as Function of the Block Size  $r \times c$  ( $r = c$ ) for the FPF Poisson Problem for a 1024 x 1024 Matrix and the Number of PEs Equals 8 and 16 Respectively.**

### 1.3 Conclusion

We showed how some ScaLAPACK routines perform on the Cray T3E. We observed the scalability of all the routines that we employed. ScaLAPACK routines are easy to use. Before calling these routines, the user has to follow an initialization step that mainly initializes the buffer, defines the processor mapping and the block size, etc.

Apart from tridiagonal systems of equations, ScaLAPACK considers (from the user's point of view) band and dense matrices in a similar way. All the entries (zero or nonzero) of the matrices are passed to ScaLAPACK. This is a drawback since for band matrices, most of the entries are zero. A data compression done by the user can save a large amount of memory space and may reduce the computational time.

More results of this work can be obtained in reference [4].

## 2. Numerical Experiment with the GEOS General Circulation Model

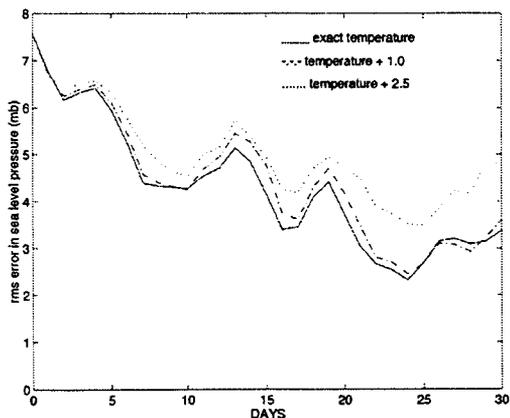
In 1969, Charney, Halem, and Jastrow [2] conducted a series of numerical experiments employing the Mintz-Arakawa two-level General Circulation Model (GCM) to test the Charney conjecture that one could infer the large scale wind fields at all latitudes if one had a continuous history of the complete temperature fields. The results of those experiments more than supported the conjecture by showing that not only the winds, but the complete state of the atmosphere including sea level pressure could be determined. These simulation studies and those that followed later by these and other authors helped to usher in a new era of remote sensing observing systems that was to become operational for the next three decades.

As computers have greatly increased in speed and memory over the past three decades, GCMs have also seen major increases in horizontal and vertical resolutions. The goal of this task was to carry out experiments similar to those in [2] with a present GCM.

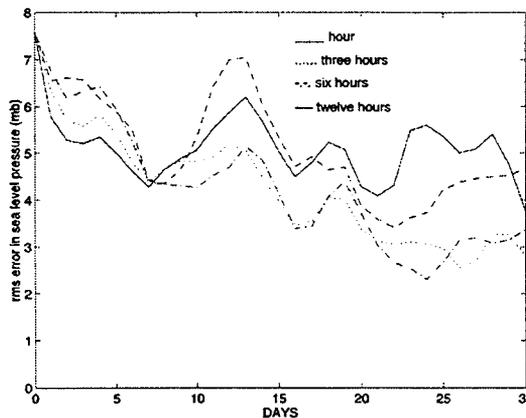
For our analysis, we employed the Goddard Earth Observing System (GEOS) Global Circulation Model [5] with a  $4 \times 5$  resolution and 20 levels. The experiments closely duplicated those of [2]. A 90-day history record (day 1 to day 90) and a 60-day perturbation record (day 30 to day 90) was produced. To obtain the

perturbation record, we introduced at day 30 a  $1^{\circ}$  random perturbation of the temperature fields. Beginning at day 60, history record temperature fields were inserted every 1, 3, 6, 12 hours into the perturbation record. Results of these experiments for temperature states with  $1^{\circ}$ ,  $2.5^{\circ}$ , and  $0^{\circ}\text{C}$  rms errors are presented in Figures 5, 6, 7, and 8 which show the plot of root-mean-square errors in sea level pressure and zonal wind (at 400mb) between the history record and the new inserted record.

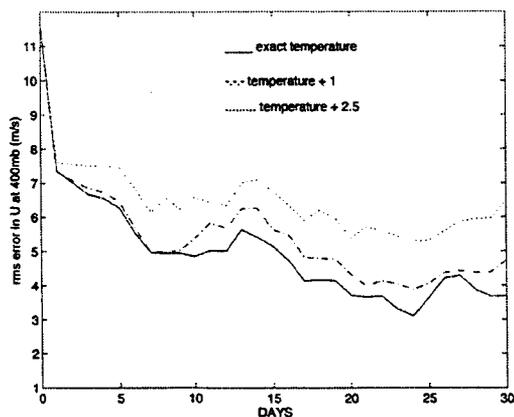
These results with a contemporary model, confirm that a history of accurate temperature profiles can be utilized to infer the complete state of the atmosphere as mentioned in reference [2].



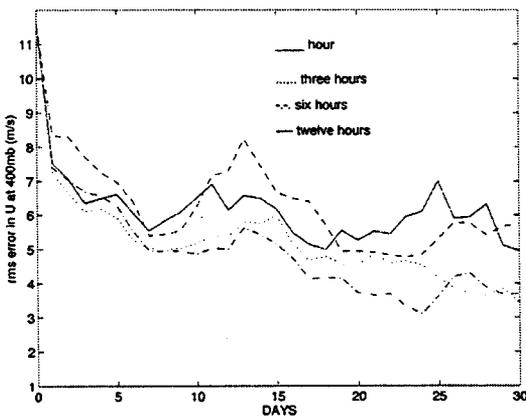
**Figure 5: The rms Error in Sea Level Pressure, in Case Where Temperatures with Random Error Perturbations of  $0^{\circ}$ ,  $1^{\circ}$  and  $2.5^{\circ}\text{C}$  are Inserted Every 6 hr at all Grid Points.**



**Figure 6: The rms Error in Sea Level Pressure, in Case Where Temperatures  $0^{\circ}\text{C}$  with Random Perturbation Error are Inserted Every 1, 3, 6, 12 Hours at all Grid Points.**



**Figure 7: The rms Error in Zonal Wind (m/sec) at 400 mb, in Case Where Temperatures with Random Error Perturbations of  $0^{\circ}$ ,  $1^{\circ}$  and  $2.5^{\circ}\text{C}$  are Inserted Every 6 hr at all Grid Points.**



**Figure 8: The rms Error in Zonal Wind (m/sec) at 400 mb, in Case Where Temperatures with  $0^{\circ}\text{C}$  Random Perturbation Error are Inserted Every 1, 3, 6, 12 Hours at all Grid Points.**

Part of this work was presented at AA Fest: Symposium on General Circulation Model Development: Past, Present, and Future [3].

## References

- [1] Blackford, L. S., Choi, J., Cleary, A., D'Azenedo, E., Demmel, J., Dhillon, I., Dongarrn, J., Hammarling, S., Henry, G., Petitet, A., Stanley, K., Walker, D., and Whaley, R. C. (1997). *ScaLAPACK Users Guide*. SIAM. Philadelphia, PA. (an online version is available at [http://www.netlib.org/scalapack/slug/scalack\\_slug.html](http://www.netlib.org/scalapack/slug/scalack_slug.html)).
- [2] Charney, J., Halem, M., and Jastrow, R. (1969). Use of incomplete historical data to infer the present state of the atmosphere. *Journ. Atmos. Sci., Part 2*, 26(4), 1160-1163.
- [3] Halem, M., Kouatchou, J., and Hudson, A. (1998). A retrospective analysis of the Charney et al., 1969 numerical experiments with the Mintz-Arakawa General Circulation Model. Presented at *AA Fest: Symposium on General Circulation Model Development: Past, Present, and Future*. In Celebration of Professor Akio Arakawa's Career, Univ. of California, Los Angeles.
- [4] Kouatchou, J. (1997). Performance of ScaLAPACK Routines on the CRAY T3E. (an online publication available at [http://farside.gsfc.nasa.gov/RIB/repositories/PI\\_Kouatch/SCALAPACK/scalapack.html](http://farside.gsfc.nasa.gov/RIB/repositories/PI_Kouatch/SCALAPACK/scalapack.html)).
- [5] Takacs, L., Molod, A., and Wang, T. (1994). Documentation of the Goddard Earth Observing System (GEOS) General Circulation Model-Version 1. *NASA Technical Memorandum 104606*.

---

---

# **APPLIED INFORMATION TECHNOLOGY BRANCH**

---

**Yelena Yesha**, Branch Head

*Digital Libraries Technology*

**Yair Amir**, Johns Hopkins University  
**Susan Hoban**, University of Maryland Baltimore County

*Global Legal Information Network (GLIN)*

**Nabil Adam**, Rutgers University  
**Konstantinos Kalpakis**, University of Maryland Baltimore County

*Executive Secretariat to the U. S. Global Change Research Program's Data and  
Information Management Working Group*

**Les Meredith**, Senior Scientist

*Executive Secretariat to the Committee on Environmental and Natural Resources (CENR)  
Task Force on Observations and Data*

**Sushel Unninar**, University of Maryland Baltimore County

*Pilot EOS Direct Readout Ground Systems Support*

**Fran Stetina**, Fran Stetina and Associates

*NASA and the Private Sector*

**Murray Felsher**, Associated Technical Consultants

*Support for NASA/NOAA Collaboration in Earth Science Modeling*

**Miodrag Rancic**, University of Maryland Baltimore County

*Development of an Implicit 2D Adaptive Mesh Refinement and De-refinement  
Magnetohydrodynamics Code*

**Dinshaw Balsara**, University of Illinois

*A Scalability Model for ECS's Data Server*

**Daniel Menascé**, George Mason University; **Mukesh Singhal**, Ohio State University

*Numeric Simulation of a Volcanic Jet-plume*

**Santiago Egidio Arteaga**

*Volumetric Display of Earth and Space Science Data*

**David Ebert**, University of Maryland Baltimore County

*Putting Log Data to Work: Mass Storage Performance Information System*

**Lisa Singh**, Northwestern University

## DIGITAL LIBRARIES TECHNOLOGY

CESDIS has been tasked with conducting research in areas related to digital library technology, specifically in areas which will complement the work proposed by the investigator teams funded by NASA Cooperative Agreement Notice CAN-OA-94-01. Work was performed in this reporting year by Yair Amir of Johns Hopkins University and Susan Hoban of the University of Maryland Baltimore County.

### *Combining Satellite Communication in Commedia*

**Yair Amir**  
**Johns Hopkins University**  
**Department of Computer Science**  
**(yairamir@cs.jhu.edu; <http://www.cs.jhu.edu/~yairamir>)**

### **Statement of Work**

Using the Internet currently, it is possible to pass a message between almost any two points within the U.S. with a latency of about 80 milli-seconds (turn around time), with a relatively high probability of success. Preliminary measurements for satellite communication show that latency of about half a second (turn around time) will be experienced for each satellite hop. This drawback creates an interesting problem for protocols that are designed to achieve interactivity. Satellite communication may provide high bandwidth with access to almost any point on the Earth, including places where Internet connection is not yet supported or which lacks the necessary bandwidth for systems such as Commedia, a crossplatform infrastructure for multimedia conferencing. A subcontract has been in place with Johns Hopkins University for research by Dr. Amir on the possibility of utilizing satellite communication within Commedia. A report on his project follows.

### **Report**

This research program has been conducted in the framework of the NASA Center of Excellence in Space Data and Information Sciences (CESDIS). The research encompassed two distinct parts. In the first year, we studied the impact of high latency channels on the protocols and technologies used in the Commedia project. In the second year, we have designed and implemented new protocols that circumvent the problems of high latency links in the system.

We designed and built an infrastructure that allows collaboration using the web over high latency links. Utilizing a novel replication scheme, we managed to mask most of the effects of high latency. The scheme automatically directs users (web browsers) to the best replica of the replicated web server—typically the one on their side of the high latency link. We re-engineered our group communication protocols to cope with high latency links as part of the network. This work reached an advanced stage, but is not yet completed. Some of the project results were demonstrated, both at Hopkins and at CESDIS several times during the course of the project.

## Summary of Findings

### Investigating the Impact of High Latency Links

The first step in this research was to understand the problems imposed by high latency links on collaboration over the network. To do that, we designed and built a collaboration infrastructure that included several types of media sources and players. This was based on our existing group communication protocols. After some experiments, we narrowed down the list of sources and media players to simple eight bits audio, Connectix uncompressed video, and MPEG1 video/audio streams.

We developed programs that play the audio, the Connectix, and the MPEG1 streams as native applications in the Unix and Windows operating systems. In addition, we developed Java applets programs to play Connectix video inside a web browser. Although the quality of the Connectix video was far from the MPEG1's, it was quite impressive to demonstrate the potential of web-based collaboration (8 frames per seconds, 8 bits for color).

We deployed the existing group communication protocols in the wide area network we decided to use. This was not a perfect solution, but it worked, enabling us to gather valuable information on the behavior of the protocols in real life over high latency links. We pushed 160Kbits/sec reliable multicasting between all of the machines in this network. The network configuration, detailed below, contained 19 machines from Hopkins, UMBC, GSFC, Rutgers, and DIMACS. We have used this testbed continuously for six months. Following is the network layout of our testbed:

# cnds.jhu.edu domain	# cs.umbc.edu domain	# gsfc.nasa.gov domain
5 128.220.221.255	3 130.85.100.255	3 128.183.0.0
commedia 128.220.221.1	topdog 130.85.100.62	cesdis3 128.183.38.27
com1 128.220.221.11	stavro 130.85.100.121	cesdis7 128.183.38.31
com2 128.220.221.12	retriever 130.85.100.32	what 128.183.38.63
com3 128.220.221.13		
com5 128.220.221.15		
# rutgers.edu domain	# dimacs.rutgers.edu domain	
3 128.6.42.255	5 128.6.75.255	
cimic 128.6.42.134	dimacs 128.6.75.16	
cimic1 128.6.42.127	lunar 128.6.75.43	
adam 128.6.42.5	iyar 128.6.75.51	
	av 128.6.75.54	
	brownin 128.6.75.22	

We have used three web servers (at Hopkins, UMBC, and DIMACS) in our experiment as the potential hubs for collaboration using the web technology. This part of the research was completed when we had a limited working version of Commedia, with multicast protocols, group communication services, very simple media protocols, and representative applications. This testbed was running on high latency links over the Internet, connecting several local area networks. We have collected valuable data that directed us in designing the new protocols.

### An Infrastructure for Collaboration over High Latency Links

We have investigated the behavior of the protocols in our testbed. In particular, we were interested in the implications of latency and omission loss over the high latency links. Based on the testbed results, we have identified the following key issues that need to be addressed by a new protocol:

- Efficient utilization of wide area links for message dissemination. We have noticed that when high latency links are involved, it is very important to construct the best routing tree. A bad decision might be reflected in considerable loss of performance.

- The ability to cope with non-negligible packet loss rates over high latency links. We discovered that it is very important to recover lost packets directly from the nearest neighbor. In particular, end-to-end reliability (which is the most prevalent method) will carry a very high cost.
- The ability to limit the domain over which messages are disseminated to only the sites that have active receivers for these messages. This has to be done without affecting the ability to have open group semantics and without compromising the Virtual Synchrony semantics for reliability and ordering guarantees.
- The ability not to block the sending of messages at the same time, from many sources, as well as the ability not to block delivery of these messages. In practice, most of the streaming information, such as video and audio streams, only require at most FIFO delivery (e.g., MPEG stream). This requires the ability to multicast AND the ability to deliver without incurring latency due to other messages.

The above properties led us to the design of a protocol that is plugged into the existing architecture of Commedia. The protocol uses the following techniques:

- Combining a local area network protocol with a protocol designed for the wide area link.
- Routing packets over the wide area network using distinct routing trees, each of which is optimized for each source site.
- Limiting the scope of data dissemination only to the necessary sites. All of the sites will get control information (as opposed to data) in order to maintain the ordering and reliability guarantees.
- Allowing each site to generate messages independently of any other site. To vastly increase the scalability of the protocol, we decided to maintain the current ring mechanism within each site. Our current research shows that this structure poses no performance penalties on the local area network. Overall, this technique allows us to handle streams (unreliable, reliable, and FIFO) without incurring latency beyond the propagation delay.
- Using an object-based approach for the different kinds of control and data messages. This allows us to customize packets (size and rate) for each link of the relevant routing tree.
- Solving the reliability problem on a hop by hop basis.

The implementation of the protocol is still underway. Several parts of it are already working within our framework and were demonstrated. We hope to complete a stable version of the protocol in the next few months. We plan to publish our results at that time.

It appears that web technology is a very promising infrastructure for collaboration. Using the freely available, always improving web browsers, it is fairly simple to build a collaboration framework using the web server as a central connecting point and data repository. This architecture is working well when the clients are in the vicinity of the web server. When collaborators are separated by satellite links, this solution is much less desirable.

Replicating the web server on both sides of the high latency link seems to be a good solution. This creates two challenges: how to efficiently replicate the data (including live streams), and how to seamlessly direct each client to the best copy. Most of the existing web replication architectures involve a cluster of servers that reside at the same site. These architectures improve performance by sharing the load between the different replicas, and improve availability by having more than one server. However, they cannot address the performance and availability problems embedded in the network, especially when high latency links are involved.

Our architecture, in contrast, incorporates wide-area replication, replicating the web server (data and processing power) on both sides of the high latency link. This may be achieved by using the group communication mechanism.

We have implemented three alternative methods to automatically direct the user's browser to the best replica:

- The HTTP redirect method: This method is implemented using web server-side programming at the application level.
- The DNS round trip times method: This method is implemented at the Domain Name Service (DNS) level, using the standard properties of DNS.
- The shared IP address method: This method is implemented at the network routing level, using the standard Internet routing.

Selecting the *best* replica takes into account the following considerations:

- Network topology: which replica is closest to the client, network-wise. This will strongly favor a web browser on this side of the satellite link.
- Server availability: which servers are currently active. If no server is active on this side of the link, it is better to have a server on the other side rather than not being able to participate in the collaboration at all.
- Server load: which server is currently able to return the most rapid response.

This infrastructure is complete. A paper describing this work was recently accepted for publication in the DISC'98 conference. The web replication software will be available on the Commedia web page by September 1998.

## Publications

Amir, Y., Peterson, A., and Shaw, D. (1998) Seamlessly selecting the best copy from Internet-wide replicated web servers. *Proceedings of the 12th International Symposium on Distributed Computing*. Andros, Greece. This paper is available as a CESDIS technical report.

## **NASA Digital Library Technology Project Support**

**Susan Hoban**  
**University of Maryland Baltimore County**  
**Department of Computer Science and Electrical Engineering**  
**(shoban@pop900.gsfc.nasa.gov)**

### **Profile**

Dr. Hoban received a B.S. in astronomy, an M.S. in physics, and a Ph.D. in astronomy, all from the University of Maryland. Prior to joining CESDIS through a subcontract with the University of Maryland Baltimore County, Dr. Hoban was a Principal Scientist with Hughes STX, providing support to the GSFC Digital Library Technology Project as the Assistant Manager. She served as a guest lecturer for the Maryland Space Grant Consortium, teaching a course entitled *Introduction to the Internet for K-12 Educators*. As the Principal Investigator on a project funded through NASA's Innovative Developments in Education in Astronomy Science, Dr. Hoban became Dr. Sue in *Astronomy On Line: Ask Dr. Sue*. In this capacity she developed science education curricula as well as its World Wide Web implementation. Dr. Hoban was also instrumental in developing a homepage for NASA's Chief Scientist, Dr. France Cordova.

Dr. Hoban's association with NASA began when she was selected to participate in the NASA Graduate Student Researchers Program for work in charged coupled device imaging (astronomical observations and analysis). As a National Academy of Sciences Research Associate, she performed research on infrared spectroscopy of astronomical sources and served as the IRAF Data Reduction Package Manager for installation, maintenance, and user assistance. This work was continued as a research scientist with USRA's Goddard Visiting Scientist Program. Dr. Hoban's research interests include image processing of remotely sensed data, two-dimensional data analysis (spectral and spatial), and multi-wavelength studies of comets and young planetary systems.

### **1. Digital Library Technology (DLT) Project**

- a. Support is provided, as needed, to Dr. Nand Lal (Code 933) in managing the NASA HPCC/IITA Digital Library Technology Project. This activity includes reporting, assessment, proposal writing, research, and development in digital libraries. The IITA Program is being phased out over FY98, and the current activities will transition into the NASA HPCC Learning Technologies Project (LTP).

The DLT project is a component of the LTP Digital Audio Testbed (DAT). Hoban participates in coordination sessions with participants from other NASA centers. The DAT conducted a successful transatlantic demonstration of RealMedia. Real-time audio and video were successfully broadcast among schools in Paris, France; Washington, DC; and Brooklyn, NY. NASA Administrator Daniel Goldin participated from Washington, and First Lady Hilary Rodham Clinton participated from Paris.

As part of supporting the DLT project, Hoban traveled to Lewis Research Center (11/4/98) to deliver a summary of the DLT project at the LeRC Middleware Meeting. She also attended a site visit at Carnegie Mellon University, which is funded through the DLT project as part of the NSF-NASA-DARPA Digital Library Initiative (DLI). She also attended the DLI Principal Investigators meeting January 5-6, 1998, held in Berkeley, CA, as well as the final meeting of the IITA Principal Investigators, which was held in Portland, ME, June 1-3, 1998.

b. **FITNESS: Facilitating Information Technology iNfusion into Earth and Space Science**

This activity is part of the DLT project. As the term of the cooperative agreements comes to a close, the DLT project office has initiated the FITNESS effort to place the technologies developed by the investigator teams into the Earth and space science communities. This brokering effort involves interfacing with the technologists and the science communities.

One activity undertaken as part of the FITNESS effort was the Information Technology Workshop 2, which was held at Goddard on September 24-26, 1998. The Workshop featured DLT investigators, as well as investigators from NASA's Office of Space Science Applied Information Systems Research program. About 50 people attended the workshop. G. Flanagan and M. Meyett/CESDIS participated in the planning and implementation of this workshop.

A second FITNESS activity involved the brokering of a connection between DLT Principal Investigator J. Percival/University of Wisconsin and the National Undergraduate Research Observatories (NURO) in an attempt to improve transmission rates for NURO users with the Progressive Image Transmission software. Several tests were conducted by the users, and feedback was provided to Percival. All parties considered this activity a success.

Hoban worked with N. Lal/GSFC and H. Burrows/HSTX to prepare a DDF proposal for FITNESS. This proposal was not funded.

Hoban attended meetings of the Space Science Data System Technical Working Group. This group was formed out of Code S at NASA Headquarters to develop a framework for integrating all Space Science technologies and data. The DLT/FITNESS effort is proposing to develop the web site for the Space Science Data System.

- c. Background work has begun for the preparation of a new solicitation which will be sponsored by the Learning Technologies project.

## **2. GLIN: Global Legal Information Network**

A joint effort between the Library of Congress and NASA, this project is developing the infrastructure to place the legal instruments of all countries on line. Hoban is responsible for coordinating the technical development that NASA provides, through CESDIS, for the Library of Congress. As part of this task, she attends the periodic meetings of the GLIN technical team and the Ad Hoc Advisory Council.

Hoban assisted M. Halem/930 in the preparation of his presentation for the GLIN Director's meeting, which was held in August at the Library of Congress.

## **3. ELIS: Environmental Legal Information System**

Kalpakis and Hoban (CESDIS), with colleagues from the Center for International Environmental Law, the Law Library of Congress, and the Earth and Space Data Computing Division of NASA's Goddard Space Flight Center (W. Campbell/935 and J. P. Gary/930), successfully proposed to the MTPE Earth Science Information Partners 3 program. The proposal is entitled, "Integrating Environmental and Legal Information Systems." We proposed to identify remote sensing data that may be applicable to the interpretation and enforcement of environmental laws, and develop a system for integrating these data with existing on-line databases of environmental legal information. This work will be an enhancement to the Global Legal Information Network, a joint effort by the Library of Congress and NASA. We will also develop a model piece of environmental legislation which incorporates remote sensing data ab initio, to be used as a teaching tool at the American University and other law schools. The system will be called ELIS (Environmental

Legal Information System). Susan developed a web site for the project:  
<http://cesdis.gsfc.nasa.gov/~hoban/elis>.

This project is funded for 5 years. Awards were announced in December 1997, and the cooperative agreement was signed in May 1998. Work began shortly thereafter. Project members attended the NRC Workshop on the ESIP Federation, Feb. 23-25, in Washington, DC, and the first meeting of the ESIP Federation, May 5-7, 1998, also in Washington, DC. Kalpakis and Hoban submitted an abstract to the first EARSel WORKSHOP on IMAGING SPECTROSCOPY.

#### **4. Scientific Information Environment for SOFIA (Stratospheric Observatory For Infrared Astronomy)**

We have written a white paper (<http://cesdis.gsfc.nasa.gov/~hoban/sofia/sie.htm>) outlining a scientific information environment for NASA's Stratospheric Observatory For Infrared Astronomy. This system would coordinate data storage, retrieval and processing for the astronomical community.

We have also prepared a proposal, submitted to the SOFIA Data Archive team (via Dr. Mark Morris, UCLA), which is also available on the web (<http://cesdis.gsfc.nasa.gov/~hoban/sofia/>).

#### **5. GIBN: Global Interoperable Broadband Network**

The GIBN effort is one of 11 projects comprising the G7 Information Society. Hoban is responsible for coordinating the GIBN Trans-Pacific Digital Library Experiment and development and implementation of the web site for the project (<http://dit.gsfc.nasa.gov/gibn/>). This project will be a demonstration of several digital library activities over high-performance networks between the United States and Japan. This effort is funded by NASA Headquarters, through Lewis Research Center.

#### **6. HST Observations of R-Aqr**

Our proposal (Hollis/NASA, Hoban and Knoll/STScI) to observe the R-Aqr binary jet system with the NIC-MOS camera on board the Hubble Space Telescope has been accepted. The observations will determine the physical characteristics of the dust in the jet. Observations are scheduled for the fall.

#### **7. Advances in Digital Libraries ADL98**

Hoban served on the ADL98 organization committee, and organized a panel with N. Lal and H. Burrows entitled "Adoption of Digital Library Technologies by Various Communities." She also co-authored a paper entitled, "Socio-economic Effects of Electronic Publishing," with G. H. Burrows/RSTX and N. Lal/Code 933. The meeting was held in Santa Barbara, April 22 - 24, 1998.

#### **8. NGST Data Archive Study Committee**

Hoban was asked to serve on the NGST Data Archive Study Committee. This committee is charged with conducting a study of the state-of-the-art of scientific data systems, and with preparing a document which makes recommendations to NGST for the development of the NGST data archive. This document will be delivered to NGST at the end of FY98.

## 9. Other

Hoban participated in the preparation of a proposal to NSF with Y. Yesha/CESDIS, Roger Ghanem (JHU) and others entitled "Applications of HPCC for Identification, Tracking and Control of Environmental Pollution." This proposal was not funded.

She attended the CSOC planning meeting at USRA in Houston.

Attended Hughes STX "Scientific Data Centers Workshop", held at GSFC, 11/17 & 11/18.

Attended the Digital Earth Workshops, held in Code 930, NASA/GSFC in April and June, 1998.

## GLOBAL LEGAL INFORMATION NETWORK

The Global Legal Information Network (GLIN) is an international, non-commercial, cooperative network of government agencies working in conjunction with the Law Library of the U. S. Library of Congress to create a database of international law documents which will be available to member countries throughout the world and which will facilitate international cooperation and joint ventures. The Library of Congress and NASA have signed a Memorandum of Understanding to establish a framework for coordinating cooperative efforts on updating and enhancing the technological infrastructure of GLIN. The intent is for the work to be conducted through collaborative and cooperative research by the Law Library, NASA GSFC, industry, academia, participating GLIN members, and other relevant international bodies.

In order to more efficiently collect and disseminate current legal information, a prototype system has been established to acquire, process, and retrieve digitized legal texts. The application of advanced digital technology is necessary to maintain the GLIN database and to increase the speed and flexibility of the system as the volume and complexity of the data expands. Upgrades and enhancements to GLIN are desired in order to share the benefits and burdens of obtaining, processing, and retrieving legal texts among cooperative partners throughout the world. Nabil Adam (Rutgers University) and Konstantinos Kalpakis (University of Maryland Baltimore County) have contributed to the CESDIS portion of this effort. Their reports follow.

### *Information Extraction Applications for GLIN*

**Dr. Nabil R. Adam  
Rutgers University  
Center for Information Management, Integration, and Connectivity  
(CIMIC)**

## 1. Introduction

This report discusses the work undertaken at Rutgers CIMIC for the CESDIS and Library of Congress work on the Global Legal Information Network (GLIN). The specific work focuses on applying Information Extraction (IE), a form of Natural Language Processing, to the problem of law summary classification and retrieval. For this work, we developed an incremental modeling methodology that reuses an existing hier-

archical classification scheme followed by a systematic method for identifying common concepts within and between classes of documents. Concepts are words or phrases that appear in specific linguistic contexts (e.g., as specific parts of speech or sentence fragments) and are expressed by a set of semantic and syntactic constraints. We then train a series of IE "extractors" based on this model that are capable of identifying these sets of concepts. The concept identification process is used to determine if a novel document should be assigned to a class. Novel documents are passed down the hierarchy and are processed by extractors at each node. Successful extraction of key concepts leads to a class assignment. In the case of GLIN, this results in index terms being assigned to the GLIN summary.

Information retrieval (IR) is achieved by gathering the instantiated concept definitions (sets of constraints) and using them to form an index of the document. This index can then be queried using a standard user interface. However, rather than return all of the documents that contain a specific query word, our IR system first returns a set of class/concept pairs that are indicative of the query terms supplied. The user may then filter the query further by choosing some or all of the class/concept pairs. This results in much smaller and more precise result sets.

In August 1996, we received access to a collection of approximately 50,000 GLIN summaries that were classified/indexed using terms from the GLIN thesaurus. Using the modeling methodology just described, we modeled a subset of the GLIN documents as a hierarchy of classes. The 18 classes modeled represent about 2% of the total index terms found in GLIN (18/700), while the documents associated with these index terms represent over 10% of the total documents found in our test collection. Within the classes, 32 unique concepts were found.

Based on this hierarchy, we then trained a series of IE systems to recognize these concepts. Following the standard IR practices, we used 80% of the documents in a class for training and tested using the remaining 20%. The IE software was licensed from the University of Massachusetts at Amherst and adapted for use on this project. The modeling, training, and software development effort took just over 4 months to complete. Much of this initial time was spent in developing the methodology and software required to automate much of the work. Using the incremental modeling methodology, a new class can be modeled and added to the overall system in approximately 4 person hours.

A classifying system was built using the hierarchy and trained extractors. A web interface was designed to allow a user to type or paste in a new GLIN summary and have that summary classified with up to 18 different GLIN thesaurus terms. A web interface to the GLIN classifier can be found at the URL: <http://cimic.rutgers.edu/~holowcza/glin/ling/>.

## **2. Experimental Results**

After creating the extractors, each one was run on a test set of summaries. 100 of the summaries came from within the sub-domain (relevant texts) and 100 were randomly chosen from outside of the sub-domain (irrelevant texts). Recall and Precision measures were recorded. The system favors recall with results ranging from 70 to 100%. Precision measured ranged from 63 to 100%. In most cases, a classic recall/precision tradeoff was identified. Future work may focus on adjusting the training tolerance error  $r$  values (a parameter to the training function) to try and improve recall and precision for some of the classes.

## **3. Information Retrieval Application**

We also created an IR application as described previously. We classified 5,000 GLIN summaries using our classifier and then formed document indexes from the instantiated concept definitions (constraints). The index attributes include the sentence number, segment number (within the sentence), class, concept, phrase type (noun phrase, prepositional phrase), the actual words used, and a document ID (pointer to the actual document). 5,000 documents created 12,129 records (note that more than one CN definition can

apply to the same document). A WWW forms interface and several CGI scripts were also written for the IR application. The URL for the GLIN query application is: <http://cimic.rutgers.edu/~holowcza/glin/ling/query.html>.

#### 4. Other Events

Our work has led to the following publications and presentations:

1. Holowczak, R. D. Extractors for Digital Library Objects. Ph.D. Dissertation, Rutgers University. May, 1997.
2. Holowczak, R. D. and Adam, N. R. Information Extraction based Multiple-Category Document Classification for the Global Legal Information Network. Proceedings of the Ninth Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-97). July, 1997. Providence, Rhode Island.
3. Holowczak, R. D. Extractors for Digital Library Objects. Presentation given to Columbia University Department of Computer Science. February, 1997.
4. A number of short talks on Extractors for Digital Libraries were given:
  - Sarnoff Research Center, Princeton, NJ - Presentation for research center staff.
  - Rutgers University, Newark, NJ September 22, 1997 - Presentation for visiting faculty and provost from UMBC.

#### 5. Monograph *Digital Libraries for the 21st Century*

Towards the end of Summer 1997, we began work on a book entitled *Digital Libraries for the 21st Century*, co-authored by Nabil Adam, Milton Halem, Richard Holowczak, and Yelena Yesha. The book presently has six chapters under development that include the following:

- I. Introduction
- II. History and State of Digital Libraries Today
- III. The Role of Digital Libraries in the 21st Century
- IV. Digital Libraries for Science, the Arts, and Commerce
- V. Challenges and Issues Facing Digital Libraries' Evolution
- VI. Summary and Future Directions

A draft of the Introductory section has been completed and additional materials are currently being gathered for the remaining chapters.

In chapter 1, we give an introduction to the book including our motivation for the themes of the book, contributors and sources of information, and a description of each chapter.

In chapter 2, we discuss the history and current state of digital libraries today. We trace the roots of the digital library back to electronic card catalogs that provide rudimentary index search capabilities. We then review approaches to text storage and retrieval where the full text of documents is stored and indexed. In both of these first two cases, the indexes and document collections typically reside on a single host. Distributing documents among many hosts in a networked environment is thus the next focus of the chapter. Here the Internet and World Wide Web set the stage for distributed, multimedia document collections. Finally, we conclude our historical review with a look at modern multimedia information systems.

In the second half of chapter 2, we review a number of commercial efforts and research projects dedicated to building digital libraries. Government agencies, universities, corporations, and non-profit organizations around the world have initiated such projects. Due to the rapid proliferation of these efforts, we can only provide a snapshot of the state of research and development as it stands today.

In chapter 3, we provide our views on the role of the digital library in the 21st century. In describing this "ideal" picture, we examine the digital library from the perspective of the user communities and discuss how they will interact with digital libraries. We will also describe characteristics of the digital library such as how content will be stored, accessed, searched for, delivered, and secured. Finally, we end chapter 3 with our views on the roles of digital libraries in society.

We foresee future digital libraries that cater to specific needs of groups of individuals or that specialize in storing, maintaining, and presenting particular topical areas. In chapter 4, the requirements of and uses for a number of specialized digital libraries are discussed. These include libraries specializing in the arts, science and engineering, and in support of commerce. The requirements cited include both technical and non-technical considerations. Our references for this chapter include our personal and professional interactions with leading physical scientists, researchers, museum curators, and library directors at many of the leading public and private sector organizations.

We begin chapter 5 with a review of our assumptions for future digital libraries. We then discuss research directions that, in light of these assumptions, must be pursued in order to achieve the kinds of sophisticated interactions with digital libraries we envision. The research directions discussed include the storage, indexing, integration, search, retrieval, and presentation of digital library content, ontologies, knowledge bases, and intelligent, adaptable systems that allow the digital library to cater to the users with diverse backgrounds and expertise, universal access to digital library content, and ensuring security and privacy for users of the digital library.

**Konstantinos Kalpakis**  
**University of Maryland Baltimore County**  
**Department of Computer Science and Electrical Engineering**  
**(kalpakis@cs.umbc.edu)**

My primary focus during this period was development work for GLIN. I undertook a number of developed activities related to the GLIN project. My efforts during this period were on developing a sequence of prototypes, experimenting with various approaches. The main line of the approach was to utilize the services provided by traditional relational database management systems in order to develop a GLIN prototype that addressed the requirements of the Law Library of the Library of Congress. The initial approach to using Postgress and Inquiry, though shown that it can be done through a prototype, had certain drawbacks, that lead me to discard that approach. The next approach was to opt for DB2 or Oracle 8. I experimented with both of them, and both were shown to be appropriate. Based on the desire of the Law Library, the Oracle platform was selected. I developed a prototype system based on Oracle 8, using a combination of Java and Javascript to develop the various modules needed, while using the JDBC protocol to communicate with the database. The option to use the PL/SQL and Javascript was not selected, though quite appealing, since that would have lead into making the prospect of migrating into a non-Oracle platform infeasible. At this point, a prototype is running on Windows NT and Solaris platforms, as a Java application. Even though I was targeting that the prototype could also be used through the Web on standard Web browsers, due to limitations of the Netscape and Explorer browsers, currently, only a limited set of functions are fully available. I am exploring ways to get around those issues. A version of the prototype was demonstrated at the GSFC Technology Showcase in March 1998. Besides the development/prototyping work which was the main thrust of my effort for that period, various experiments were performed on bilingual text storage and retrieval, indexing and retrieval processing times, and capacity estimation. However, these efforts were not completed in this period.

As an extension to the basic GLIN prototype, in cooperation with colleagues from CESDIS, NASA GSFC, the Law Library, and the American University, we submitted a proposal, in response to CAN-97-05, with the

title "Integrating Legal and Environmental Information Systems" to the NASA's MTPE program. This proposal was selected for funding in the Spring of 1998, and is currently under way.

## **EXECUTIVE SECRETARIAT TO THE DATA AND INFORMATION MANAGEMENT WORKING GROUP OF THE U.S. GLOBAL CHANGE RESEARCH PROGRAM**

The Data and Information Management Working Group (DIMWG) acts as the data management arm of the U.S. Global Change Research Program (USGCRP) and provides an informal mechanism for interagency coordination and cooperation. Working Group agencies are the Department of Commerce, the Department of Defense, the Department of Energy, the Department of the Interior, the Environmental Protection Agency, NASA, the National Science Foundation, and the U.S. Department of Agriculture. The Department of State and the National Academy of Sciences serve as liaison members. The Data and Information Management Working Group has six subgroups and more than 50 active participants. The DIMWG supports collaboration between computer and Earth scientists involved in database, data management, and data distribution research by facilitating access to global change-related data and information in useful forms.

This task was assigned to CESDIS through the Global Change Data Center (GCDC) in the NASA Goddard Earth Sciences Directorate (Code 900). It requires the provision of Executive Secretariat support to the Data and Information Management Working Group including the guidance and coordination necessary to ensure future accomplishments which can be endorsed by the National Academy of Sciences and which enhance the level of general cooperation and participation of the DIMWG agencies. Les Meredith and is responsible for providing the support required by this task.

**Les Meredith, Senior Scientist  
(les@usra.edu)**

### **Profile**

Dr. Meredith holds Bachelors, Masters, and Ph.D. degrees from the State University of Iowa. He is a Fellow of the American Association for the Advancement of Science, a Fellow of the Royal Astronomical Society, and a member of the American Geophysical Union, the American Physical Society, Phi Beta Kappa, and Sigma Xi.

Dr. Meredith's contributions to space science span more than 40 years and include employment as Head of Rocket Sonde Branch and Meteor and Aurora Section of the Naval Research Laboratory and a variety of positions at NASA Goddard Space Flight Center including Space Science Division Chief, Deputy Director of Space and Earth Sciences, Assistant Director, Acting Director, Director of Applications, and Associate Director. He spent a year as Liaison Scientist for Space Science in Europe with the Office of Naval Research in London, four years as the General Secretary of the American Geophysical Union, and more than five years as its Group Director for meetings and advocacy.

Dr. Meredith is the recipient of the NASA Exceptional Scientific Achievement Medal (1965), the NASA Outstanding Leadership Medal (1975), the Senior Executive Service Presidential Meritorious Award (1981), and the NASA Distinguished Service Medal (1987).

## Report

1. Developed the agenda, sent out meeting notices, produced all material for the meeting, briefed the chair, wrote the minutes, and followed up on the action items established for the CENR's Subcommittee on Global Change Research, SGCR, Data Management Working Group, DMWG. The DMWG met about monthly. Between meetings, I worked with members of the DMWG and the National Academy of Sciences on special issues, responded to requests for help, and performed the many actions needed to keep the DMWG interagency coordination process productive and viable.
2. In my role as Program Associate for Data Management of the SGCR, I actively participated in their almost weekly executive planning meetings. In particular, I represented the SGCR's data management interests and regularly responded to questions and action items.
3. Drafted the DMWG's sections that were included in the SGCR's publication "Our Changing Planet-FY1999." This included a section that summarized the accomplishments of the DMWG in the past year. Importantly, it also included the first identification of the DMWG's performance measures. Performance measures were required by OSTP for FY1999 for all elements of the SGCR and will be used to judge its success.
4. Initiated a DMWG process to draft language for an agency to use in grants if the agency wants the recipient to make the data produced available. This language was approved by the DMWG and sent to the SGCR for its endorsement. It was also sent to the DMWG's contact in Canada who had asked for help on this issue.
5. Initiated a DMWG process to create a mechanism to make it possible for datasets to be cited and for the individuals and organizations responsible for them to get credit for their work. This required the establishment of citation guidelines and has resulted in a document listing the Global Change-related data sets the agencies made available in 1997. The plan is that this document will be published by the SGCR and also be incorporated as a part of the DMWG's Global Change Data and Information System, GCDIS, Web page. Already this dataset citation process has resulted in additional agencies identifying the data that they have. Publication of such a dataset citation list for 1998 has been made an SGCR performance measure. This dataset citation capability could well be one the DMWG's actions that has the greatest long-term importance to both dataset providers and users.
6. Attended most meetings of the new National Assessment Working Group, NAWG, of the SGCR. National assessments of the potential effects of climate change have become a major new part of the SGCR program. One result of this attendance was my initiation of having the DMWG, through DOE, given the responsibility for the NAWG Web page. Another was getting the Chair of the NAWG to review his program for the DMWG. This resulted in his invitation to the DMWG to make a formal proposal for its support to them. I drafted this proposal which was presented at a joint NAWG/DMWG meeting by the DMWG Chair. So far this has resulted in the NAWG requesting the DMWG to draft appropriate data management policies for them. I have drafted these policies and they are now in the review process.

## **EXECUTIVE SECRETARIAT TO THE COMMITTEE ON ENVIRONMENTAL AND NATURAL RESOURCES (CENR) TASK FORCE ON OBSERVATIONS AND DATA**

The function of the Secretariat is to act on behalf of the CENR Task Force as the primary CENR interface for international consultations on scientific planning and implementation of the Global Observing System and its related data management system. This includes coordination with the international efforts underway by the Global Terrestrial Observing System (GTOS), the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), the Committee on Earth Observation Satellites (CEOS), the World Climate Research Programme (WCRP), and the International Geosphere-Biosphere Programme (IGBP).

This task was assigned to CESDIS through the Global Change Data Center (GCDC) in the NASA Goddard Earth Sciences Directorate (Code 900). It requires the provision of all the necessary technical and administrative support to assist the CENR Executive Director in implementing the responsibilities of the Secretariat. This includes coordinating the activities of the Task Force and its working groups, planning and coordinating U.S. participation in the International Global Observing System in accordance with the strategy outlined in the OSTP concept paper on the GOS, coordinating relevant observations and data management budget justification and advocacy material among the CENR subcommittees for submission to the Task Force, and coordinating with the Task Force's Data Management Working Group to promote effective access data management systems for CENR relevant global, regional, state, and local environmental and natural resources data.

Sushel Unninayar is responsible for providing the support required by this task. He works with CESDIS through a subcontract with the University of Maryland Baltimore County.

**Sushel Unninayar  
University of Maryland Baltimore County  
Department of Computer Science and Electrical Engineering  
(sushel@cesdis.usra.edu)**

Primary activities included the following: (1) Development of the Integrated Global Observing Strategy (IGOS); (2) Interagency coordination regarding the CENR Task Force on Observations and Data Management (TFODM), and the U. S. Secretariat for Global Observing Systems; (3) Interagency coordination regarding the U. S. Global Change Research Program (USGCRP), in particular the Working Group on Observation and Monitoring (WGOM) headed by NASA; (4) Development of NASA's solar terrestrial/climate relations research through invited participation in the expert Panel established for the same; (5) Participation as member of the U. S. delegation to the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) to advise on Earth and environmental science issues and the preparation of UNISPACE-III; (6) Development of the GCRP long term scientific strategy plan and the program cross cut areas identified by the OMB and the OSTP; (7) Coordination with the World Climate Research Program and the Global Climate Observing System (GCOS) Program of the World Meteorological Organization; and strategy for the U. S. participation in the 15 year Global Ocean-Atmosphere-Land System (GOALS) project—a component of the international Climate Variability and Predictability Program (CLIVAR) Brief details follow:

1. IGOS activities covered interactions with the international Committee on Environment Observing Satellites (CEOS) which undertook the task of further developing and coordinating the space-based com-

ponent of IGOS through a set of specific projects such as climate and global climate change, the ozone issue, greenhouse gases, and terrestrial environmental monitoring vegetation and ecosystems, and natural disasters among others. In parallel, a multi-agency (of the U. N.) Global Observing Systems Space Panel (GOSSP) was established under the auspices of WMO/GCOS. GOSSP operated through virtual working groups covering satellite observing requirements, and observing/monitoring systems and networks. As a contribution of the U. S. to this international activity I drafted the first annual report for GOSSP in a format suitable for easy, periodic updates in the future.

Surface-based and in-situ components of IGOS are more complex because of the lack of international infrastructures for the coordination of networks and systems for most parameters and variables. In this regard, I completed the first "Compendium of Requirements and Systems for the Global Observing Systems." The Compendium was published by NASA (in January 1997) and also submitted to the U. N. as a U. S. contribution to the International Workshop on In-Situ Global Observing Systems held in Geneva, Switzerland. This compendium has since been used extensively in 1997 and 1998 by various organizational entities, including: The National Academy of Sciences (NAS), and the various working groups of the Global Change Research Program, The World Climate research Program (WCRP), and the World Meteorological Organization (WMO).

Work continues on the development of a global system for In-Situ observations needed for national (GCRP) and international programs. It is likely to continue for the foreseeable future.

2. CENR/TFODM and U. S. Secretariat for Global Observing Systems Programs: Continued coordination with agencies on national and global/international observing systems. Re-alignment of the activities of the TFODM to meet the needs of the U. S. Global Change Research Program. Coordination involved in the establishment of a new Working Group on Observations and Monitoring to serve the purposes of the CENR, IGOS and the GCRP. This represents the convergence of activities to develop a cohesive consolidated/integrated infrastructure to develop plans for long-term research and operational observing systems. Work in this area continues today and will for the foreseeable future.

NOTE: There is a strong overlap between this activity and that identified under item 1 above.

3. The U. S. Global Change Research Program (GCRP): The GCRP is an interagency effort to address a broad range of global change issues including: global climate change, the Ozone hole and ozone depletion, terrestrial and marine ecosystems, greenhouse gases, the impacts of short-term and long-term climate change, terrestrial and marine ecosystems, and others.

Following the international Kyoto Summit (attended by VP-Gore) led by the U. N. under the auspices of the U. N. Framework on Climate Change (UNFCCC) to which the U. S. is a signatory party, the OMB and the OSTP directed the GCRP to address the issues of: the carbon budget and carbon cycles, as also the convergence of climate change on seasonal-to interannual time scales with that of decades-to centuries, and ecosystems/impacts.

NASA has a substantial investment in the GCRP: approximately 70% of its total budget of about \$2 billion. The entire budget of the Earth Science Enterprise is identified as a NASA contribution to the USGCRP. A substantial activity has resulted from this new OMB/OSTP initiative, beginning in early 1998. Work in this area continues to date. A long-term scientific strategic plan needs to be developed as also the plans to reorganize the GCRP for the year FY2000. In particular, I have developed the sections of the plans on Observations and monitoring which comprises most of NASA's efforts in this area and most of its budget as well. I have also reviewed and commented on all other sections of the GCRP—many of which are highly deficient at this time.

4. Solar-Terrestrial/climate interactions: At the request of NASA HQ, I was initially involved (late 1997) in this project to jump start the Panel established by the National Academy of Sciences to conduct a study on the subject. Later (early 1998) I was invited as an expert to NASA's Panel to review propos-

als submitted to NASA for funding. The review meeting was held in April 1998. This project has attracted considerable scientific attention over the past two years, after lying dormant for about 20 years. A resurrection appears to be in progress on account of our current and planned capabilities (NASA led) to observe solar variability via space-based systems; in particular, multi-spectral observations.

5. UNISPACE-III and the U. N. Committee on the Peaceful Uses of Outer Space (UN- COPUOS): Invited in Fall 97 to be a scientific advisor to the U. S. delegation to the U. N. Committee on the Peaceful Uses of Outer Space (COPUOS). The U. N. Office of Outer Space Affairs is located in Vienna, Austria. Primary activities involved the planning of UNISPACE-III, the next major international conference on outer space activities to be held in July 1999. The central theme for this conference will be, by international agreement, Earth and environmental sciences and applications. For this I developed background scientific papers as NASA's input to the U. N. I was nominated by NASA HQ to be the scientific advisor to the U. S. delegation to COPUOS during their planning sessions in February 98, and June 98. Work continues on the planning of UNISPACE-III with NASA being the lead agency for the organization of parallel scientific and technical workshops and symposia during UNISPACE-III. Other matters involve reviewing the draft conference statement and conference report, which according to U. N. format needs to be approved by member countries before the conference. Activities included national and international interagency coordination, as well as coordination with various international associations involved/participating in remote sensing and outer space activities.
6. Long-term strategy for the global change research program: By statute the GCRP is required to have a long-term scientific strategy and plan. Through interagency coordination (and working groups), work was begun (97/98) in developing the long-term plan for science as well as observations and monitoring, the major component of the GCRP budget to address scientific needs. This has involved and continues to involve interactions with the GCRP WG on Observations, the OMB and OSTP. In parallel, the activity involves developing the program plans for FY2000. The GCRP WG on observations and monitoring is chaired by NASA. This activity overlaps somewhat with item# 1 and 2 detailed before, even though their objectives are somewhat different. Major issues involve the stability and continuity of operational observing systems, the transitioning of research and experimental systems and networks, and the continued development of new sensor technology and data management systems. Particularly problematic is the degradation of existing surface and in-situ operational systems which are outside the jurisdiction of the GCRP even though they play a vital role in the collection and provision of data for global change research. To address the new focused issues raised by the OMB/OSTP, substantial use will need to be made of NASA's next generation satellite systems such as TRMM (already launched), SeaWifs (already launched) and EOS AM, PM and CHEM. In addition, in-situ observations will also be required for measurement of terrestrial and ocean (including coastal zones) bioparameters and the fluxes of gases involved in the carbon budget/cycle. Integrating and combining space-based and in-situ observations will be a challenging task in some cases, while there has been demonstrable success in others (e.g., sea surface temperature. GRCP related work has involved a substantial interagency coordination activity.
7. WCRP, GCOS, GOALS and CLIVAR: CLIVAR is a major program of the international WCRP to address the broader issue of climate variability on all time scales. Various sub-programs of CLIVAR are directly aligned with the research objectives of the USGCRP. They also support the scientific work required by the Intergovernmental Panel on Climate Change (IPCC) which carries periodic assessments of the science of climate change and associated impacts. During most of 97/98 I was directly involved as a scientific advisor to the National Research Council (National Academy of Science) Panel on The Global Ocean-Atmosphere Land Systems (GOALS) project. In particular, the formulation of the US scientific strategy for participation in GOALS. GOALS is proposed as a 15 year research and experimental program directed at improving seasonal to interannual climate predictions. It will rely heavily on existing and proposed observing technologies. Embedded within the 15 year duration of GOALS, short-term field experiments are also proposed to investigate ocean-atmosphere-land interaction processes. A substantial modeling activity is called for. Other important programs of CLIVAR

include research on decadal and longer time scales as also investigations into climate history over the recent past using paleo records derived from tree rings, ice cores, coral cores, pollen, and fossil evidence. These data are crucial for the validation of climate models.

In addition to coordination activities in the Washington area, travel included two meetings (Feb. 98, June 98) of the UN/COPUOS in Vienna, Austria, meetings with the WCRP in Geneva, Switzerland (Feb. 98), and the presentation of a paper on Global Change (June 98) at the International Symposium on Remote Sensing of the Environment, Tromso, Norway.

## **Pilot EOS Direct Readout Ground Systems Support**

**Fran Stetina**  
**Fran Stetina and Associates**  
**(stetina@gsti.com)**

### **Objective**

This task provides technical support to develop the system design and implementation plans for NASA GSFC code 935 Regional Validation Centers to become pilot EOS Direct Readout Ground Receiving Stations and for these Centers to become regional MTPE product validation centers.

### **Background**

To effectively conduct research in global change problems and issues, it is necessary to solicit the cooperation of the broadest user community. To meet this long term outreach objective, NASA GSFC code 935 has developed the concept of Regional Validation Centers. This concept has been accepted as an effective approach to support the long term objectives of NASA's Mission to Planet Earth and to effectively transfer NASA's information technology to the broadest user community as well as to solicit the help of a broader community to both use and evaluate MTPE data products.

A number of these regional center are being implemented as prototype centers to test the effectiveness of new information system technologies under real operating conditions. During the next few years, emphasis will be concentrated on developing information system technologies to enhance reception and delivery of EOS direct readout data and validation of the products derived from this data for regional applications.

Understanding and evaluating new technologies such as hyperspectral instruments for agriculture and forestry monitoring will be in the forefront. In addition the use of unmanned aerial vehicles will play a substantial role in reducing costs and providing a valuable enabling and innovative technology.

Two of the Regional Applications Centers will continue to be important testbeds for new technology initiatives, these are:

1. University of Hawaii

In Hawaii a consortium of state, government, university, and private sector organizations is developing a concept called the Pacific Disaster Center(PDC); a regional validation center has been co-located with the PDC. Efforts undertaken in support of this activity are to define the relationship between the RVC and PDC.

## 2. University of Southwestern Louisiana

At the University of Southwestern Louisiana in Lafayette, a regional validation center has been established to concentrate on providing value-added weather products to the oil and gas industry. Emphasis will be on fusing all available weather products and providing new value-added products to minimize the impact of severe storms on the operations of the oil and gas industry as well as to determine the impacts of severe weather on the fragile coastal wetland areas.

These Centers would contribute to the efficient and effective utilization of human and natural resources and the development of an information infrastructure to support knowledgeable decision making. Such an infrastructure must not only gather and store data, but it must contain sufficient processing power and intelligence to produce useful output products. The system must facilitate rapid retrieval and distribution of information so that decision making can be done based on objective criteria using expert knowledge and simple visualization techniques. This philosophy requires a systems design approach which emphasizes integration, automation, user friendly interfaces, and a thorough understanding of the user's requirements.

Implementation of systems with these features is based on 10 years of project management experience for NASA GSFC in implementation of satellite weather receiving systems and ground processing. Specifically it includes the development of a modular system concept called SAMS, Spatial Analysis and Modeling System. The SAMS system has been defined as a potential model for the development of the MTPE Regional Validation and Calibration Center.

One of the key components of such a system is a real time direct readout capability. Thus, the design and development of the Regional Environmental and Technology Center concept (Regional Validation Center), has been defined as an important objective of NASA's Mission to Planet Earth.

As Co-Pi for the development of SAMS, I am uniquely qualified to apply this information to the design and development of the Regional Data Center Prototype System.

### Scope

The activities to be undertaken under this task include hardware and software system design which are required to develop a Prototype Regional Validation Center to support MTPE. Included in this concept is the need to develop a core EOS direct readout capability and general support of end-to-end system software to provide EOS core instrument algorithms and basic mission products. The system concept should include an archiving and distribution capability. In addition, strategies should be developed to test EOS Direct Readout System components and concepts in an operational environment. This includes the use of aircraft high spatial and spectral resolution instruments to support algorithm development and evaluation, integration of in-situ measurements to validate remote sensing measurements, and the integration of a Geographic Information System. In addition the system should include the design of a local user analysis system to interface with the Regional Data Center.

### Task Elements

- Provide expert advice to determine user requirements for EOS direct readout core instrument algorithms and products.
- Develop project plans to utilize hyperspectral instruments to facilitate the development of regional EOS MODIS algorithms.
- Determine weather product requirements for various applications which will be implemented at Regional Validation Centers for both operational and research users.

- Assist in defining MTPE core algorithm processing capability for a direct readout facility.
- Define the relationship and operating scenarios between the Pacific Disaster Center and the NASA Hawaii Regional Validation Center.
- Provide expert advice in defining EOS direct readout system concepts, define end-to-end system components and functions. Utilize SAMS concept to determine the requirements of Regional Validation Center.
- Develop operational scenarios for Direct Readout System and its interfaces with the GSFC EOSDIS.
- Provide expert advice in the development of strategies to develop and test various components of the Regional Data Center System using existing operational facilities at the University of Southwestern Louisiana and the University of Hawaii.
- Assist in the development of an implementation plan for the use of unmanned aerial vehicles to support MTPE regional algorithm development and MTPE product validation.
- Represent the NASA GSFC Regional Validation Center manager in meetings and conferences as required.

The Earth Alert personal warning system has been defined as a potential important technology which has significant value to the Hawaii Pacific Disaster Center. A number of activities relating to bringing this technology to a successful commercial product line and introduction of this capability to The Hawaii Civil Defense and to FEMA have been undertaken under task #61.

## **Accomplishments**

- Under the activities of this task a Memorandum of Understanding between GSFC and FEMA has been implemented.
- A strategy for the testing of the Earth Alert personal warning system has been developed, in partnership with FEMA, NOAA Weather Radio, and A State Emergency Operation Center.
- A plan has been developed to extend this warning system to a more general information dissemination system called Weather Anywhere. The system will be demonstrated at the Annual Air Traffic Control Association Conference in Atlantic City in November 1998.
- A plan is being developed to implement various GSFC information system technologies to support the emergency management community.
- An MOU has been developed between Freewing Aerial Robotics Corporation and GSFC to test their UAV for GSFC remote sensing applications. The plane will be test flown at Wallops Flight Center in September 1998.
- Various low cost aircraft hyperspectral instruments have been identified and are under field investigations to provide information which will be useful in developing EOS direct readout data products for regional applications.
- Two pilot EOS direct readout satellite stations are being implemented at GSFC for future use at RAC's.
- New instruments which will fly on the Freewing UAV have been identified and field demonstrations have been proposed and are planned.

## NASA SCIENCE AND THE PRIVATE SECTOR

**Murray Felsher**  
**Associated Technical Consultants (ATC)**  
**(felsher@tmn.com)**

The essence of the activity between Associated Technical Consultants and the National Aeronautics and Space Administration (NASA) through a subcontract between ATC and USRA's Center of Excellence in Space Data and Information Sciences (CESDIS) was to establish a mechanism for a suitable and ongoing interface between NASA and the remote sensing private sector, as well as provide, by example, an indication to NASA that its own remote sensing science and technology, and that of its contracting personnel, can be applicable and of substantial interest to other federal agencies. NASA, recognizing the depth of ATC contacts with the commercial remote sensing community, requested that I participate in several related activities as well. The following are accomplishments and results of these efforts.

### **1. NASA/Industry Workshop on MTPE's Commercial Strategy**

Initial ATC efforts were devoted to working with Mission to Planet Earth (MTPE, now Earth Science) senior staff in planning and developing a program and agenda for the subject Workshop, held 22-23 July 1996 at the Greenbelt, Maryland Marriott. NASA Administrator Goldin keynoted the workshop, and more than 70 senior industry representatives and 50 senior NASA managers were invited and attended. This was the first time that NASA had attempted such a workshop on this scale and at that management level. I served on the Program/Organizing Committee of the Workshop, and there provided the industry point-of-view. I was also responsible for nominating and inviting the majority of the industry participants in two of the four Workshop panels: the Data Provider and the Value-Added industry panels. Ground rules allowed for both government and industry participants to undertake frank, in-depth discussions that have served as the basis for building current government/industry levels of trust.

### **2. Workshop on Water Monitoring, Remote Sensing, and Advanced Technologies**

In order to allow NASA to highlight the "relevance for public good" of its remote sensing science and technology programs, ATC proposed a strategy for a Workshop involving a sister federal agency, EPA. The Workshop took place 11-13 December 1996 at the Holiday Inn Southwest, in Washington, DC. The expressed purpose of the Workshop was to expose technical and management personnel of both agencies to (1) NASA's remote sensing science and technology, and (2) EPA's water resources monitoring requirements and databases. The goal of the Workshop was mutual education and the opportunity to explore future collaboration in water monitoring/remote sensing research and applications. The success of the workshop has subsequently led to several joint activities by EPA and NASA.

### **3. Visits to Industry by Government Representatives**

In order to provide NASA Headquarters and NASA center personnel engaged in earth remote sensing programs a sense and appreciation of those same and similar activities as concurrently being undertaken by the private sector, ATC planned and organized a series of visits to local (Washington DC-area) remote sensing industry sites. Included were all segments of the remote sensing continuum. That continuum ranges from companies primarily residing in the space segment—actually acquiring space-derived imagery—through the ground segment. The latter includes both value-added firms—firms that enhance the space-derived data and transform it into useful image information, and firms that provide hands-on training for image analysts. Specifically, NASA Code YS Earth Science Director and his headquarters program

managers were invited to participate in the visitation program. It was felt that by introducing government science, technology, and management personnel to corporate methodologies, industry research priorities, and management organization and techniques, the NASA executives would be exposed to a whole new set of work parameters to which they were not at all familiar. The converse was true, as well. NASA, as well as the rest of the administration is on record as seeking a closer working relationship with the private sector. For NASA this meant seeking new and creative fiscal instruments, as Cooperative Research and Development Agreements (CRADAs), and the establishment of other joint government/industry activities. In addition, by bringing NASA science managers into key commercial remote sensing workplaces, the industry executives were themselves able to closely question key NASA officials, face-to-face, on a one-on-one basis—sometimes for the very first time. The visits also allowed NASA to appreciate its own R&D efforts and how the industry views those efforts. NASA had the opportunity to incorporate, into its own R&D planning, the private sector's vision of those facets of NASA R&D/technology that could be labeled as pre-competitive, and thus within NASA's purview. An attempt was continuously made to define that point (or range) at which such R&D could be called commercially viable, and thus outside of NASA's purview. Such agreement proved to be elusive, but a first step has been taken. At the very least, the visits permitted industry and government to delineate and minimize programmatic gaps and duplication of effort between the two institutions. As such, it served as an excellent starting point for upcoming cooperative goals, such as ensuring a maximum return of budgetary and intellectual investment for both parties.

#### **4. Monthly NARSIA Briefings**

At the behest and request of NASA, I was tasked, through the membership of the North American Remote Sensing Industries Association (NARSIA) to initiate and chair a monthly briefing seminar series. These one- and two-day briefing sessions were held from July 1997 to January 1998. Intense and comprehensive briefings of NASA's entire (then) Mission to Planet Earth program were presented to industry representatives. High ranking NASA Headquarters and Center officials briefed an eager set of industry participants. Although the briefings themselves were on-the-record, with hard-copy of the NASA viewgraphs being made available to all, ample time was provided for an intense dialogue, and the resulting discussions were informal and off-the-record. The resulting frank and candid give-and-take went far to establish a strong rapport between the NASA and commercial executives present. The result of this exercise has been a solid sense of trust that has developed between this cadre of government and industry individuals. The point was made and accepted that NASA must not position itself to compete with the U. S. remote sensing industry. For its part, industry should provide NASA with whatever non-proprietary information it needs to allow NASA to accomplish its goal of achieving and maintaining U. S. primacy, among space-faring nations, in remote sensing science and technology research.

#### **5. Applications**

Significant assertions were made by industry, during both the visits and monthly briefings, in an attempt to convince NASA Headquarters to move beyond funding remote sensing basic research, per se, and to initiate a serious program in remote sensing applications research. NASA, it was felt, had to establish itself to both the general public as well as the remote sensing industry as being not only a highly motivated research entity, but also a government organization that conducts programs of immediate relevance to the daily lives of the citizenry. Such relevance, albeit just recognizable in far-field efforts as basic Global Change Research, becomes very much more cogent when the programmatic areas are clearly in the field of applications research. These include, but are not limited to: natural hazards mitigation and assessment; environmental monitoring; land use planning; coastal zone management; facilities siting; agricultural crop yield and production estimates; forest inventory; mineral/petroleum exploration and assessment; and the like. At our briefing sessions, one of our initial briefers was Dr. Ghassem Asrar, then Senior Scientist at NASA/YS. Dr. Asrar has since been named Associate Administrator for Earth Science, and we are delighted to have learned that one of his first actions has been to create an Applications Division within the Office of Earth Science. We will do whatever is necessary to assure that this new Division will receive the

support it needs from industry to become a viable force in moving NASA remote sensing research into new areas of relevance.

## 6. A NASA Commercial Office

One of the successes of the newly-created (1997) National Imagery and Mapping Agency (NIMA), an agency of the Department of Defense, has been the close working relationship quickly established between NIMA and its industrial base. Much of the credit for this success can be placed with the (then) NIMA Director, Admiral Jack Dantone. Admiral Dantone was a staunch supporter of the companies that served as his agency's systems integrators, contractors, and suppliers. His speeches were laced with pleas for closer ties between his managers and the industries that served them. The parallel between NIMA Director Dantone and NASA Administrator Goldin is, in that respect, altogether striking. Mr. Goldin has often gone on record as advocating more creative joint activities with industry. But perhaps most important, insofar as NIMA's rapid success in incorporating his industrial base into his own vision, was the creation, by the Admiral, of a Commercial Office whose Head reported directly to him. By designating a commercial ombudsman, as it were, on so high an administrative level, the word quickly moved through the NIMA bureaucracy that industry is to be taken seriously. Our discussions with NASA have resulted in a recommendation that Mr. Goldin likewise establish a Commercial Office within the Administrator's organization. The Director of that office would rapidly be able to bring to the attention of all NASA offices, within the roles and missions of those offices, those aspects of commercial activities that would enhance both NASA and the private sector organizations that serve it.

## SUPPORT FOR NASA/NOAA COLLABORATION IN EARTH SCIENCE MODELING

The goal of the Earth and Space Science (ESS) Project in the NASA High Performance Computing and Communications (HPCC) Program is to accelerate the development and application of high performance computing technologies to meet the Grand Challenge needs of the U. S. Earth and space science community. One approach being taken by ESS is to provide testbed access to Guest Investigators in the broader Earth and space science community to prepare for the next generation of high-end production computers.

NOAA expects to procure a Class-8 supercomputer for operational weather prediction work which will be installed at NASA GSFC in FY99. This system is planned to have a scalable parallel architecture, but NOAA's current operational codes are designed and optimized for execution on conventional vector machines.

The purpose of the task reported upon here is to support the NASA effort by providing ESS with the services of a senior Earth modeler who is a member of the ESS Inhouse Team of computational scientists who will provide expert guidance, support, and enhanced communications and cooperation between the ESS Project and the selected Guest Investigators. The same individual is to assist NOAA in carrying out code development/conversions and performance experiments using the current generation scalable parallel SGI/Cray T3E installed at GSFC in support of HPCC/ESS Grand Challenge Investigators.

**Miodrag Rancic**  
**University of Maryland Baltimore County**  
**Department of Computer Science and Electrical Engineering**  
**mrancic@ciga.gsfc.nasa.gov**

## 1. Quasi-uniform Grids for Global Models of the Atmosphere on MPPS

Ever since I joined CESDIS in November 1997, I have tried to put my scientific objectives in accordance with the HPCC project to which I now belong, and to establish a sound foundation for my future work.

Technically, I continued my research on nonstandard, quasi-uniform spherical grids for global models of the Earth's atmosphere, which I began while working at NCEP (Rancic et. al 1996, Purser and Rancic 1997, 1998). These grids (conformal, gnomonic and smoothed cubic and octagonal) appear to generally have some special advantages, particularly when applied on the massively parallel processors (mpps). In the past, I was developing and testing these grids using only the grid-point Eulerian numerical technique. Here, at CESDIS, I have begun the development of a rational strategy for a global semi-Lagrangian model on mpps the using concept of quasi-uniform grids. At the same time, however, I have continued to search for the most appropriate solution for certain problems which 3D integrations revealed that exist around the singular points of these grids within the Eulerian grid-point method.

Generally, I am trying to combine the technique of grid overlapping (blending) (e.g., Chesshire and Henshaw 1990) with the concept of quasi-uniform grids, for both Eulerian and semi-Lagrangian approaches. More precisely, I am developing two new global models of the atmosphere for integration on the high performance massively parallel computers, with the following major features:

- a semi-Lagrangian global model; the gnomonic cubic grid; medium resolution; for modeling of climate;
- a Eulerian, grid-point, nonhydrostatic global model; conformal octagonal grid; a very high resolution; for weather forecasting.

## 2. Semi-Lagrangian Model

Semi-Lagrangian techniques, which have been dynamically developed in meteorology in the last two decades, allow using the time-step which is no longer restricted by CFL linear stability conditions. Though treatment of the poles did not represent a serious problem for this technique on the vector processors, this situation has dramatically changed on the massively parallel platforms. One original solution to avoid these problems is to apply a gnomonic cubic grid and the method of blending of the domains, suggested by Ronchy et al. (1996), in combination with the cascade interpolation method of Purser and Leslie (1991).

Blending on the gnomonic cube requires only application of 1D Lagrangian interpolations (or, alternatively, any 1D interpolations). Similarly, the cascade method for SL schemes consists of a sequence of 1D interpolations that are applied in order to estimate value of advected variables at the departure points. This reduces the amount of calculations, (and, on mpps, proportionally the number of communications), from  $N \times N$  to  $2N$  (in 2D case). At the same time, this approach is establishing the road which should be followed in order to achieve global conservation and monotonicity of advected fields within the SL model (Rancic, 1995). I have developed such an advection algorithm, as the starting point for this project, and it is now being tested on the Cray T3E.

The other important features of the dynamics of this model are:

- A so-called "strong conservative form" of the governing equations (Sharman et al. 1988), written in the general curvilinear coordinates on the sphere, where the historic variables are longitudinal and latitudinal velocity components.
- PVD (potential vorticity and divergence) formulation, following approach of Bates et al. (1995);
- A hybrid sigma-theta vertical coordinate (e.g., Konor and Arakawa, 1997).

So far, I have derived the strong conservative form for the shallow water equations on the sphere. The PVD formulation will require solving of a specific nonlinear elliptic solver on the cubic grid, and this part of the project is being done in collaboration with Dr. J. Robert Purser from NCEP.

### 3. Eulerian Nonhydrostatic Model

The core of the major regional model of NCEP (Eta) will be used for this project. A full 3D hydrostatic version of this model, which employs an octagonal grid, has already been finished and tested.

The observed problems around singularities will be solved in the manner that may be referred to as "patching" of the singularities, by using a blending approach. Eight patching domains, each covering the region around a singular point, will be relatively smaller than the rest of the computational domains, so that in addition to calculation, which should be consequently finished faster, the processors assigned to the patches will also prepare blending. Thus, unlike the standard longitude-latitude grid where the major portion of PEs has to wait for a relatively small number of PEs around the poles to finish polar filtering, here a small number of processors is supposed to finish the job somewhat faster than the rest of them.

This should be an interesting new way to deal with load balancing in a situation where a very large number of processors is used for calculation.

The remaining important issue concerning this project is the treatment of the vertical acceleration and the choice of the vertical coordinate. I have developed a consistent numerical approach, based on the method suggested by Laprise (1992), which allows using a pressure-based vertical coordinate. Presumably, this approach, among other possible advantages, should allow a relatively simple adding of nonhydrostatic effects to the hydrostatic model. In addition, I am following the latest developments in this area, and I might also consider using a quasi-nonhydrostatic formulation, depending on the results that a combined group from NCAR and NOAA will derive.

### Publications, Presentations, Other Activities

On February 26 I conducted a poster presentation (Goddard Teas and Posters) in the Atrium of Building 28, with the title: "Quasi-uniform square grids for global simulation of atmosphere on massively parallel processors".

During this period, a paper that I co-authored and one workshop presentation of mine, have been published:

Purser, R. J. and Rancic, M. (1998). Smooth quasi-homogeneous gridding of the sphere. *Quart. J. Roy. Met. Soc.*, 124, 637-647.

Rancic, M. and Baillie, C. (1996) Cubic and octagonal spherical grids on MPPs. In G.-R. Hoffmann and N. Kreitz (Ed.), *Making its Mark*, World Scientific, 492.

I have also reviewed an article submitted to *Monthly Weather Review*.

## References

- Bates, J. R., Li, Y., Brandt, A., McCormick, S. F. and Ruge, J. (1995) A global shallow-water numerical model based on the semi-Lagrangian advection of potential vorticity. *Quart. J. Roy. Meteor. Soc.*, 121, 1981-2005.
- Cheshire, G., and Henshaw, W. D. (1990) Composite overlapping meshes for solution of partial differential equations. *J. Comp. Phys.*, 90, 1-64.
- Konor, C. S., and Arakawa, A. (1997) Design of an atmospheric model based on a generalize vertical coordinate. *Monthly Weather Review*, 125, 1649-1673.
- Laprise, R. (1992) The Euler equations on motion with hydrostatic pressure as an independent variable. *Monthly Weather Review*, 120, 197-207.
- Purser, R. J., and Leslie, L. M. (1991) An efficient interpolation procedure for high-order three-dimensional semi-Lagrangian models. *Monthly Weather Review*, 119, 2492-2498.
- Purser, R. J., and Rancic, M. (1997) Conformal octagon: An attractive framework for global models offering quasi-uniform regional enhancement of resolution. *Meteor. Atmos. Phys.*, 62, 33-48.
- Rancic, M. (1995) An efficient, conservative, monotonic remapping for semi-Lagrangian transport algorithms. *Monthly Weather Review*, 123, 1213-1217.
- Rancic, M., Purser, R. J. and Mesinger, F. (1996) A global shallow water model using an expanded spherical cube: Gnomonic versus conformal coordinates. *Quart. J. Roy. Meteor. Soc.*, 122, 959-981.
- Ronchy, C., Iacano, R., and Paolucci, P. S. (1996) The 'cubed sphere': A new method for solution of partial differential equations in spherical geometry. *J. Comput. Phys.*, 124, 93-114.
- Sharman, T. D., Leller, T. L., and Wurtele, M. G. (1988) Incompressible and anelastic flow simulation on numerically generated grids. *Monthly Weather Review*, 116, 1124-1136.

## OTHER RESEARCH AND DEVELOPMENT PROJECTS

### *Development of an Implicit 2D Adaptive Mesh Refinement and De-refinement Magnetohydrodynamics Code*

Dinshaw Balsara  
National Center for Supercomputing Applications  
University of Illinois  
(u10956@ncsa.uiuc.edu)

#### Statement of Work

This task required CESDIS to provide consulting support to develop a linearized Riemann solver for numerical MHD for work with NASA scientists which was to be based on developing code that concretized ideas contained in a paper by Dr. Dinshaw Balsara entitled *The Linearized Formulation of the Riemann Problem for Adiabatic and Isothermal MHD*. Dr. Balsara was retained on a consulting basis to perform this work.

#### Report

In this work I was asked to focus primarily on two issues—the pressure positivity preserving strategy for the MHD equations and the CGL equations.

First I will describe the work that was done to build a pressure positivity conforming methodology for MHD:

1. I wrote out the entropy advection equation. Using that and the other equations of MHD, I obtained a full hyperbolic system of equations for MHD.
2. I analyzed the above eigensystem. I coded up the eigenvectors into an eigenvector subroutine. I made sure it works right.
3. I built a Riemann solver that incorporates this eigenvector package and computes fluxes for this modified system of MHD. A suitable entropy fix was built to go along with this Riemann solver.
4. I built a set of switches that make the decision on when to turn on the Riemann solver that ensures pressure positivity in the `oned_tvd.f90` solver subroutine.
5. I built and ran a few different problems to show that strong Alfvénic discontinuities can be accurately advected by the code. This was done for very small plasma betas showing that the code works.

For the CGL equations the following tasks were carried out:

1. Built eigenvector module for CGL equations.
2. Built eigenvalue module for CGL equations.
3. Tested (1) and (2) for orthonormality and completeness.

4. Built symmetrized linearized Riemann solver for the CGL equations. Built it so that it naturally incorporates an Einfeldt scheme in its body.
5. Built Muscl scheme with interpolation on the primitives. The Hancock half step was incorporated.
6. Incorporated it into the AMR module from Goddard.
7. Incorporated the Riemann solver into the CGL module in a way that allows the constant and fluctuating parts of the magnetic field to be split off. The Riemann solver also gives the primitives at the zone boundaries.

## **Publications**

The following papers are available as CESDIS technical reports.

### **TR -98- 216**

#### **Analysis of the Eigenstructure of the Chew, Goldberger, and Low System of Equations**

The Chew, Goldberger, and Low (CGL) system of equations applies to several situations in magnetospheric physics. It is based on making a double adiabatic approximation for the thermal pressure. In this paper we derive the eigenvalues and a complete set of left and right eigenvectors for the CGL system. The system admits eight eigenvalues, seven of which have analogues in ideal MHD. An eighth eigenvalue turns out to correspond to a new kind of advected wave. This wave produces magnetic fluctuations, but the magnetic pressure is balanced by the corresponding thermal pressure fluctuation produced by the fact that the thermal pressures are anisotropic. This wave corresponds to a linearly degenerate wave. The eigenvectors for the magnetosonic waves become singular in certain limits. These are identified and eigenvector regularization is done where needed. Intuitive insights pertaining to the nature of the waves are developed. This is especially true for the eighth wave. In the regime of validity of the double adiabatic approximation, the wave speeds show a strict ordering. This makes the CGL system amenable to numerical solution using upwind schemes. The linear degeneracy of the eighth wave suggests that it might be treated differently in the context of upwind schemes. Several important parallels as well as some important points of difference between the CGL system of equations and ideal MHD equations are pointed out throughout the paper.

### **TR - 98 - 217**

#### **Maintaining Pressure Positivity in Magnetohydrodynamic Simulations**

Higher order Godunov schemes for solving the equations of Magnetohydrodynamics (MHD) have recently become available. Because such schemes update the total energy, the pressure is a derived variable. In several problems in laboratory physics, magnetospheric physics, and astrophysics, the pressure can be several orders of magnitude smaller than either the kinetic energy or the magnetic energy. Thus small discretization errors in the total energy can produce situations where the gas pressure can become negative. In this paper we design a linearized Riemann solver that works directly on the entropy density equation. We also design switches that allow us to use such a Riemann solver safely in conjunction with a normal Riemann solver for MHD. This allows us to reduce the discretization errors in the evaluation of the pressure variable. As a result, we formulate strategies that maintain the positivity of pressure in all circumstances. We also show via test problems that the strategies designed here work.

TR - 98- 218

## **A Staggered Mesh Algorithm Using High Order Godunov Fluxes to Ensure Solenoidal Magnetic Fields in Magnetohydrodynamic Simulations**

The equations of Magnetohydrodynamics (MHD) have been formulated as a hyperbolic system of conservation laws. In that form it becomes possible to use higher order Godunov schemes for their solution. This results in a robust and accurate solution strategy. However, the magnetic field also satisfies a constraint that requires its divergence to be zero at all times. This is a property that cannot be guaranteed in the zone centered discretizations that are favored in Godunov schemes without involving a divergence cleaning step. In this paper we present a staggered mesh strategy which directly uses the properly upwinded fluxes that are provided by a Godunov scheme. The process of directly using the upwinded fluxes relies on a duality that exists between the fluxes obtained from a higher order Godunov scheme and the electric fields in a plasma. By exploiting this duality, we have been able to construct a higher order Godunov scheme that ensures that the magnetic field remains divergence free up to the computer's round-off error. Several stringent test problems have been devised to show that the scheme works robustly and accurately in all situations. In doing so it is shown that a scheme that involves a collocation of magnetic field variables that is different from the one traditionally favored in the design of higher order Godunov schemes can nevertheless offer the same robust and accurate performance of higher order Godunov schemes provided the properly upwinded fluxes from the Godunov methodology are used in the scheme's construction.

### ***A Scalability Model for ECS's Data Server***

**Daniel A. Menascé**  
**George Mason University**  
**Department of Computer Science**  
**(menasce@cne.gmu.edu)**

**Mukesh Singhal**  
**Ohio State University**  
**Department of Computer and Information Science**  
**(singhal@cis.ohio-state.edu)**

## **Statement of Work**

The objective of this study is to carry out an analysis to determine if the current ECS Data Server design is scalable to the near and far term data volume requirements of EOSDIS.

## **1. Introduction**

Perhaps one of the most important examples of large-scale data intensive, geographically distributed information systems is NASA's Earth Observing System (EOS) Data and Information System (EOSDIS). EOS is a NASA mission aimed at studying the planet Earth. A series of satellites with scientific instruments aboard will be collecting important data about the Earth's atmosphere, land, and oceans over a period of 15 years. This mission will generate an estimated terabyte/day of raw data which will be processed to generate higher level data products [2]. Raw data received from the satellites is first stored as Level 0 (L0) data which may then be transformed after successive processing into levels 2 through 4 (L2 - L4). Data

received from the satellites and data products generated from them will be stored at various Distributed Active Archive Centers (DAACs) located throughout the United States. An important component of a DAAC is the Data Server—the subsystem that stores and distributes data as requested by EOSDIS users.

The Data Server stores its information using a hierarchical mass storage system that uses a combination of automated tape libraries and disk caches to provide cost-effective storage for the large volumes of data held by the Data Server. Performance studies and workload characterization methods and software for hierarchical mass storage systems are reported in [3, 5, 6, 7, 8].

In this report, we present a model for the scalability analysis of the Data Server subsystem of the EOSDIS Core System(ECS). The goal of the model is to analyze whether the planned architecture of the Data Server will support an increase in the workload with the possible upgrade and/or addition of processors, storage subsystems, and networks. This analysis does not contemplate new architectures that may be needed to support higher demands.

The remaining sections of this report are organized as follows. Section two provides a summary of the architecture of ECS's Data Server as well as high level description of the Ingest and Retrieval operations as they relate to ECS's Data Server. This description forms the basis for the development of the scalability model of the data server. Section three presents the scalability model and the methodology used to solve it. This section describes the structure of the scalability model, input parameters, algorithms for computing parameters of the scalability model solver, algorithms for solving the scalability model, and the assumptions and rationale behind these assumptions. The scalability model takes into account the proposed hardware and software architecture. The model is quite general and allows the modeling of data servers with numerous configurations.

## **2. Ingest and Retrieval Operations**

This section provides a high level description of the Ingest and Retrieval workloads of the ECS's Data Server. This description forms the basis for the development of a model to analyze the scalability of the Data Server. The scalability analysis entails determining whether the current architecture of the ECS Data Server supports an increase in the workload intensity with possibly more processing and data storage elements of possibly higher performance.

### **2.1 Subsystems of the Data Server**

The following subsystems of the Data Server will be considered for the purpose of the scalability analysis considered in this study:

#### **Software Configuration Items:**

- Science Data Server (SDSRV): responsible for managing and providing access to non-document Earth science data.
- Storage Management (STMGT): stores manages and retrieves files on behalf of other SDPS components.

#### **Hardware Configuration Items:**

- Access Control and Management (ACMHW): supports the Ingest and Data Server subsystems that interact directly with users. Of particular interest here is the SDSRV.
- Working Storage (WKSHW): provides high performance storage for caching large volumes of data on a temporary basis.

- Data Repository (DRPHW): provides high capacity storage for long-term storage of files.
- Distribution and Ingest Peripherals (DIPHW): supports ingest and distribution via physical media.

## 2.2 Ingest Data Operation

The diagram in Figure 1 depicts the flow of control and data for the Ingest process. We have not included Document Repository nor the Document Data Server due to their small impact on scalability if compared with ingest of L0 data. Circles in the diagram represent processes. The labels in square brackets beside each process indicate the hardware configuration item they execute on. Bolded labels indicate hardware configuration items that belong to the Data Server.

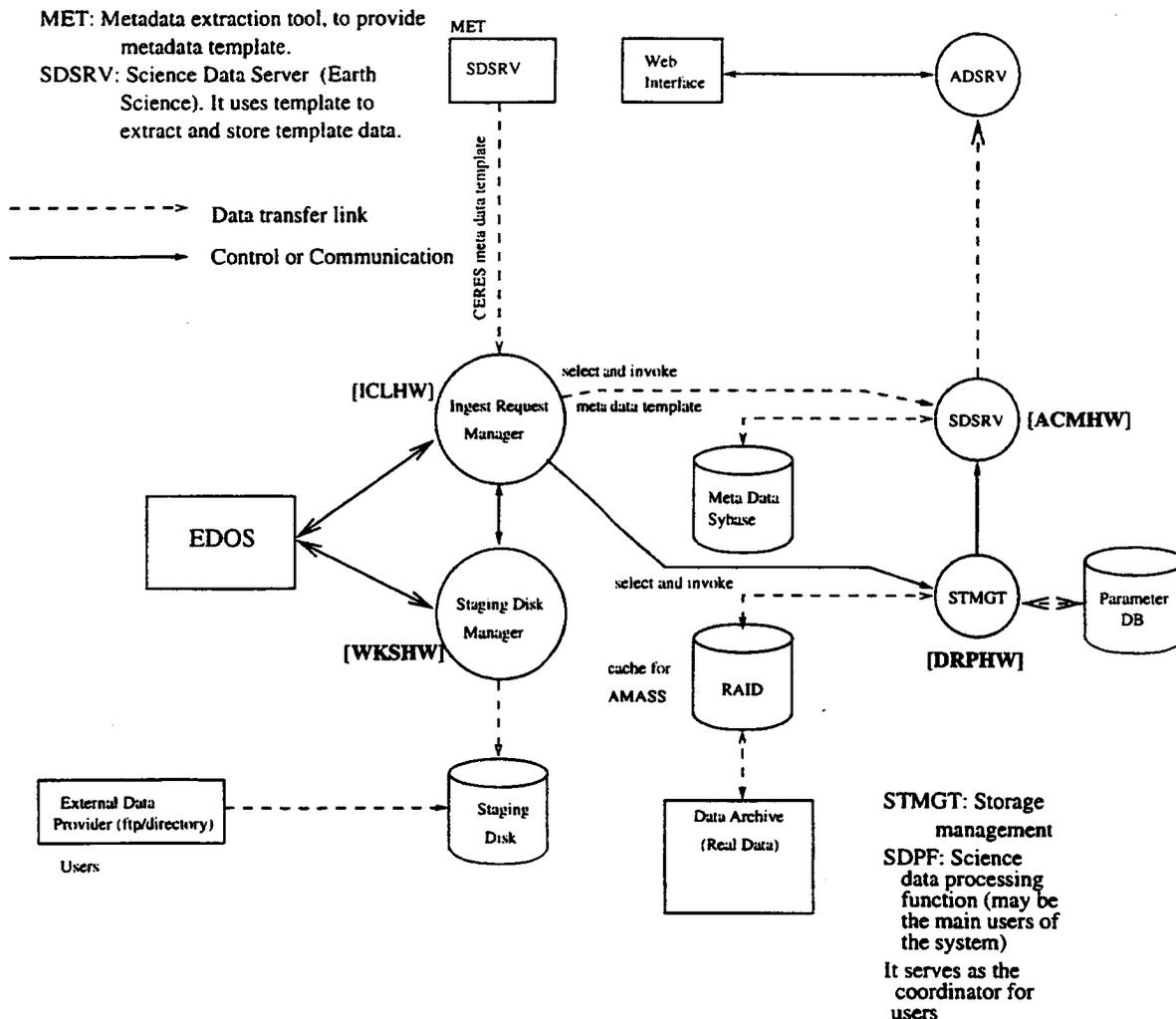
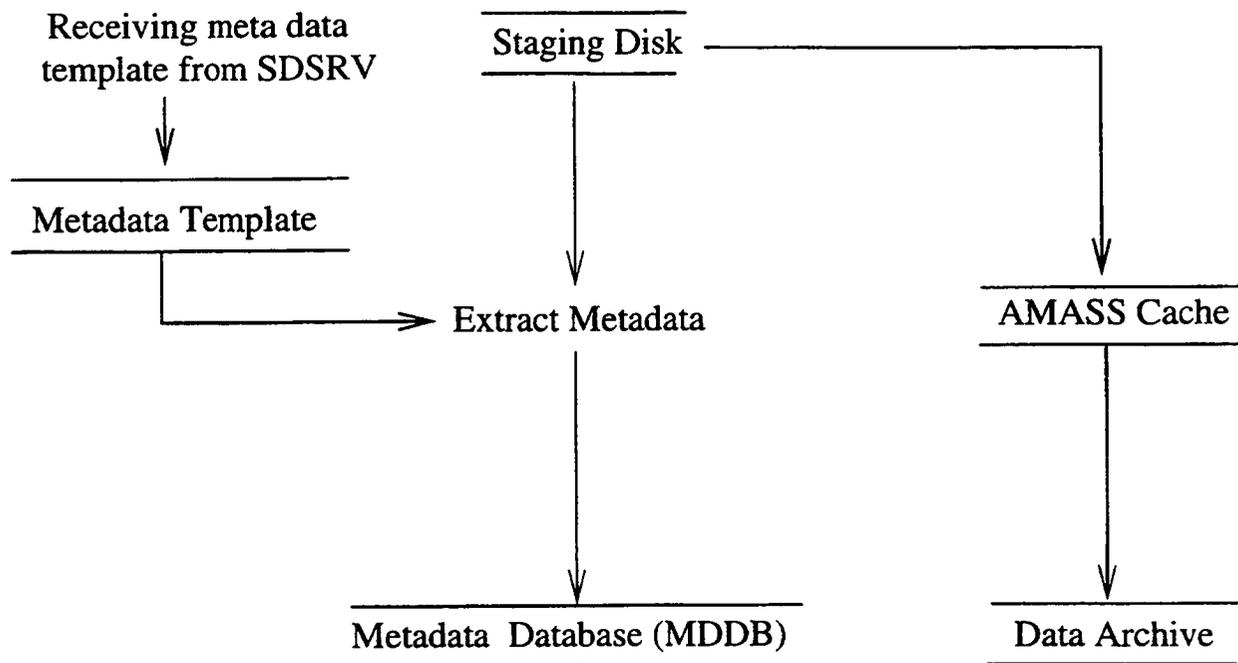


Figure 1: L0 Ingest Control and Data Flow

The main aspects of the diagram of Figure 1 are discussed below:

- Incoming L0 data is first stored into the system on the Staging Disk, and then into AMASS's cache—the hierarchical mass storage systems' disk cache for files. The metadata are extracted and entered into a Metadata database managed by Sybase and the actual data are archived. This is depicted in Figure 2.



**Figure 2: Data Flow Diagram for Ingest Data**

- The SDPF (Science Data Processing Function) represents the users of the Ingest system and negotiates with the Ingest Request Manager for coordination of transferring data into the Ingest system.
- Data are initially entered through an interactive GUI interface, or most of the time from external data providers through ftp or direct transfer of files, if that is done on the same local network, into the Staging Disk.
- The actual data is then transferred into AMASS' disk cache. From the cache, the data migrates to robotically mounted tapes managed by AMASS. The metadata extracted from the data is stored into a Metadata Database managed by Sybase.
- The SDSRV (Science Data Server) gives the metadata templates to the Ingest Request Manager for it to extract metadata.
- There are two and sometimes three SDSRV's and one STMG (Storage Management) processes. The Ingest Request Manager process selects which SDSRV to use.

The scalability analysis will, among other things, determine possible performance bottlenecks. The staging disk, the AMASS disk cache, and the metadata extraction process are likely candidates for bottlenecks.

### 2.3 Retrieval Operation

This section examines the retrieval and processing operation on L1+ data. Figure 3 depicts the flow of control and data for this operation. Circles in the diagram represent processes. The labels in square brackets beside each process indicate the hardware configuration item they execute on. Bolded labels indicate hardware configuration items that belong to the Data Server.

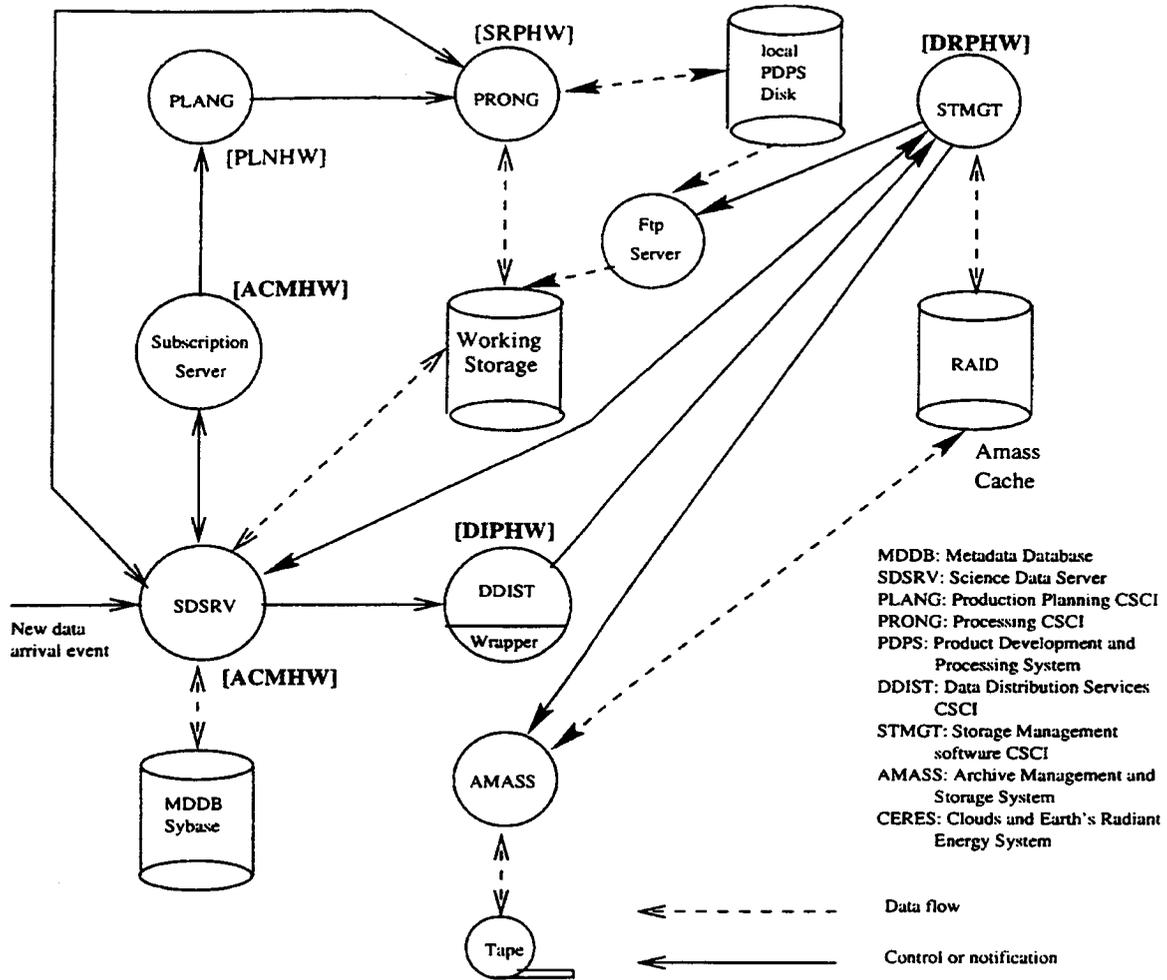


Figure 3: A Flow Diagram of Data Retrieval and Processing

The retrieval operation proceeds in the following three stages:

Stage 1: Checking data and deciding what processing is required

- SDSRV initiates the retrieval process by notifying the Subscription Server of the new data arrival.
- The Subscription Server performs a subscription check for this data and performs an appropriate notification, e.g., email notification, etc.
- The Subscription Server notifies PDPS PLANG of new data arrival.
- PLANG figures out (e.g., retrieves) a processing plan and based on the processing plan, passes the processing request to PRONG.

- PDPS PRONG connects to the appropriate SDSRV (may not be the SDSRV which initiates the retrieval and processing operations).

#### Stage 2: Retrieving data

- The SDSRV requests that the Data Distribution Services CSCI (DDIST) retrieves the data files
- SDSRV  $\longrightarrow$  requests DDIST  $\longrightarrow$  requests STMGT. The STMGT retrieves the files from AMASS archive into the AMASS cache if it is already not present in the cache.
- SDSRV notifies PRONG of data (identified by UR) availability.

#### Stage 3: Processing data and archiving, both data and metadata

- PRONG transfers the retrieved data from the Working Storage to local PDPS disk. (If the AMASS cache and Working Storage are on different devices, then data must be first moved from the former to the latter.)
- PRONG processes the retrieved data to produce a higher level product.
- PRONG processes the data to a higher-level product and extracts metadata from the higher-level data using the Metadata Extraction Tool and populates the target metadata template and writes a metadata file (on MDDB Sybase).
- PDPS PRONG sends an insert request to SDSRV.
- SDSRV  $\longrightarrow$  requests STMGT  $\longrightarrow$  requests AMASS. The AMASS file manager archives the files. Archiving is done in two steps:
  - STMGT copies data from PDPS (local disk) to Working Storage via an ftp command.
  - data are copied from the Working Storage to AMASS cache (and then to AMASS archive).
- SDSRV inserts metadata in the Metadata Database (MDDB) and then notifies PRONG that the archival operation has been completed.

## 2.4 Assumptions

The various software processes shown in the previous subsection were mapped into the different hardware configuration items for the GSFC, EDC, and LaRC DAACs. The following assumptions were made when developing the scalability model.

- Processing of "Ingest data" and "Data retrieval and processing" constitute the main load on the Data Server. Thus, we modeled only these two operations.
- We did not model users' requests for data to be subsetted or subsampled nor did we model compressed data.
- In data retrieval operations, PLANG retrieves a processing plan from a database (e.g., Sybase).
- The AMASS cache and the working storage may be implemented on the same disk.

- Servers that are not potential bottlenecks were not considered in the model. Examples include the "subscription server" and PDPS.
- We assume that mean arrival rate of both types of requests (ingest data and data retrieval) and service demands of these requests at various service stations are available or can be easily estimated.

### 3. A Scalability Model

We now describe our scalability model for the ECS's Data Server and our methodology for solving this model. We describe the structure of the scalability model, input parameters, algorithms for computing parameters of the scalability model solver, and algorithms for solving the scalability model. We describe our assumptions and rationale for these assumptions.

The scalability model is based on our understanding of the architecture of ECS Data Server and the Ingest and Retrieval operations described in the previous section. The sole purpose of the model is to analyze the scalability of the Data Server, i.e., to determine whether the current architecture of the ECS Data Server can support an increase in the workload intensity.

#### 3.1 A Framework for Scalability Analysis

Figure 4 gives the structure of the scalability model. The "Scalability Model Generator" collects information from three input files (these files define the modeling information on the ECS's data server and the workload) and processes this information to create an output file which contains inputs to the "Scalability Model Solver". This solver uses queuing network [4] techniques to obtain desired performance measures such as response times per workload, device utilizations, bottleneck indications, and queue lengths.

The first input file to the Scalability Model Generator, "Hardware Objects", defines the hardware resources (e.g., processors, disks, networks, and tape libraries) of the Data Server. The second input file to the Scalability Model Generator, "Workloads and Execution Flow", completely characterizes the workload that drives the Data Server. The third input file to the Scalability Model Generator, "Processes", defines the parameters of the software modules that will be executed on hardware servers by arriving requests for service (i.e., the workload).

The Scalability Model Generator reads information in these three files, processes this information, and generates an output file that contains the service demands for every resource in the queuing network model of Service demand is the total service time of a request of a certain workload type at a given device. The service demand does not include any time waiting to get access to the device. Waiting times are obtained by solving the model. The equations that form the basis of computation of service demands are presented in section 3.3. The Scalability Model Solver reads information about the service demand from this file and solves the queuing network model for desired performance measures. The underlying equations that form the basis for a solution are described in section 3.4.

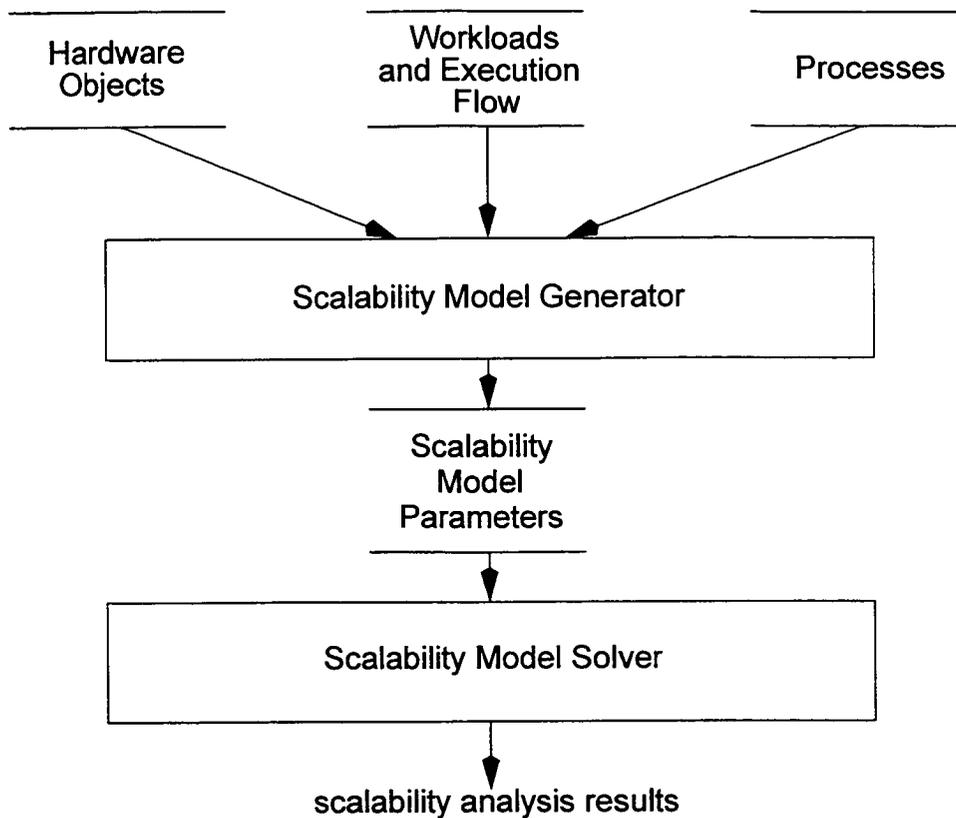


Figure 4: Scalability Model Framework

### 3.2 Parameters for the Scalability Model

The parameters used in the scalability model are:

- $P$  : set of processes.
- $NCPUs_s$  : number of processors of server  $s$ .
- $SP_{int_s}$  : SPECint95 rating of server  $s$ .
- $SP_{fp_s}$  : SPECfp95 rating of server  $s$ .
- $TypeSP_p$  : type (e.g., int or fp) of the SPEC rating used to specify the computation demand for process  $p$ .
- $SP_p$  : SPEC rating of the machine used to measure the computation demand of process  $p$ .
- $ComputeDemand_p$  : compute demand of process  $p$  measured at a machine with SPEC rating  $SP_p$ , in seconds.

- $PExec_{p,w}$  : probability that process  $p$  executes in workload  $w$ .
- $Seek_{d,s}$  : average seek time of single disk  $d$  of server  $s$ , in seconds.
- $Latency_{d,s}$  : average rotational latency of single disk  $d$  of server  $s$ , in seconds.
- $TransferRate_{d,s}$  : transfer rate of single disk  $d$  of server  $s$ , in MBytes/sec.
- $Hit_{da,s}$  : cache hit ratio for disk array  $d$ .
- $RAIDSeek_{da,s}$  : average seek time at any of the disks that compose disk array  $da$  at server  $s$ , in seconds.
- $RAIDLatency_{da,s}$  : average rotational latency at any of the disks that compose disk array  $da$  at server  $s$ , in seconds.
- $RAIDRate_{da,s}$  : transfer rate of any of the disks that compose disk array  $da$  at server  $s$ , in Mbytes/sec.
- $NTDrives_{t,s}$  : number of tape drives of tape library  $t$  at server  $s$ .
- $NRobots_{t,s}$  : number of robots of tape library  $t$  at server  $s$ .
- $Rewind_{i,t,s}$  : rewind time of tape drive  $i$  of tape library  $t$  at server  $s$ .
- $MaxTSearch_{i,t,s}$  : maximum search time of tape drive  $i$  of tape library  $t$ , in seconds at server  $s$ .
- $TapeRate_{i,t,s}$  : transfer rate of tape drive  $i$  of tape library  $t$  at server  $s$ , in Mbytes/sec.
- $Exchanges_{t,s}$  : number of tape exchanges per hour for each robot of tape library  $t$  at server  $s$ . (Each exchange involves putting the old tape in the tape library and loading the new tape into the tape drive.)
- $FilesPerMount_{p,t,s}$  : average number of files accessed per mount by process  $p$  at tape library  $t$  at server  $s$ .
- $FileSizePerMount_{p,t,s}$  : average size of files accessed by process, in Kbytes  $p$  per mount at tape library  $t$  at server  $s$ .
- $Bandwidth_n$  : bandwidth of network  $n$ , in Mbps.
- $NType_n$  : type of network  $n$ .
- $NumBlocks_{d,s,p}$  : number of blocks accessed by process  $p$  at single disk  $d$  at server  $s$ .
- $BlockSize_{d,s,p}$  : block size for each access to single disk  $d$  at server  $s$  by process  $p$ , in KBytes.

- NumBlocksRead<sub>da,s,p</sub> : number of blocks read by process  $p$  from disk array  $da$  at server  $s$ .
- NumBlocksWritten<sub>da,s,p</sub> : number of blocks written by process  $p$  to disk array  $da$  at server  $s$ .
- StripeUnitSize<sub>da,s</sub> : size of the stripe unit for disk array  $da$  at server  $s$ , in Kbytes.
- Server<sub>p</sub> : server in which process  $p$  is allocated.
- $\lambda_w$  : arrival rate of workload  $w$ , in requests/sec.
- $P_w$  : set of processes executed by workload  $w$ .
- $\pi_w = \{(p, x) \mid p \in P \text{ and } x = Pr[p \text{ is executed in workload } w]\}$  : process flow within workload  $w$ .
- PNet<sub>n,w</sub> : probability that network  $n$  is traversed by workload  $w$ .
- Volume<sub>n,w</sub> : total data volume transferred through network  $n$  by workload  $w$ , in Kbytes.

The input parameters for the Scalability Model Solver are:

- $D_{i,p,w}$  : average service demand of process  $p$  in workload  $w$  at device  $i$ , i.e., the total time spent by the process at the device for workload  $w$ . This time does not include any queuing time.
- $\lambda_w$  : average arrival rate of requests of workload  $w$  that arrive to ECS's Data Server.

### 3.3 Algorithms for Computing the Scalability Model Solver Parameters

In this section, we derive expressions for computing service demands for workloads at various types of devices. The service demand at a device due to a task is defined as the multiplication of the visit count of the task to the device and the service time of the task per visit to the device. The service demand represents the total average time spent by the task at the device.

#### 3.3.1 Computation of Service Demands for Processors

The service demand that a task in workload  $w$  presents at a server  $s$  due to the execution of a process  $p$  is given by:

$$D_{s,p,w} = \frac{\text{ComputeDemand}_p \times \text{PExec}_{p,w}}{\text{ScaleFactor}(p,s)} \quad (1)$$

where

$$\text{ScaleFactor}(p,s) = \begin{cases} \text{SPint}_s / \text{SP}_p & \text{if TypeSP}_p = \text{int} \\ \text{SPfp}_s / \text{SP}_p & \text{if TypeSP}_p = \text{fp} \end{cases} \quad (2)$$

Since  $ComputeDemand_p$  is given for a processor of certain rating,  $ScaleFactor(p, s)$  is used to normalize the process service time to the speed-rating of the current processor. The service demand,  $D_{s,w}$ , of a workload  $w$  at the CPU of server  $s$  is then

$$D_{s,w} = \sum_{\forall p \in P_w \mid s = Server_p} D_{s,p,w} \quad (3)$$

### 3.3.2 Computation of Service Demands for Single Disks

The service demand that a task in workload  $w$  presents to a disk  $d$  at a server  $s$  due to the execution of a process  $p$  is given by

$$D_{d,s,p,w} = PExec_{p,w} \times NumBlocks_{d,s,p} \times \left[ Seek_{d,s} + Latency_{d,s} + \frac{BlockSize_{d,s,p}}{TransferRate_{d,s} \times 1000} \right] \quad (4)$$

The term " $Seek_{d,s} + Latency_{d,s} + \frac{BlockSize_{d,s,p}}{TransferRate_{d,s} \times 1000}$ " denotes the time the disk takes to fetch one block of data.

The service demand,  $D_{d,s,w}$ , of a workload  $w$  at disk  $d$  of server  $s$  is then

$$D_{d,s,w} = \sum_{\forall p \in P_w \mid s = Server_p} D_{d,s,p,w} \quad (5)$$

### 3.3.3 Computation of Service Demands for Disk Arrays

The computation of service demands for disk arrays is involved and is done in several steps. The number of blocks that a process  $p$  reads at a disk (i.e., the number of stripe accesses) in disk array  $da$  at server  $s$  is given by

$$NumBlocksReadPerDisk_{da,s,p} = \left[ \frac{NumBlocksRead_{da,s,p} \times BlockSize_{d,s,p}}{5 \times StripeUnitSize_{da,s}} \right] \quad (6)$$

The numerator denotes the total volume of information read from all five disks in the disk array and the denominator denotes the volume of information read from all five disks in a single stripe group access.

The service time to process each stripe request at each disk is given by the following equation: (The first subexpression indicates that the seek time is amortized over all stripe unit accesses.)

$$\text{ServiceTimePerDisk}_{da,s,p} = \frac{\text{RAIDSeek}_{da,s}}{\text{NumBlocksReadPerDisk}_{da,s,p}} + \text{RAIDLatency}_{da,s} + \frac{\text{StripeUnitSize}_{da,s}}{\text{RAIDRate}_{da,s}} \quad (7)$$

The service demand due to read requests at a disk in disk array  $da$  at server  $s$  due to execution of process  $p$  in workload  $w$  is given by the following equation: (Since a disk array has a data cache, term  $(1 - \text{Hit}_{da,s})$  denotes the probability that data to be read is not available in the cache and a read access will have to be made.)

$$\text{ReadServiceDemandPerDisk}_{da,s,p,w} = \text{NumBlocksReadPerDisk}_{da,s,p} \times \text{ServiceTimePerDisk}_{da,s,p} \times \text{PExec}_{p,w} \times (1 - \text{Hit}_{da,s}) \quad (8)$$

Now the service demand  $D^r_{da,s,p,w}$ , due to read requests at disk array  $da$  at server  $s$  due to execution of process  $p$  in workload  $w$  is given by the following equation:

$$D^r_{da,s,p,w} = \frac{H_5 \times \text{ReadServiceDemandPerDisk}_{da,s,p,w}}{1 - \text{USingleDisk}_{da,s,p,w}} \quad (9)$$

where  $H_5 = \sum_{j=1}^5 1/j = 2.28$  and  $\text{USingleDisk}_{da,s,p,w}$  is given by Eq. (14). The term  $H_5$  shows up in the expression because a read request at the disk array is complete only after the last read at its disks is done. This approximation is based on [5].

The service demand,  $D^r_{da,s,w}$  of a workload  $w$  at the disk array  $da$  of server  $s$  is then

$$D^r_{da,s,w} = \sum_{\forall p \in P_w \mid s = \text{Server}_p} D^r_{da,s,p,w} \quad (10)$$

The computation of the service demand due to write requests at disk array  $da$  at server  $s$  due to the execution of process  $p$  in workload  $w$  is similar. The computation of the number of blocks that a process  $p$  writes at a disk (i.e., the number of stripes written) in the disk array  $da$  at server  $s$  is somewhat different and is given by the following equation: (The  $(4/5)$  term in the denominator is due to the fact that a parity block is generated for every four blocks written onto the disks. Thus 25% additional data is generated.)

$$\text{NumBlocksWrittenPerDisk}_{da,s,p} = \left\lceil \frac{\text{NumBlocksWritten}_{da,s,p} \times \text{BlockSize}_{d,s,p}}{(4/5) \times \text{StripeUnitSize}_{da,s}} \right\rceil \quad (11)$$

$$\text{WriteServiceDemandPerDisk}_{da,s,p,w} = \text{NumBlocksWrittenPerDisk}_{da,s,p} \times \text{ServiceTimePerDisk}_{da,s,p} \times \text{PExec}_{p,w} \quad (12)$$

$$D_{da,s,p,w}^w = \frac{H_5 \times \text{WriteServiceDemandPerDisk}_{da,s,p,w}}{1 - \text{USingleDisk}_{da,s,p,w}} \quad (13)$$

where

$$\text{USingleDisk}_{da,s,p,w} = \text{PExec}_{p,w} \times \lambda_w \times [(\text{NumBlocksReadPerDisk}_{da,s,p} + \text{NumBlocksWrittenPerDisk}_{da,s,p}) \times \text{ServiceTimePerDisk}_{da,s,p}] \quad (14)$$

The service demand,  $D_{da,s,w}^w$ , of a workload  $w$  at the disk array  $da$  of server  $s$  is then

$$D_{da,s,w}^w = \sum_{\forall p \in P_w \mid s = \text{Server}_p} D_{da,s,p,w}^w \quad (15)$$

### 3.3.4 Computation of Service Demands for Tape Libraries

The computation of the service demands for tape drives and robots at a tape library is involved and is done in several steps.

The total average seek time that a process  $p$  experiences at tape drive  $i$  in tape library  $t$  at server  $s$  is given by the following equation: (The factor "1/2" is due to the fact that the first file access will result in searching half the tape on the average and the factor "1/3" shows up because the remaining file accesses will require searching 1/3 of the tape on the average.)

$$\text{AverageSeekTime}_{i,t,s} = \text{MaxTSearch}_{i,t,s} \times [1/2 + (\text{FilesPerMount}_{p,t,s} - 1) / 3] \quad (16)$$

The average tape mount time in seconds at tape drive  $i$  in tape library  $t$  at server  $s$  is given by

$$\text{MountTime}_{i,t,s} = 3,600/2 \times \text{Exchanges}_{t,s} \quad (17)$$

The time that tape drive  $i$  in tape library  $t$  at server  $s$  takes to serve a file access request is given by

$$\text{TapeDriveServiceTime}_{i,t,s} = \text{AverageSeekTime}_{t,s} + \frac{\text{FilesPerMount}_{p,t,s} \times \text{FileSizePerMount}_{p,t,s}}{\text{TapeRate}_{i,t,s}} + \text{Rewind}_{i,t,s} \quad (18)$$

The average robot service time is then

$$\text{RobotServiceTime}_{t,s} = 2 \times \text{MountTime}_{i,t,s} \quad (19)$$

So, the service demand at the tape drive  $i$  of tape library  $t$  of server  $s$  due to the execution of process  $p$  in workload  $w$  is

$$D_{i,t,s,p,w}^{\text{tapedrive}} = P_{\text{Exec}_{p,w}} \times \text{TapeDriveServiceTime}_{i,t,s} / N_{\text{TDrives}_{t,s}} \quad (20)$$

The service demand at the tape drive  $i$  of tape library  $t$  of server  $s$  due to workload  $w$  is

$$D_{i,t,s,w}^{\text{tapedrive}} = \sum_{\forall p \in P_w \mid s = \text{Server}_p} D_{i,t,s,p,w}^{\text{tapedrive}} \quad (21)$$

The average service demand at any robot of tape library  $t$  of server  $s$  due to the execution of process  $p$  in workload  $w$  is

$$D_{t,s,p,w}^{\text{robot}} = P_{\text{Exec}_{p,w}} \times \text{RobotServiceTime}_{t,s} / N_{\text{Robots}_{t,s}} \quad (22)$$

The service demand at any robot of tape library  $i$  of tape library  $t$  of server  $s$  due to workload  $w$  is

$$D_{t,s,w}^{\text{robot}} = \sum_{\forall p \in P_w \mid s = \text{Server}_p} D_{t,s,p,w}^{\text{robot}} \quad (23)$$

### 3.3.5 Computation of Service Demands for Networks

The service demand of workload  $w$  presents at network  $n$  is given by the following equation: (The term "Volume $_{n,w}$ /Bandwidth $_n$ " denotes the time taken by the network to transfer the data for a task in workload  $w$ .)

$$D_{n,w}^{network} = \frac{PNet_{n,w} \times Volume_{n,w} \times 8}{Bandwidth_n \times 1000} \quad (24)$$

### 3.4 The Scalability Model

The scalability model uses quelling network (QN) models to determine the degree of contention at each of the devices that compose ECS's Data Server. The QN model used in this case is a multiclass open QN [4] with additional approximations to handle the case of disk arrays and to handle the instances of simultaneous resource possession that appear when modeling automated tape libraries [3]. The QNs used also allow for load dependent devices. Load dependent devices are used in the model to handle the following situations:

- **Symmetric multiprocessors:** this case is characterized by a single queue for multiple servers. In this case, the service rate  $\mu(k)$  of the CPU as a function of the number of requests  $k$  is given by  $k \cdot \mu$  for  $k \leq J$  and  $J \cdot \mu$  for  $k > J$  where  $J$  is the number of CPUs and  $\mu$  is the service rate of each CPU.
- **Collision-based LANs:** in this case, the throughput of the LAN decreases as the load increases due to an increase in the number of collisions. This can be modeled by using an appropriate service rate function  $\mu(k)$  as a function of the load on the network [1].

An open multiclass QN is characterized by the number  $R$  of classes, the number  $K$  of devices, by a matrix  $D = [D_{i,r}]_{i=1, \dots, K, r=1, \dots, R}$  of service demands per device per class, and by a vector  $\vec{\lambda} = (\lambda_1, \dots, \lambda_R)$  of arrival rates per class. For each device, one has to indicate its type. The following types of devices are allowed in the QN model:

- **Delay devices:** no queues are formed at these devices.
- **Queuing Load Independent (LI) devices:** queues are formed at these devices, but the service rate of the device does not depend on the number of requests queued for the device.
- **Load Dependent device (LD):** queues are formed at these devices, but the service rate of the device depends on the number of requests queued for the device. In the case of load dependent devices, one has to provide the service rate multipliers (see [4]) for each value of the number of customers. In most cases this is true for multiprocessors and collision-based LANs. The value of the service rate multipliers saturates very quickly with the number of requests. Therefore, we only need to provide a small and finite number of service rate multipliers for each LD device.
- **Disk Array:** this is a special type of device used to model disk arrays (see Figure 5 for a depiction of this type of device).

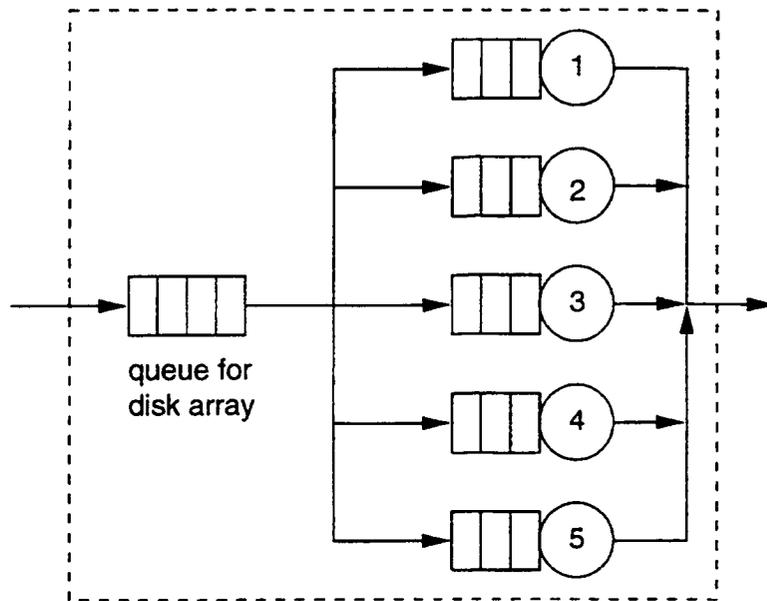


Figure 5: Disk Array

The output results of an open multiclass QN are:

- $R'_{i,r}(\vec{\lambda})$ : average residence time of class  $r$  requests at device  $i$ , i.e., the total time—including queuing and service—spent by requests of class  $r$  at device  $i$ .
- $R_r(\vec{\lambda})$ : average response time of requests of class  $r$ .  $R_r(\vec{\lambda}) = \sum_{i=1}^K R'_{i,r}(\vec{\lambda})$ .
- $U_i(\vec{\lambda})$ : utilization of device  $i$ .
- $n_{i,r}(\vec{\lambda})$ : average number of requests of class  $r$  present at device  $i$ .
- $n_i(\vec{\lambda})$ : average number of requests of at device  $i$ .  $n_i(\vec{\lambda}) = \sum_{r=1}^R n_{i,r}(\vec{\lambda})$ .

The basic equations for open multiclass QNs are (see [4]):

$$U_{i,r}(\vec{\lambda}) = \lambda_r D_{i,r}$$

$$U_i(\vec{\lambda}) = \sum_{r=1}^R U_{i,r}(\vec{\lambda})$$

$$\bar{n}_{i,r}(\vec{\lambda}) = \frac{U_{i,r}(\vec{\lambda})}{1 - U_i(\vec{\lambda})}$$

$$R'_{i,r}(\vec{\lambda}) = \begin{cases} D_{i,r} & \text{delay device} \\ \frac{D_{i,r}}{1 - U_i(\vec{\lambda})} & \text{LI device} \end{cases}$$

$$R_r(\vec{\lambda}) = \sum_{i=1}^K R'_{i,r}(\vec{\lambda})$$

$$n_i(\vec{\lambda}) = \sum_{r=1}^R n_{i,r}(\vec{\lambda})$$

The extension to LD devices is given in [4].

#### 4. Concluding Remarks

In this report, we derived the algorithms and expressions to be used to convert data describing the software and hardware architecture of ECS's Data Server into a scalability model. The model will be used to verify how well the Data Server supports an increase in workload intensity while maintaining reasonable performance. The scalability model is based on queuing network models that are automatically generated from the description of the architecture and the workload.

#### References

- [1] Lazowska, E. D., Zahorjan, J., Graham, G. S., and Sevcik, K. C. (1984). *Quantitative System Performance: Computer System Analysis Using Queuing Network Models*. Prentice Hall. Englewood Cliffs, N.J.
- [2] Kobler, B., Berbert, J., Caulk, P., and Hariharan, P. C. (1995). Architecture and Design Storage and Data Management for the NASA Earth Observing System Data and Information System (EOSDIS). *Fourteenth IEEE Symposium on Mass Storage Systems (Second International Symposium)*. Monterey, CA.
- [3] Menascé, D. A., Pentakalos, O. I. and Yesha, Y. (1996). An Analytic Model of Hierarchical Mass Storage Systems with Network-Attached Storage Devices. *Proceedings of the 1996 ACM Sigmetrics Conference*. Philadelphia, PA.
- [4] Menascé, D. A., Almeida, V. A. F., and Dowdy, L. W. (1994). *Capacity Planning and Performance Modeling: from Mainframes to Client-server Systems*. Prentice Hall. Englewood Cliffs, N.J.
- [5] Pentakalos, O. I., Menascé, D. A., Halem, M., and Yesha, Y. Analytical Performance Modeling of Hierarchical Mass Storage Systems. *IEEE Transactions on Computers*, 46(10), 1103-1118.
- [6] Pentakalos, O. I., Menascé, D. A., and Yesha, Y. (1997) Pythia and Pythia/WK: Tools for the Performance Analysis of Mass Storage Systems. *Software Practice and Experience*.

[7] Pentakalos, O. I. and Menascé, D. A. (1996) Automated Clustering Based Workload Characterization for Mass Storage Systems. *Fifth NASA Goddard Space Flight Center Conference on Mass Storage Systems and Technologies*. College Park, MD.

[8] Pentakalos, O. I., Menascé, D. A., Yesha, Y., and Halem, M. (1995) An Approximate Performance Model of a Unitree Mass Storage System. *Fourteenth IEEE Symposium on Mass Storage Systems (Second International Symposium)*. Monterey, CA.

## ***Numeric Simulation of a Volcanic Jet-Plume***

**Santiago Egido Arteaga**  
**University of Maryland College Park**  
**(arteaga@umbc.edu)**

The purpose of this research has been to simulate numerically on parallel computers the volcanic plume produced during a Plinian eruption. In order to decrease the computational effort required, we assume radial symmetry of the plume.

We have developed a model for this phenomenon consisting of the Navier-Stokes equations, the equation of state of ideal gases, energy conservation, convection-diffusion of the chemicals which appear in the simulation, the hydrostatic equilibrium condition for the initial atmosphere and buoyancy terms, and the Rankine-Hugoniot equations for shockwaves.

For the numerical simulation, we first discretized the problem using conventional finite difference methods; in particular, we implemented Lax-Friedrich and Lax-Wendroff discretizations. We spent some time eliminating some instabilities associated with inaccurate boundary conditions. But in the end these methods were determined to be inappropriate, because they allowed too much and too early mixing of air and volcano jet. The resulting initial flow obtained was an explosion caused by the heating of the atmosphere.

To prevent this we needed a boundary-free method, separating the air from the jet. Since these formulations have complicated parallel implementations and can require too much time, we developed our own implicit boundary-free method, which does not keep track of exactly where the boundary air-jet interface is, but rather of how much air and jet there is in each discretization cell. This model prevents the explosive behavior observed with the previous discretizations and improved greatly the numerical results.

Additionally, various amounts of time were invested in learning gas dynamics, in particular shock waves, and familiarizing ourselves with the J90 and Cray T3D computers, the NCAR graphics package, and the MPI communications library.

We have also spent some time optimizing the programs for execution on the Cray T3E computer, leading to improvements in execution speed of some routines by a factor of ten, and multiplying the performance of the whole program by a factor of four (further improvements are possible). The parallel efficiency of our implementation is very good; we routinely use 99.5% of the conceded processor time, and in some test runs we have achieved 99.99% efficiency. We have also written programs to graphically display the numerical results.

## ***Volumetric Display of Earth and Space Science Data***

**David Ebert**  
**University of Maryland Baltimore County**  
**([ebert@umbc.edu](mailto:ebert@umbc.edu))**

My work on this contract was to provide expertise in volumetric display software to help design a new software architecture for a revolutionary new display device. My work centered on the following topics:

- Theoretical design of software architecture.
- Exploration of using traditional graphics hardware to aid in driving this new display technology.
- Proposal for future development directions.

Below I have summarized my activities for these projects.

### **1. Theoretical Software Architecture**

After exploring many possible approaches to the development and design of software for this new architecture, I have conclude that the software development should take place at three levels: hardware specific driver, hardware translation layer, device-independent layer. The separation into these layers will allow the development of software capable of driving the prototype display, but also allow easy modification at the lower levels for different architectures as the development continues.

### **2. Use of Traditional Graphics Hardware**

Creating images in a 3D glass cube is a challenging problem with great computational demands. I have analyzed the current state of the art of hardware graphics engines to determine their suitability to aid in this process. These architectures can be utilized in several fashions:

1. Multiple passes through the Z-buffer for each depth of 3D display. This would allow traditional polygons, lines, and characters to be displayed in the 3D glass cube. The performance issue here is the speed of Z-buffer operations for the graphics engine.
2. Use of the hardware frame-buffer for storing a 3D volume to be rendered in the cube. This will not have a great performance increase over main-memory storage.
3. Use of texture-memory hardware to store the 3D volume. This allows the use of hardware translation tables to perform very quick operations on the volume before display in the 3D glass cube.

### **3. Proposal for Future Development Directions**

I propose a two-path development of software for this architecture. The first path is to develop a simple display driver for the device and display actual NASA data in the cube (e.g., vector flow field from a magneto-hydrodynamics simulation by D. Aaron Roberts). This will show the capabilities of the machine and also allow exploration of and experimentation with techniques to best utilize the display technology. Many issues can be resolved through this exploration and development. The second path is the 3-level software architecture that will create a flexible, extensible software architecture for this new device.

## **Putting Log Data to Work: Mass Storage Performance Information System**

**Lisa Singh**  
**Northwestern University**  
**Department of Electrical and Computer Engineering**  
**(lisa-singh@nwu.edu)**

### **Problem Statement**

The Science Computing Branch (SCB) manages the world's most active openly accessible mass storage system, consisting of over 45 terabytes of user data and supporting over 1000 users. The system consists of eight robotic silos and computing platforms from multiple vendors running various operating systems. Currently detailed logs capture the activity of the mass data storage and delivery system (MDSDS) internal to the different systems. However, because these logs are large, ill-structured, and inaccessible for querying, the overall performance of the MDSDS, as well as statistics about user access patterns, cannot be determined.

### **Project Goals**

The goal of this project is to design and implement a system that evaluates the performance of the SCB massive storage system. Performance results can include the ability to submit ad hoc queries on a data warehouse containing log data, the creation of reports that identify summary information, and finally, the search for hidden patterns that exist within the data itself, i.e., data mining.

### **1. High Level Tasks**

My role on this project is to assist in the development of a data warehouse and a data mining tool for the log data. The high level tasks necessary to meet these goals are as follows:

#### **1. *Data cleansing***

Ensuring that all the data is in one consistent format that still accurately represents the original dataset. Some specific data cleansing issues include dealing with missing data values, removing redundant data, and using a consistent format for every attribute data value.

#### **2. *Data warehousing***

The accumulation of large amounts of heterogeneous and distributed data into a single data repository. Typically, this data cannot be accessed by system users. Instead its purpose is to handle large, data intensive queries.

#### **3. *Query and report generation***

Develop an application to query the data warehouse and produce reports with summary statistics.

#### **4. *Data transformation***

Converting the data in the data repository into a format that a data mining tool can utilize. Typical data transformation issues include determining useful background knowledge, using dimensionality transformation to reduce the effective number of variables under consideration, and projecting data onto easily solvable solution spaces.

## 5. Data mining

Designing tools that extract global patterns from transformed data and analyze potential relationships (associations) within these extracted patterns. The information attained in this process is inductive and may be used to identify user clusters or system performance groupings, suggest potentially interesting correlations based on current access patterns, or predict future trends.

## 2. Accomplishments

We are currently at step 3 in the high level process. The remainder of this report will go through a detailed description of my individual tasks and accomplishments associated with tasks 1 and 2.

Other team members determined the schema or design of the initial prototype database. This design was then translated into a set of database tables. Once all the tables were designed, I set up the tables in our prototype database. For the initial prototype, we are using a Sybase relational database.

Possibly the most tedious task of this process is extracting useful data from the log files. There are four types of log files in our initial prototype. For each file, I wrote a separate extraction program. Each of these programs is written in PERL. The programs take log files as input and output data files that can be directly loaded into a database.

At this stage the data was ready to be loaded into the Sybase database. I wrote scripts that populated all the log related data tables. I also created meta-data tables and administrative tables, where meta-data is defined as data about data. Examples of meta-data tables include attribute naming conventions and foreign key constraints. Because these tables are rarely updated and contain only a small amount of data, I populated them manually. Eventually, this will also need to be automated.

Once the relational database design was complete, I began working on the data warehouse model. For performance issues, we chose to transform our standard relational schema to a star schema. I helped create the initial star schema model. I then began working on setting up the data warehouse. This setup involves creating tables and loading data into these new sets of data warehouse tables.

## 3. Future Plans

Now that the data models, extraction scripts and loading scripts have been completed, it is time to determine the amount of disk space that will be used by the new data warehouse. Therefore, given a particular log, we need to determine the number of bytes the database records consume. Logs are much larger than the database records associated with each log entry. Therefore, this analysis boils down to determining the compression ratio, i.e., given a log of a particular size, what is the likely size of the data file. I determined some initial estimates for this. I am now beginning a more formal disk space analysis study.

The next major task will be producing a comparative analysis of the performance of different prototype systems for our warehouse. I will be comparing an Oracle 8 and a Redbrick data warehouse. The performance analysis should help us determine which system is better for our data set. Once this is accomplished, the final data warehouse can be populated.

---

---

# **ADMINISTRATION BRANCH**

---

**Nancy Campbell**, Senior Administrator – Branch Head

**Georgia Flanagan**, Administrative Assistant 3 (Conference Management)

**L'Tanya Pierce**, Administrative Assistant 3 (Financial)

**Michele Meyett**, Administrative Assistant 2 (Web Site Administration, Database Management, Presentation Graphics, Desktop Publishing)

**Dawn Segura**, Administrative Assistant 2 (Financial/Subcontract Support)

This branch is responsible for supporting the CESDIS Director, Senior and Staff Scientists, Technical Specialists, funded project personnel and graduate students, and USRA's corporate office. Branch personnel:

- Serve as the liaison among funded research personnel, NASA scientific and administrative personnel, and USRA accounting and procurement personnel,
- Monitor subcontracts and consulting agreements,
- Monitor the contract's Small and Small/Disadvantaged Business Plan,
- Prepare and monitor task budgets,
- Prepare contract reports,
- Obtain Contracting Officer permission for foreign travel by staff and university scientists,
- Obtain Contracting Officer permission for equipment purchases with contract funds and report purchases to Goddard's property personnel,
- Assist with conference planning and provide on-site support at conference, workshop, and seminar locations,
- Assist foreign national visitors in gaining access to Goddard,
- Provide peer review support to NASA program personnel for proposals submitted in response to NASA Research Announcements and Cooperative Agreement Notices,
- Maintain CESDIS Web site,
- Provide desktop publishing assistance for paper preparation, the CESDIS newsletter, and presentation graphics,
- Make travel arrangements and provide assistance with travel voucher completion,
- Perform functions of remote site data entry for USRA's centralized accounting system including payroll, purchasing, and accounts payable.

## BRANCH ACTIVITIES

### Seminar Series

CESDIS sponsors seminars by visiting scientists from universities, government laboratories, and the public sector. These presentations are open to everyone at Goddard as well as interested off-site attendees. Announcements of speakers and dates are posted on the CESDIS Website. Seminar presentations during this reporting year are listed below. Abstracts appear in Appendix C.

- William Arms. Corporation for National Research Initiatives. *Interoperability Research in Digital Libraries*.
- Bharat Bhargava. Purdue University. *Large Scale Distributed Database Systems: Experiments and Observations*.
- Krzysztof Gorski. Theoretical Astrophysics Center (Copenhagen, Denmark). *High Resolution Mapping on the Sphere for Space and Earth Applications*.
- David Harel. Weizmann Institute of Science (Rehovot, Israel). A 3-part series: (1) *Some Thoughts on Statecharts, 13 Years Later*, (2) *Computers Are Not Omnipotent*, (3) *On the Aesthetics of Diagrams*.
- Benjamin Kedem. University of Maryland College Park. *Bayesian Spatial Prediction in Skewed Random Fields*.
- Zvi Kedem. Courant Institute of Mathematical Sciences, New York University. *MILAN: Prototyping a New Methodology for Reliable Parallel Processing on Distributed Environments*.
- Hao Le. Flashback Imaging, Inc. *Volumetric Imaging Model*.
- Jorge Pinzon. University of California, Davis. *Spatial and Spectral Feature Extraction*.
- Lisa Singh. Northwestern University. *Mining Semi-structured Data Using a Concept Library*.
- Jennifer Trelewicz. Arizona State University. *Transforms for Digital Holographic Data Storage, a Progress Report*.
- Victor Vianu. University of California, San Diego. *Active Databases for Electronic Commerce*.
- Ouri Wolfson. University of Illinois, Chicago. *Location Management in Moving Objects Databases*.

### Workshop on Data Mining, Warehousing, and Large Data Recovery

This by-invitation-only workshop was held at Goddard on August 19-21, 1997 for the purpose of identifying areas within NASA to which data mining and warehousing technology could be applied. As stated in the workshop announcement, the underlying premise was as follows: NASA continues to collect increasing amounts of Earth and space science data. Providing distributed access to archived data is only the first step. Developing data warehouses containing a variety of data sets in a form that facilitates further scientific analysis and automatically mining the warehoused data for new insights and discoveries is necessary for the data NASA collects to be efficiently used and exploited.

Presentations were made by the following individuals:

- Chaitanya Baru. San Diego Supercomputer Center. *Warehousing Scientific and Very Large Data Sets*.
- Robert Grossman. Magnify, Inc. and the University of Illinois, Chicago. *An Introduction to Data Mining*.
- Jiawei Han. Simon Fraser University (Canada). *OLAP Mining: an Integration of OLAP with Data Mining*.
- Alberto Mendelzon. University of Toronto. *Commercial Products: State of the Art*.
- Ramakrishnan Srikant. IBM Almaden Research Center. *Data Mining*.
- Jennifer Widom. Stanford University. *Data Warehousing: Overview and Research Achievements*.

A more complete discussion of data mining and warehousing with selected references appears in the workshop materials comprising Appendix A.

## **Image Registration Workshop**

This 2-day workshop was organized by CESDIS Senior Scientist Jacqueline Le Moigne and was sponsored by CESDIS, the GSFC Applied Information Sciences Branch of the Earth and Space Data Computing Division, and the Washington/Northern Virginia Chapter of the IEEE Geoscience and Remote Sensing Society. Its purpose was to explore promising approaches to image registration for various domains of applications such as medical, military, and/or space imagery.

A complete list of speakers and the titles of their presentations may be found in Appendix B. A copy of the workshop proceedings may be obtained by contacting the CESDIS administrative office at 301-286-4403.

## **CESDIS Science Council**

The CESDIS Science Council met on August 12, 1997 at Goddard. Presentations on work-in-progress were given by Yelena Yesha, Harold Stone who spoke of his collaborative work with Jacqueline Le Moigne who could not be present, Don Becker, Richard Lyon, Nathan Netanyahu, Susan Hoban, and Kostas Kalpakis. A portion of the afternoon was devoted to an open discussion of the future of CESDIS by interested participants since the second 5-year contract was due to expire on July 5, 1998. (Ultimately the existing contract was extended for two years through July 5, 2000.)

The next regularly scheduled meeting of the CESDIS Science Council will be in the Fall of 1998.

---

---

# ***APPENDIX A***

---

## **Workshop on Data Mining, Warehousing, and Large Data Recovery**

**August 19 - 21, 1997**

**NASA-CESDIS Workshop on  
Data Mining, Warehousing and Large Data Recovery  
NASA Goddard Space Flight Center  
August 19-21, 1997**

## **Workshop Goals**

In this workshop we will first review data mining and data warehousing technology, the current state of the art both from a research and a commercial product point of view, and the research challenges posed by it. We will then focus on specific challenges relevant to NASA, such as mining and warehousing very large data sets and scientific data.

## **Data Mining**

Data mining is the automatic discovery patterns, associations, changes, anomalies and significant structures in large data sets. Data intensive computing is concerned with statistically and numerically intensive queries on large data sets. Traditional data analysis is assumption driven in the sense that a hypothesis is formed and validated against the data. Data mining in contrast is discovery driven in the sense that patterns are automatically extracted from data.

Data mining and data intensive computing are emerging as key enabling technologies for a variety of scientific, engineering and business problems. For data mining and data intensive computing, data management issues must be balanced against numerical issues and the input/output bandwidth of the system must be balanced against the processing power of the system.

## **Data Warehousing**

A data warehouse is a "subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organization decision making." (W. H. Inmon, *Building the Data Warehouse*, John Wiley, 1996). Or perhaps "a single, complete, and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use in a business context" (B. Devlin, *Data Warehouse: from Architecture to Implementation*, Addison-Wesley, 1997). Whichever definition we adopt, the database industry and research community have been focusing more and more on this topic over the past few years.

As the traditional problems of on-line transaction processing (OLTP) become well understood, and the technology matures, there has been a shift of attention to on-line analytic processing (OLAP): how to turn the masses of operational data that organizations accumulate into information that can be exploited to make better decisions. Data warehousing aims to consolidate and integrate data in a form that analysts can explore and manipulate easily and efficiently.

## **Opportunity**

NASA continues to collect ever increasing amounts of Earth and space science data. Providing distributed access to archived collected data is only the first step. Developing data warehouses containing a variety of data sets in a form that facilitates further scientific analysis and automatically mining the warehoused data for new insights and discoveries is necessary for the data NASA collects to be efficiently used and exploited.

## Timeliness

Recently there has been an increased focus on data mining and data warehousing by researchers. Some of the key technologies have already been incorporated into commercial products. In addition, several testbeds for high performance data warehousing and data mining are being developed. Currently, it is very timely for NASA to expose the research and commercial communities to its specific problems and challenges in these areas.

**NASA-CESDIS Workshop on  
Data Mining, Warehousing and Large Data Recovery  
NASA Goddard Space Flight Center  
August 19-21, 1997  
Bldg. 28, Room E210**

### Agenda

**Tuesday, August 19  
Data Mining  
Chair: Dr. Robert Grossman**

- |                 |                                                                                                                               |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------|
| 9:00 - 9:15 am  | Welcome<br>Milton Halem, NASA Goddard Space Flight Center and<br>Yelena Yesha, CESDIS/University of Maryland Baltimore County |
| 9:15 - 11:30 am | Introduction to Data Mining,<br>Robert Grossman, Magnify, Inc. and University of Illinois at Chicago                          |
| 1:00 - 3:00 pm  | A Survey of Data Mining Algorithms and Techniques<br>Rakesh Agrawal, IBM Almaden Research Center                              |
| 3:30 - 5:30 pm  | A Database Perspective on Data Mining<br>Jiawei Han, Simon Fraser University                                                  |

**Wednesday, August 20  
Data Warehousing  
Chair: Dr. Alberto Mendelzon**

- |                  |                                                                                                 |
|------------------|-------------------------------------------------------------------------------------------------|
| 9:00 - 11:00 am  | Data Warehousing: Overview and Research Achievements<br>Jennifer Widom, Stanford University     |
| 11:00 - 12:00 pm | Commercial products: State of the Art<br>Alberto Mendelzon, University of Toronto               |
| 1:00 - 3:00 pm   | Warehousing Scientific and Very Large Data Sets<br>Chaitan Baru, San Diego Supercomputer Center |
| 3:00 - 5:00 pm   | Research Challenges,<br>all - discussion                                                        |

**Thursday, August 21**  
**Discussion**  
**Chair: Dr. Milton Halem**

9:00 - 11:00 am      Challenges and Opportunities for NASA in Data Mining and Data Warehousing  
Rakesh Agrawal, Chaitan Baru, Robert Grossman, Jiawei Han,  
Alberto Mendelzon, Yelena Yesha, and Jennifer Widom

**Data Mining Speakers**

Dr. Rakesh Agrawal  
IBM Almaden Research Center  
ragrawal@almaden.ibm.com  
<http://www.almaden.ibm.com/cs/quest/publications.html>  
+1-408-927-1734

Dr. Robert Grossman  
Magnify, Inc.  
and University of Illinois at Chicago  
rlg@magnify.com  
<http://www.magnify.com> and <http://www.lac.uic.edu>  
+1 312 214 4120

Dr. Jiawei Han  
Simon Fraser University  
han@cs.sfu.ca  
<http://fas.sfu.ca/cs/people/Faculty/Han/>  
<http://fas.sfu.ca/cs/research/groups/DB/sections/publication/kdd/kdd.html>  
+1-604-291-4411

**Data Warehousing Speakers**

Dr. Alberto Mendelzon  
University of Toronto, coordinator

Dr. Jennifer Widom  
Stanford University

Dr. Chaitan Baru  
San Diego Supercomputer Center

**Data Mining References**

M.-S. Chen, J. Han, P.S. Yu. (1997) Data Mining: An Overview from Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*.  
<http://fas.sfu.ca/cs/research/groups/DB/sections/publication/kdd/kdd.html>

K. Koperski, J. Adhikary, J. Han. (1996, June). Spatial Data Mining: Progress and Challenges. Paper presented at SIGMOD'96 AND Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'96), Montreal, Canada. <http://fas.sfu.ca/cs/research/groups/DB/sections/publication/kdd/kdd.html>

R. Agrawal, J.C. Shafer. (1996, January). Parallel Mining of Association Rules: Design, Implementation and Experience. (IBM Research Report RJ 10004). To appear in *IEEE Transactions on Knowledge and Data Engineering*. <http://www.almaden.ibm.com/cs/quest/publications.html>

R. Agrawal, A. Arning, T. Bollinger, M. Mehta, J. Shafer, R. Srikant. (1996, August). The Quest Data Mining System. *Proceedings of the 2nd Int'l Conference on Knowledge Discovery in Databases and Data Mining*, Portland, Oregon. <http://www.almaden.ibm.com/cs/quest/publications.html>

R. Grossman. (1996). The Terabyte Challenge: An Open, Distributed Testbed for Managing and Mining Massive Data Sets. *Proceedings of 1996 IEEE-ACM Conference on Supercomputing*. IEEE Computer Society Press, 1996. See also <http://www.lac.uic.edu/hpcc-grossman.html>.

S. Bailey, R. L. Grossman. (1997). Dynamic Similarity: Mining Collections of Trajectories. *Proceedings of the 1997 Workshop on Managing and Mining Massive Data (M3D 1997)*, to appear. See also [http://www.magnify.com/white\\_papers.html](http://www.magnify.com/white_papers.html)

Fayyad, Haussler, Stolorz. (1996). KDD for Science Data Analysis: Issues and Examples. 2nd Int'l Conf. on Knowledge Discovery and Data Mining (KDD'96).

## Data Warehousing References

### Overview

Chang, Moon, Acharya, A., Shock, Sussman, A., and Saltz, J. (1997) Titan: A High-Performance Remote-sensing Database. Int'l Conf. on Data Engineering '97.

Byard, J. Schneider, D. (1996). The Ins and Outs (and everything in between) of Data Warehousing. ACM SIGMOD 1996 Tutorial Notes. Available in <http://www.redbrick.com/rbs-g/whitepapers/sigmod96.pdf>

Chaudhuri, S., Dayal, U. (1997, March). An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record 26 (1). [http://bunny.cs.uiuc.edu/sigmod/sigmod\\_record/9703/chaudhuri.ps](http://bunny.cs.uiuc.edu/sigmod/sigmod_record/9703/chaudhuri.ps)

Widom, J. (1995). Research Problems in Data Warehousing. Int'l Conference for Information and Knowledge Management '95. <ftp://db-stanford.edu/pub/papers/warehouse-research.ps>

Zhugue, Garcia-Molina, Wiener. (1996). The Strobe Algorithms for Multi-Source Warehouse Consistency. PDIS 1996. <http://www-db.stanford.edu/pub/papers/strobe.ps>

### OLAP

Gray, J., Bosworth, A., Layman, A., Pirahesh, H. (1996). Data cube: a relational aggregation operator generalizing group-by, cross-tabs and subtotals. Int'l Conf. on Data Engineering '96.

Harinarayan, V., Rajaraman, A., Ullman, J.D. (1996) Implementing Data Cubes Efficiently. ACM SIGMOD '96 (best paper award). <http://www-db.stanford.edu/pub/papers/cube.ps>

---

---

# ***APPENDIX B***

---

## **Image Registration Workshop**

**November 20-21, 1997**

# Image Registration Workshop

November 20 - 21, 1997

NASA Goddard Space Flight Center  
Greenbelt, MD, USA

## General Chair

Jacqueline LeMoigne  
USRA/CESDIS  
NASA/GSFC - Code 930.5  
Greenbelt, MD 20771  
301-286-8723  
301-286-1777 (fax)

## Technical Program

Rama Chellappa, UMCP  
Samir Chettri, GST  
Robert Crompt, NASA/GSFC  
Tarek El-Ghazawi, GWU  
Nazmi El-Saleous, UMCP  
Emre Kaymaz, KTT  
Bao-Ting Lerner, KTT  
B.S. Manjunath, UCSB  
Manohar Mareboyana, BSU  
David Mount, UMD  
Nathan Netanyahu, UMCP  
John Pierce, KTT  
Srinivasan Raghavan, HSTX  
Aya Soffer, UMBC  
Harold Stone, NEC  
James Tilton, NASA/GSFC  
Eric Vermote, UMCP  
Wei Xia-Serafino, HSTX

## Workshop Coordinator

Georgia Flanagan  
USRA/CESDIS  
NASA/GSFC - Code 930.5  
Greenbelt, MD 20771  
301-286-2080  
301-286-1777 (fax)  
georgia@cesdis.usra.edu

## Call For Abstracts

NASA Goddard Space Flight Center, USRA's Center of Excellence in Space Data and Information Sciences (CESDIS), and the Washington/Northern Virginia Chapter of the IEEE Geoscience and Remote Sensing Society are pleased to announce the first workshop in image registration. This workshop will explore promising approaches to image registration for various domains of applications, such as medical, military, or space imagery.

## Scope of Workshop

1. General Techniques and Algorithms for Image Registration
2. Applications
  - Aerial Imagery
  - Satellite Image Geo-Registration
  - Medical Image Registration
3. Evaluation Metrics: Accuracy, Computational Requirements, Applicability, Autonomy, ...

## Submission Requirements

Prospective authors are invited to propose papers in any of the technical areas listed above. To submit a proposal:

- send two copies of a 500- word abstract and a cover cover sheet stating paper title, technical area(s), contact author's name, address, telephone and fax numbers, and electronic address to:

**Jacqueline LeMoigne**  
**USRA/CESDIS, Code 930.5**  
**NASA Goddard Space Flight Center**  
**Greenbelt, MD 20771**  
**lemoine@cesdis.gsfc.nasa.gov**

Workshop proceedings will be in the form of a CESDIS/NASA Conference Publication.  
*The best four papers will be published in a Special Issue of Pattern Recognition on image registration.*

## Important Dates

Paper abstracts due:	June 20, 1997
Notification:	July 21, 1997
Camera Ready:	September 19, 1997
Workshop:	November 20-21, 1997

URL: <http://cesdis.gsfc.nasa.gov/IRW/>

# Image Registration Workshop

## Call for Participation

---

---

NASA Goddard Space Flight Center  
Greenbelt, Maryland, USA  
November 20 - 21, 1997  
<http://cesdis.gsfc.nasa.gov/IRW>

---

---

## Agenda

Sponsored by: USRA/CESDIS; NASA Goddard Space Flight Center's Applied Information Sciences Branch and Earth & Space Data Computing Division; and the WASH/NOVA Chapter of the IEEE Geoscience & Remote Sensing Society.

## November 20, 1997

- 8:30-8:45            Opening Remarks, Jacqueline Le Moigne, USRA/CESDIS
- 8:45-9:30            Invited Talk - Lisa Brown, IBM T. J. Watson Research Center  
A Survey of Image Registration Techniques
- 9:30-10:10          Session I - General Methods  
Session Chairs: David Mount and Samir Chettri

*Image Registration by Non-Linear Wavelet Compression and Singular Value Decomposition.* J. Pinzon, S. Ustin, C. Castaneda, University of California, Davis; J. Pierce, K-T Tech, Inc.  
*An Eigenspace Approach to Multiple Image Registration.* H. Schweitzer, University of Texas at Dallas.

- 10:10-10:30          BREAK
- 10:30-12:10          Session II - General Methods and Resampling  
Session Chairs: James C. Tilton and Nathan Netanyahu

*Automatic Registration of Satellite Imagery.* L. Fonseca, Instituto Nacional de Pesquisas Espaciais, BRAZIL; B. S. Manjunath, C. Kenney, UCSB.  
*Scope and Applications of Translation Invariant Wavelets to Image Registration.* S. Chettri, GST; W. Campbell, NASA GSFC; J. Le Moigne, USRA/CESDIS.  
*A Scale Space Feature Based Registration Technique for Fusion of Satellite Imagery.* S. Raghavan, Hughes STX; R. Cromp, W. Campbell, NASA GSFC.  
*An Optical Systems Analysis Approach to Image Resampling.* R. Lyon, UMBC/CESDIS.  
*Generalized Cubic Convolution: A Technique for Restoring and Resampling Images with Non-Uniform-Sampling.* S. Reichenbach, R. Narayanan, University of Nebraska, Lincoln; J. Barker, NASA GSFC; D. Kaiser, Doane College.

- 12:10-1:30            LUNCH
- 1:30-2:15            Invited Talk - William J. Campbell, NASA Goddard Space Flight Center  
Remotely Sensed Image Geo-Registration  
Session Chair: Robert F. Cromp
- 2:15-3:15            Session III - Applications to Satellite Sensors  
Session Chairs: Nazmi El-Saleous and Eric Vermote

*Automated Navigation Assessment for Earth Survey Sensors Using Island Targets.* F. Patt, R. Woodward, GSC/SAIC; W. Gregg, NASA GSFC.  
*MODIS Land Ground Control Point Matching Algorithm.* R. Wolfe, M. Nishihama, D. Solomon, Hughes STX.  
*Algorithm Cooperation for the Automatic Registration of Satellite Images.* I. Dowman, R. Ruskone, University College of London, GREAT BRITAIN.

- 3:15-3:30            BREAK
- 3:30-5:10            Session IV - Correlation Methods  
Session Chairs: Harold Stone and Aya Soffer

- Automated and Robust Image Geometry Measurement Techniques with Application to Meteorological Satellite Imaging.* J. Carr, CARR Astronautics Corp. ; M. Mangolini, B. Pourcelot, Aerospatiale, FRANCE.
- Techniques for Multi-resolution Image Registration in the Presence of Occlusions.* M. McGuire, H. Stone, NEC Research Institute.
- Comparison of Registration Techniques for GOES Visible Imagery Data.* J. Tilton, NASA GSFC.
- Iterative Edge- and Wavelet-Based Image Registration of AVHRR and GOES Satellite Imagery.* J. Le Moigne, USRA/CESDIS; N. El-Saleous, E. Vermote, UMCP.

## **POSTER SESSION / RECEPTION**

**5:30 to 7:00 - Building 28 / Atrium**

- Alignment of Functional and Anatomical Tomograms Based on Automated and Real-Time Interactive Procedures.* U. Pietrzyk, A. Thiel, H. Lucht, A. Schuster, Max-Planck Institute for Neurological Research, GERMANY.
- Aerial Image Registration by PFANN (Point Feature and Artificial Neural Network) Matching.* J. Li, Z. Qain, Y. Zhao, Image Processing and Pattern Recognition Institute, Shanghai Jiao-Tong University, CHINA.
- Fusing Stereo Images for Photogrammetric Analysis: A Guided Approach.* G. Moore, University of Ulster at Coleraine, NORTHERN IRELAND.
- Clinical Relevance of Fully Automated Multimodality Image Registration by Maximization of Mutual Information.* F. Maes, D. Vandermeulen, G. Marchal, P. Suetens, Katholieke Universiteit Leuren, BELGIUM.
- Tracking Hurricane Paths.* N. Prabhakaran, N. Rishe, R. Athauda, Florida International University.
- Image Registration by Parts.* T. El-Ghazawi, P. Charlemwat, George Washington University; J. Le Moigne, USRA/CESDIS.
- A Robust Generalized Registration Technique for Multi-Sensor and Warped Images.* S. Mitra, M. Dickens, M. Parten, E. O'Hair, Texas Tech University.
- Assessment of Neurological Function through the Multidimensional Integration of Invasive and Non-Invasive Modalities or Sensors.* L. Bidaut, Laboratory for Functional and Multidimensional Imaging - DIM, SWITZERLAND.
- Registration of Video Sequences from Multiple Sensors.* R. Sharma, M. Pavel, Oregon Graduate Institute.
- Towards an Intercomparison of Automated Registration Algorithms for Multiple Source Remote Sensing Data.* J. Le Moigne, USRA/CESDIS, et al.
- An Efficient Registration and Recognition Algorithm via Sieve Processes.* J. Phillips, U. S. Army Research Lab; J. Huang, S. Dunn, Rutgers University
- 3D Object to 2D Image Invariance Algorithms for Image Registration.* R. Williams, U. S. Navy.

## **November 21, 1997**

- 8:30-9:15            Invited Talk - Murray Loew, George Washington University  
                           Issues in Multimodality Medical Image Registration  
                           Session Chair: Tarek El-Ghazawi

9:15-10:15            Session V - Medical Image Registration  
                          Session Chairs: Bao Lerner and Manohar Mareboyana

*On Matching Brain Volumes.* J. Gee, University of Pennsylvania.

*Automated Construction of Large-Scale Electron Micrograph Mosaics.* R. Vogt, J. Trenkle, L. Harmon, ERIM International.

*Two Stage Registration for Automatic Subtraction of Intraoral in-vivo Radiographs.* T. Lehmann, K. Spitzer, W. Oberschelp, Aachen University of Technology, GERMANY.

*Regional Registration of Texture Images with Application to Mammogram Followup.* D. Brzakovic, N. Vujovic, Lehigh University.

10:15-10:30            BREAK

10:30-12:10            Session VI - General Methods with Application to Medical Imagery  
                          Session Chairs: Chandra Shekhar and Wei Xia

*A Consistent Feature Selector Based on Steerable Filters.* M. Sallam, K. Chang, K. Bowyer, University of South Florida.

*Surface Based Matching Using Elastic Transformations.* O. Tretiak, M. Gabrani, Drexel University.

*Fast Multimodality Image Registration Using Multiresolution Gradient-based Maximization of Mutual Information.* F. Maes, D. Vandermeulen, G. Marchal, P. Suetens, Laboratory for Medical Imaging (ESAT/Radiologie), Katholieke Universiteit Leuren, BELGIUM.

*Anomaly Detection through Registration.* M. Chen, H. Rowley, D. Pomerleau, T. Kanade, Carnegie Mellon University.

12:10-1:30            LUNCH

1:30-3:10             Session VII - Theory and General Methods  
                          Session Chairs: B. S. Manjunath and John Pierce

*Registration of Deformed Images.* A. Goshtasby, Wright State University.

*Correspondence-less Image Alignment using a Geometric Framework.* V. Govindu, C. Shekhar, R. Chelappa, UMCP.

*Finding Corner Point Correspondence from Wavelet Decomposition of Image Data.* M. Mareboyana, Bowie State University; J. Le Moigne, USRA/CESDIS.

*An Efficient Algorithm for Robust Feature Matching.* D. Mount, N. Netanyahu, UMCP; J. Le Moigne, USRA/CESDIS.

*Effects of Lossy Compression on Digital Image Registration.* A. Maeder, University of Ballarat, AUSTRALIA.

3:10-3:30            BREAK

3:30-4:30             Session VIII - Computer Vision  
                          Session Chairs: Srinu Ragahavan & Emre Kaymaz

*Registration of Uncertain Geometric Features: Estimating the Transformation and its Accuracy.* X. Pennec, MIT Artificial Intelligence Lab.

*Optical Flow Estimation Using Wavelet Motion Model.* Y. Wu, T. Kanade, Carnegie Mellon University; J. Cohn, C. Li, University of Pittsburgh.

*Recovery of Motion Parameters from Distortions in Scanned Images.* J. Mulligan, NASA Ames Research Center.

---

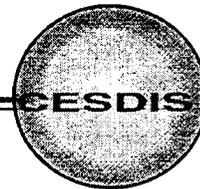
---

# ***APPENDIX C***

---

## **CESDIS Seminars**

# Seminar Announcement



Monday April 13, 1998  
Building 28, Room W230F, 11:00a.m.  
Hosted by Dr. Nabil Adam

## Interoperability Research in Digital Libraries

**Dr. William Y. Arms**  
**Vice President**

**Corporation for National Research Initiatives**

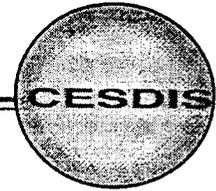
Digital library collections, such as NASA's archives, are so large and complex that they discourage extensive replication. Therefore, users draw their information from many independently managed collections. Interoperability among these independent sites has become a central research topic in digital libraries. The web provides an excellent starting point, but the simplifications that make it so successful are also barriers to sharing complex information, distributed searching, and any form of semantic interoperability. This talk discusses the current state of interoperability research and a more detailed description on work at CNRI in managing complex types of information in heterogenous digital libraries. CNRI has used a CORBA-based architecture to minimize the level of standardization necessary for exchange of information between repositories and clients, subject to access management restrictions.

*William Arms has a background in mathematics, operational research, and computing, with degrees from Oxford University, the London School of Economics, Sussex University, and Dartmouth College. He has been a pioneer in applying computing to academic activities, notably educational computing, computer networks, and digital libraries. From 1978 to 1985 he was at Dartmouth College as professor and head of computing. He then joined Carnegie Mellon University as Vice President for Computing, where his responsibilities included the Andrew project in campus-wide distributed computing, educational computing, and the university libraries. Since January 1995, he has been at the Corporation for National Research Initiatives (CNRI), where he is responsible for advanced work in digital libraries and electronic publishing.*

*For further information regarding directions,  
access to NASA Goddard Space Flight Center,  
or meeting with Dr. Arms, please contact  
Shelly Meyett at 301-286-8755.*

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement



Thursday, August 14, 1997  
Building 28, Room W230F, 10:00 a.m.

Bharat Bhargava  
Department of Computer Sciences  
Purdue University

## Large Scale Distributed Database Systems: Experiments & Observations

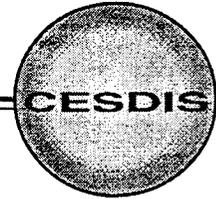
This talk identifies the research initiatives that are underway to develop applications with large dimensions. We present details of our research in communication software for building large scale (in terms of physical distribution of sites) transaction processing under the wide area network environments. Experiments that study the performance of wide area network communication will be presented. Several problems in message delivery and their impact in distributed transaction processing have been identified. We observed large variations in the communication delay and pattern of failures. We conducted experiments by connecting sites around the world. We have developed an emulation tool and used it in our experiments. We conclude that the traditional properties (ACID) for transaction model must change for success in WAN environment. For example, in Wide Area Networks, three-phase commit is not a tolerable solution. The criterion for replication consistency, serializability, and recovery requires further investigation. Ideas such as tolerable consistency, adaptability, and flexible transactions must be incorporated in the transaction model.

Some recent results for dealing with communicating multi-media documents such as digital libraries and building an adaptable video conferencing will be presented.

For further information regarding directions, access to NASA GSFC, or meeting with Dr. Bhargava, please contact Michele Meyett at 301-286-4403 or [shelly@cesdis.usra.edu](mailto:shelly@cesdis.usra.edu)

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

## Seminar Announcement



Wednesday November 26, 1997  
Building 28, Room E210, 2:00p.m.  
Hosted by Dr. Milton Halem

### **High Resolution Mapping on the Sphere for Space and Earth Applications**

**Krzysztof M. Gorski**  
**Theoretical Astrophysics Center, Copenhagen**

New generation CMB experiments aim at full sky mapping at angular resolution of a few arc minutes. Similar, or better resolution is aimed at in global Earth surface mapping and modeling. Individual spherical maps at such angular resolution comprize many millions of bins of the data. Our ability to extract science effectively from data bases of such size depends on inherent properties of the maps. I will discuss some spherical map making techniques which are commonly used at the present time in space and Earth applications, and then present a new method of spherical tessellation, its advantages in application to high resolution mapping on the sphere, and capacity to support the fast spherical harmonic transform.

*For further information regarding directions,  
accesss to NASA Goddard Space Flight Center,  
or meeting with Krzysztof M. Gorski , please contact  
Shelly Meyett at 301-286-8755.*

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement



Wednesday, August 6, 1997  
Building 28, Room E210, 2:00 p.m.

David Harel  
Weizmann Institute of Science  
Rehovot, Israel

## Some Thoughts on Statecharts, 13 Years Later *Part 1 of a 3 part series*

Statecharts were developed in 1984, as a powerful visual formalism extending state-transition diagrams. The language is used widely for the specification and design of complex reactive systems (often of concurrent, embedded, and real-time nature), that occur in the aerospace, telecommunications, automotive and control industries.

This talk will be an informal introduction to statecharts, and it will also discuss the capabilities of the STATEMATE and Rhapsody systems built around the language. Some of the issues that arose in developing the language and the tools will be discussed from both a personal and a technical point of view.

---

### The next 2 seminars are:

August 15, 1997 Bldg. 28, Room E210 2:00 pm	Computers are not Omnipotent Part 2
August 18, 1997 Bldg. 28, Room E210 10:00 am	On the Aesthetics of Diagrams Part 3

---

For further information regarding directions, access to NASA GSFC, or meeting with David Harel, please contact Georgia Flanagan at 301-286-2080 or [georgia@cesdis.usra.edu](mailto:georgia@cesdis.usra.edu)

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement



Friday, August 15, 1997  
Building 28, Room E210, 2:00 p.m.

David Harel  
Weizmann Institute of Science  
Rehovot, Israel

## Computers are not Omnipotent *Part 2*

In a cover article in April, 1984, TIME magazine quoted the editor of a software magazine as saying:

*"Put the right kind of software into a computer and it will do whatever you want it to. There may be limits on what you can do with the machines themselves, but there are no limits on what you can do with the software."*

In the talk we shall disprove this contention outright, by exhibiting a wide array of results obtained by mathematicians and computer scientists in the last 60 years. Since the results point to inherent limitations of any kind of computing device, even with unlimited resources, they have interesting philosophical implications concerning our own limitations as entities with finite mass.

Technically, we shall discuss problems that are noncomputable, as well as ones which are computable in principle but are provably intractable as far as the amount of time and memory they require. We shall discuss the famous class of NP-complete problems, jigsaw puzzles, the traveling salesman problem, timetables and scheduling, and zero-knowledge cryptographic protocols. We shall also relate these "hard" results with the "softer" ideas of heuristics and artificial intelligence.

---

### The next seminar will be:

August 18, 1997  
Bldg. 28, Room E210  
10:00 am

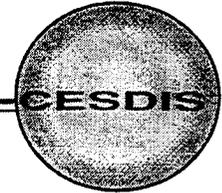
On the Aesthetics of Diagrams  
Part 3

---

For further information regarding directions, access to NASA GSFC, or meeting with David Harel, please contact Michele Meyett at 301-286-4403 or [shelly@cesdis.usra.edu](mailto:shelly@cesdis.usra.edu)

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# *Seminar Announcement*



Monday, August 18, 1997  
Building 28, Room E210, 10:00 a.m.

David Harel  
Weizmann Institute of Science  
Rehovot, Israel

## **On the Aesthetics of Diagrams *Part 3***

Given the recent move towards visual languages and visual interfaces in real-world computerised systems, the need for algorithmic procedures that produce clear and eye-pleasing layouts of complex diagrammatic entities arises in full force. This talk addresses a modest, yet still very difficult version of the problem, in which the diagrams are merely general undirected graphs with straight-line edges. We have designed a system that carries out a rather complex set of preprocessing steps, designed to produce a topologically good, but not necessarily nice-looking layout. The result is then subjected to an annealing-like beautification algorithm. The final layout is always planar for planar graphs and attempts to come close to being planar for nonplanar graphs. Future research topics will be sketched.

For further information regarding directions, access to NASA GSFC, or meeting with David Harel, please contact Michele Meyett at 301-286-4403 or [shelly@cesdis.usra.edu](mailto:shelly@cesdis.usra.edu)

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>



## Seminar Announcement

Tuesday November 18, 1997  
Building 28, Skybox, 2:00p.m.  
Hosted by Nathan Netanyahu

# Bayesian Spatial Prediction in Skewed Random Fields

**Dr. Benjamin Kedem**  
**Department of Mathematics**  
**University of Maryland, College Park**

In nearly all cases, climatological spatial data display markedly skewed distributions, a fact not always brought into serious consideration. A way out is to assume the field at hand is a transformed Gaussian random field where the transformation is 1-1 and only known to belong to a parametric family, but otherwise it is unknown. As the optimal predictor, the median of the Bayesian predictive distribution can be used because the mean of the distribution does not exist for many commonly used nonlinear transformations. The family of transformations chosen is the Box-Cox family indexed by a parameter. Using weekly rainfall amounts obtained from a network of rain gauges in Darwin, Australia, and employing Monte Carlo integration to approximate the predictive density function and its median, the Bayesian approach competes well with kriging, and the posterior of the Box-Cox parameter provides some fresh insight into the probability distribution of weekly rainfall amounts.

A web page containing the btg code will be described and delivered to all those who are interested.

*Benjamin Kedem is professor of statistics in the Dept. of Mathematics at the Univ. of Maryland College Park (UMCP). Since 1991, he has also been affiliated with the Institute of Systems Research at UMCP. He has worked extensively on time series, remote sensing, the tropical rainfall measuring mission (TRMM), data fusion and asimilation, combination of instruments, spatial prediction/interpolation, and GLM models for times series. His research, which has been supported by AFOSR, NASA, the NAVY, and NSF, spans over 70 published papers and 2 monographs. He had also directed 9 Ph.D. theses and served, in 1988, as a coast-to-coast lecturer (via satellite) on higher order crossings (HOC). Prof. Kedem is the winner of the 1984 Award for World Class Breakthrough, REFAEL, Israel, the 1986 AFOSR Achievement for his research on HOC, the 1988 IEEE Baker Award for the most outstanding IEEE journal paper of the year, and the 1997 NASA/GSFC Exceptional Achievement Award for his outstanding contributions to the TRMM.*

*For further information regarding directions,  
accesss to NASA Goddard Space Flight Center,  
or meeting with Dr. Kedem, please contact  
Shelly Meyett at 301-286-8755.*

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement



Wednesday January 21, 1998  
 Building 28, E210, 10:00a.m.  
 Hosted by Dr. Jacqueline LeMoigne

## MILAN: Prototyping a New Methodology for Reliable Parallel Processing on Distributed Environments

**Zvi M. Kedem**  
 Department of Computer Science  
 Courant Institute of Mathematical Sciences  
 New York University

The emerging computing environment will consist of a large number of time-shared machines connected by high-speed networks with subsets of individual machines possibly under the administrative control of different organizations. It is extremely difficult to utilize the aggregate power of such an environment as an effective resource for the execution of demanding applications. The problems include unpredictable behavior of the network and unpredictable availability of the machines that could be loaded unexpectedly by computations of various priorities or even crash. It is not feasible to leave such issues to an application programmer, who in general cannot anticipate the characteristics of the runtime environment.

In this presentation, I will address the utilization of an inherently distributed platform for the execution of parallel computations. I will describe the MILAN project and several of its integrated efforts, concentrating on Calypso, a prototype software system for writing and executing parallel programs on non-dedicated platforms, using standard networked machines, operating systems, and sequential compilers. It embodies at its core a unified set of techniques developed in previous theoretical research, including Eager Scheduling and the Two-Phase Idempotent Execution Strategy.

Among notable properties of Calypso are: (1) simple programming model incorporating shared memory constructs providing programmers with a virtual machine interface to the metacomputer, (2) separation of logical and execution parallelism to allow computations to scale up and down dynamically as machines join or leave an ongoing computation, and (3) transparent utilization of unreliable machines by providing dynamic load balancing and fault masking.

Calypso has been designed jointly with A. Baratloo and P. Dasgupta. It is partially based on previous theoretical research joint with Y. Aumann, K. Palem, A. Raghunathan, M. Rabin, and P. Spirakis. Current implementations run on SunOS, Solaris, Linux, Windows NT, and Windows 95.

*Zvi Kedem is currently a Professor of Computer Science in the Courant Institute of Mathematical Sciences, New York University, where he previously served also as the chair of the Department of Computer Science. He got his D.Sc. in Mathematics at the Technion - Israel Institute of Technology. His research interests included algebraic computational complexity, computer graphics, database systems, VLSI complexity, and parallel and distributed computing. He is a fellow of the ACM.*

*For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Dr. Kedem, please contact Shelly Meyett at 301-286-8755.*

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement



Tuesday December 2, 1997  
 Building 28, Room E210, 12:00p.m.  
 Hosted by Dr. Yelena Yesha

## Volumetric Imaging Model

**Hao Le**  
**Flashback Imaging Inc.**

Advances in imaging and computing technology are creating an explosion of data in image form. The need to quickly analyze this information however, is limited by inefficiencies in image display imposed by current technology. A method for rapid access and display of large image sets would considerably improve analysis and conceptualization of information contained within the images. The image

review process is often time consuming and tedious due to the lack of fast random access memory (RAM) where images normally reside. The option of using additional memory is expensive and alleviates only part of the problem. We have developed a radically different method bypassing RAM to solve this problem. The unique nature of this method is that the size of the database of images to be viewed is not limited by available RAM. The only constraint is the total size of the hard drives upon which the data is stored. Image display and animation (looping) take place directly from disk. This method enables the computer monitor to become a window of the hard drives with the capability to view large image files with unlimited resolution. The size of images does not affect the animation speed. Also, the animation can be done with the full quantitative data set. With the disk based approach, the increase in size of image data set affects only storage requirement, and not animation speed. Another important feature is that the method is successful on conventional desktop PCs. Coupling this with an easy to use User Interface, this completely new imaging model provides an opportunity for scientists and researchers as well as general public to visualize and interact with any volume of imagery.

*Hao Le, president of Flashback Imaging Inc. He received Bachelor and Master degree of Electronic Engineering from the Kyoto University, Kyoto, Japan (1975-1981) with specialization in database design for GIS and joined Environment Canada since 1982 as a computer scientist. He has been working with weather satellites and weather radars for the last 15 years. He published papers related to GIS (k-d tree, file structure, decision support systems) and satellite remote sensing applications (water surface temperature, forest fire detection, sea ice motion). He formed Flashback Imaging Inc. in 1995 to pursue his own interest in medical imaging, in particular the Visible Human Project from the National Library of Medicine. Products using data from the Visible Human Project were presented at the First Visible Human Project Conference in Bethesda, Maryland and are now permanently exhibited at the Ontario Science Centre in Toronto, Ontario.*

*For further information regarding directions, access to NASA Goddard Space Flight Center, or meeting with Hao Le please contact Shelly Meyett at 301-286-8755.*

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement

Friday, August 22, 1997  
Building 28, Room E210, 10:00 a.m.



Jorge Pinzon  
University of California, Davis

## Spatial and Spectral Feature Extraction

We present a hierarchical supervised classification technique that discriminates broad categories of surface materials in terms of ground true features, such as water, vegetation, and soils from spectral information. Subsequently, we further discriminate these materials and extract finer ground features, like chemistries, peculiar to each.

We seek to decompose the interaction at various scales between the spatial and spectral domains in the 3D domain of spatially distributed spectral data. In the spatial domain we have wavelet tools to address scale dependencies. Along the spectral axis we employ an extension of Spectral Mixture Analysis (SMA), called Hierarchical Foreground Background Analysis (HFBA). HFBA sequentially derives a series of weighting vectors for spectra that extract discriminating features at different levels of detection: (1) constituent materials, like water soil, vegetation, (2) types within constituents, like types of soil, or types of vegetation, and (3) chemistries peculiar to each type, like iron in soil, nitrogen or cellulose in vegetation. We demonstrate the information extracted by HFBA from Landsat and AVIRIS data, in contrast to a standard NDVI computation.

The direction explored is the combination of HFBA and wavelets as a supervised classification technique. In this case, the wavelet decomposition of the HFBA-represented spectral data allows us to split the spectral images related to the discriminated ground features into subimages manifesting causes of spectral changes at different scales. For example, particular components of the wavelet decomposition of the first HFBA classification image produce data which can (1) validate the categories imposed by the supervised classification, and (2) manifest clusters which can refine the classification at that level. Other components of the decomposition show the discriminatory power of the HFBA classification; for example, they reveal the extent to which the data fall outside the classification or between the classes. Extensions to unsupervised classifications are suggested by using wavelet decomposition of spectral data followed by an HFBA spectral representation. Regardless of whether training sets or a-priori information is available, a wavelet decomposition provides a means to automatically perform an unsupervised classification. For example, particular components of the wavelet decomposition of the spatially distributed spectral data manifest classes resulting from the integration of different contributing elements providing the specific levels of spectral variation needed by HFBA.

Finally, spectral redundancies were studied to compare hyper-spectral and multi-spectral information. The wavelet and HFBA decompositions provide tools which (1) allow us to simulate hyperspectral data from multi-spectral sources at different scales, (2) study how mixing is manifested at different spectral resolutions and (3) assess which targeted features may be extracted as efficiently from multi-spectral data as they could be from hyperspectral data. We can anticipate that the choice of operators derived from different combinations of wavelets basis and HFBA vectors will impact the outcomes of this study. For this purpose we have made MODIS simulations from AVIRIS spectra and compare HFBA results to study the relevance and distribution of spectral information in each waveband for particular applications.

For further information regarding directions, access to NASA GSFC, or meeting with Dr. Pinzon, please contact Georgia Flanagan at 301-286-2080 or [georgia@cesdis.usra.edu](mailto:georgia@cesdis.usra.edu)

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

## Seminar Announcement



Monday November 3, 1997  
Building 28, Room W230F, 1:00p.m.  
Hosted by Dr. Yelena Yesha

# Mining Semi-Structured Data Using a Concept Library

**Lisa Singh**  
**Electrical & Computer Engineering Dept.**  
**Northwestern University**

Knowledge discovery in databases (KDD) is the process of identifying higher level knowledge from various data sources. Although the majority of work in this area has focused on extracting knowledge from structured data, the advent of the World Wide Web (WWW) and digital libraries has generated a need for developing tools to mine semi-structured data.

Each semi-structured document contains both structured components and unstructured blocks of text. This talk will describe viable models for handling this heterogenous data in the context of data mining applications. I will then introduce an approach for efficiently generating rules by relating structured data values to concepts extracted from unstructured data. Approximation methods to improve performance will also be introduced.

*Lisa Singh is a doctoral student at Northwestern University. Her research interests include data mining of semi-structured data, sampling techniques for data mining applications, and parallel and distributed databases.*

*For further information regarding directions,  
access to NASA Goddard Space Flight Center,  
or meeting with Lisa Singh, please contact  
Shelly Meyett at 301-286-8755.*

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement



Wednesday, July 16, 1997  
Building 28, Room E210, 11:00 a.m.  
Hosted by Jacqueline Le Moigne

## Transforms for Digital Holographic Data Storage, A Progress Report

Jennifer Trelewicz  
Arizona State University

Data storage capacity requirements have grown with technological advances over the past decade. However, the current capabilities of magnetic media are near saturation of the technology. Although magneto-optical systems have shown promise, speed and density are improving more slowly as time passes. Volume holographic data storage holds the possibility of very high density storage, with highly parallel access and rapid random access. However, inter-pixel and inter-page interference have limited the practical storage density for this technology area.

This talk will discuss some linear transform methods that have been explored for use in a digital holographic data storage system. The transforms are evaluated in terms of the characteristics imposed on the data that provide resistance to channel effects. Recovery (inverse transform) methods are also discussed.

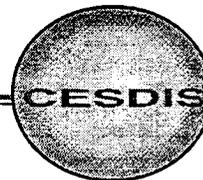
*Jennifer Trelewicz received the BS degree from Carnegie Mellon University in 1991, and the MS degree from Arizona State University in 1995. She is currently working toward the PhD in electrical engineering and the MNS in mathematics at ASU. She worked for the Motorola Government and Systems Technology Group from 1991 through 1995 as a software engineer. Her research interests include transform design and adaptive filtering.*

*Ms. Trelewicz is a member of the Phi Kappa Phi honor society, the Institute for Electrical and Electronic Engineers, the Phoenix Consultants Network, and the Society of Women Engineers. In 1996 and 1997, she received a NASA Graduate Student Researcher Program fellowship for work in holographic data storage.*

For further information regarding directions, access to NASA GSFC, or meeting with Ms. Trelewicz, please contact Georgia Flanagan at 301-286-2080 or [georgia@cesdis.usra.edu](mailto:georgia@cesdis.usra.edu)

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

## Seminar Announcement



Friday March 13, 1998  
Building 28, E210, 2:00p.m.  
Hosted by Dr. Yelena Yesha

### Active Databases for Electronic Commerce

**Victor Vianu**  
**Computer Science and Engineering**  
**University of California at San Diego**

Electronic commerce is emerging as one of the major Web-supported applications requiring database support. We introduce and study high-level declarative specifications of business models, using an approach in the spirit of active databases with immediate triggering. More precisely, business models are specified as relational transducers that map sequences of input relations into sequences of output relations. The semantically meaningful trace of an input-output exchange is kept as a sequence of log relations. We consider problems motivated by electronic commerce applications, such as log validation, verifying temporal properties of transducers, and comparing two relational transducers. Positive results are obtained for a restricted class of relational transducers called Spocus transducers (for semi-positive outputs and cumulative state). We argue that despite the restrictions, these capture a wide range of practically significant business models.

This is joint work with Serge Abiteboul (INRIA-France), Brad Fordham (Oracle) and Yelena Yesha (CESDIS).

*Victor Vianu received his PhD in Computer Science from USC in 1983. Since then, he has been on the faculty of UC San Diego and is now Professor of Computer Science. His current interests include active databases, electronic commerce, spatial databases, and querying globally distributed semistructured data. Vianu's publications include over 60 refereed research articles and a graduate textbook on database theory. He has given numerous invited talks and served as General Chair of SIGMOD and Program Chair of the PODS conference.*

*For further information regarding directions,  
access to NASA Goddard Space Flight Center,  
or meeting with Victor Vianu, please contact  
Shelly Meyett at 301-286-8755.*

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

# Seminar Announcement



Thursday, August 21, 1997  
Building 28, Room W230F  
11:30 a.m.

Ouri Wolfson  
University of Illinois at Chicago

## Location Management in Moving Objects Databases

Consider a database that represents information about moving objects and their position. For example, for a trucking company database a typical query may be: retrieve the trucks that are currently within 10 miles of truck ABT312 (which needs assistance); or for a database representing the current position of objects in a battlefield a typical query may be: retrieve the friendly helicopters that are expected to enter a given region within the next 10 minutes; or, for a satellite system: retrieve the satellites that were over Maryland on 12/13/95.

Database management system (DBMS) technology provides a foundation for efficiently answering queries about moving objects. However, there is a critical set of capabilities that have to be integrated, adapted, and built on top of existing DBMS's in order to support moving objects databases. The added capabilities include, support for spatial and temporal information, uncertainty management, rapidly changing data, and hybrid systems. The objective of our Databases fOr MovINg Objects (DOMINO) project is to build an envelope containing these capabilities on top of existing DBMS's.

For further information regarding directions, access to NASA GSFC, or meeting with Dr. Wolfson, please contact Georgia Flanagan at 301-286-2080 or [georgia@cesdis.usra.edu](mailto:georgia@cesdis.usra.edu)

<http://cesdis.gsfc.nasa.gov/admin/cesdis.seminars/seminar.html>

---

---

## ***APPENDIX D***

---

### **CESDIS Technical Reports**

**See the CESDIS Website for a  
complete set of abstracts**

**<http://cesdis.gsfc.nasa.gov/techreports.html>**

## **Nabil Adam, Rutgers University**

**TR-97-190**

**Electronic Commerce  
and Digital Libraries:  
Towards a Digital Agora**

**Nabil Adam,  
Yelena Yesha**

**January 1997**

Electronic commerce (EC) and digital libraries (DL) are two increasingly important areas of computer and information sciences with different user requirements but similar infrastructure requirements. In exploring strategic directions, we examine both requirements of the global information infrastructure that are necessary prerequisite for EC and DL [2], and specific requirements of EC and DL within the global infrastructure.

Both EC and DL are concerned with systems that support the creation of information sources and with the movement of information across global networks. EC supports effective and efficient business interactions and transactions that take place on behalf of consumers, sellers, intermediaries, and producers, while DL supports effective and efficient interaction among knowledge seekers. A digital library may require the transactional aspects of EC to manage the purchasing and distribution of its content while a digital library can be used as a resource in electronic commerce to manage products, services, providers and consumers. EC and DL share a common infrastructure in the networking, security, searching and advertising, negotiating and matchmaking, contracting and ordering, billing, payment, production, distribution, accounting, and customer service mechanisms that support such distributed information systems [31].

In a generic EC/DL model, providers (information providers, merchants, retailers, wholesalers) make multimedia objects available to consumers (customers, information seekers, users) in exchange for payment. An EC/DL system itself is characterized as a collection of distributed autonomous sites (servers) that work together to give the consumer the appearance of a single cohesive collection. Each site may store a large number of multimedia objects (documents, images, video, audio, software, structured data). This content may be stored in a variety of formats and on a variety of media such as disk, tape or CD-ROM and typically originates from a variety of providers who may wish to control its use (retrieval or modification) or to add value. Consumers are assumed to have a wide variety of domain expertise and computer proficiency which must be taken into account by designers of EC/DL systems.

Section 2 examines EC and DL research requirements in six key subareas, which section 3 provides case studies that describe three electronic commerce research projects (USC-ISI, CommerceNet, First Virtual) and six digital libraries projects sponsored by an NSF/ARPA/NASA initiatives.

**TR-97-194**

**Globalizing Business,  
Education, Culture  
Through the Internet**

**Nabil Adam,  
Baruch Awerbuch,  
Jacob Slonim,  
Peter Wegner,  
Yelena Yesha**

**February 1997**

Globalization occurs at both the national and international levels. Infrastructure is initially developed and regulated at the national level, since most utilization of the telecommunication infrastructure is within rather than among nations. Many of the technical and social questions arising at the national level are relevant to international globalization, while some issues such as interoperability among heterogeneous multilingual components occur primarily at the international level.

The technology of globalization is being driven by commercial incentives for improving the efficiency of business enterprises as well as societal concerns with improving the quality of life. We examine electronic commerce to illustrate business enterprises and education to illustrate the impact of globalization on the quality of life.

Underlying globalization is a set of technologies for human-computer interaction, finding and filtering infor-

mation, security, negotiating and matchmaking, integration and interoperability, and networking. We discuss a few of these technologies.

<b>TR-97-199</b>	<b>Information Extraction based Multiple-Category Document Classification for the Global Legal Information Network</b>	<b>Nabil Adam, Richard D. Holowczak</b>	<b>March 1997</b>
------------------	------------------------------------------------------------------------------------------------------------------------	---------------------------------------------	-------------------

This paper describes a prototype application of an information extraction (IE) based document classification system in the international law domain. IE is used to determine if a set of concepts for a class are present in a document. The syntactic and semantic constraints that must be satisfied to make this determination are derived automatically from a training corpus. A collection of IE systems are arranged in a classification hierarchy and novel documents are guided down the hierarchy based on a subset of the Global Legal Information Network domain.

<b>TR-97-200</b>	<b>The Global Legal Information Network ("GLIN")</b>	<b>Nabil Adam, Burt Edelson, Tarek El-Ghazawi, Milt Halem, Kostas Kalpakis, Nick Kosura, Rubens Medina, Yelena Yesha</b>	<b>December 1996</b>
------------------	------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------

The current globalization of the marketplace generates a greater need for cultures to learn more about one another so that decisions regarding international transactions or associations are based on trustworthy information. Additionally, many nations feel a sense of commonality not only with their immediate neighbors but also with distant trading or cultural partners. These expanding bonds help fuel the growth of common markets and greater cultural ties. Information, particularly legal information, is an essential element of these international ties because critical issues surrounding such relationships are resolved using this information. Legal researchers no longer can rely solely on the laws of a single nation to solve a legal problem; they must be able to access the law of several nations.

Fortunately, information technology has made possible faster, more accurate searches of larger and more current volumes of information. The result has been broader researching capabilities in the area of multinational comparative legal studies. Additionally, legal researchers appear to be expanding their language capabilities, as reflected in other nations. This technology may find application to worldwide databases within our lifetimes due to the great progress that has been made in machine translation.

<b>TR-97-201</b>	<b>Modeling and Analysis of Workflows Using Petri Nets</b>	<b>Nabil Adam, Vijayalakshmi Aturi, Wei-Kuang Huang</b>	<b>April 1997</b>
------------------	------------------------------------------------------------	-----------------------------------------------------------------	-------------------

A workflow system, in its general form, is basically a heterogeneous and distributed information system where the tasks are performed using autonomous systems. Resources, such as databases, labor, etc. are typically required to process these tasks. Prerequisite to the execution of a task is a set of constraints that reflect the applicable business rules and user requirements.

In this paper we present a *Petri Net* (PN) based framework that (1) facilitates specification of workflow applications, (2) serves as a powerful tool for modeling the system under study at a conceptual level, (3) allows for a smooth transition from the conceptual level to a testbed implementation and (4) enables the analysis, simulation and validation of the system under study before proceeding to implementation. Specifically, we consider three categories of task dependencies: control flow, value, and external (temporal).

We identify several structural properties of PN and demonstrate their use for conducting the following type of analyses: (1) identify inconsistent dependency specifications among tasks; (2) test for workflow safety, i.e. test whether the workflow terminates in an acceptable state; (3) for a given starting time, test whether it is feasible to execute a workflow with the specified temporal constraints.

## Dinshaw Balsara, University of Illinois

**TR-98-216**                      **Analysis of the Eigenstructure, of the Chew, Goldberger and Low System of Equations**                      **Dinshaw Balsara, Daniel Spicer**                      **September 1997**

The Chew, Goldberger and Low (CGL) System of equations applies to several situations in magnetospheric physics. It is based on making a double adiabatic approximation for the thermal pressure. In this paper we derive the eigenvalues and a complete set of left and right eigenvectors for the CGL system. The system admits eight eigenvalues, seven of which have analogues in ideal MHD. An eighth eigenvalue turns out to correspond to a new kind of advected wave. This wave produces magnetic fluctuations but the magnetic pressure is balanced by the corresponding thermal pressure fluctuation produced by the fact that the thermal pressures are anisotropic. This wave corresponds to a linearly degenerate wave. The eigenvectors for the magnetosonic waves become singular in certain limits. These are identified and eigenvector regularization is done where needed. Intuitive insights pertaining to the nature of the waves are developed. This is especially true for the eighth wave. In the regime of validity of the double adiabatic approximation the wave speeds show a strict ordering. This makes the CGL system amenable to numerical solution using upwind schemes. The linear degeneracy of the eighth wave suggests that it might be treated differently in the context of upwind schemes. Several important parallels as well as some important points of difference between the CGL system of equations and ideal MHD equations are pointed out throughout the paper.

**TR-98-217**                      **Maintaining Pressure Positivity in Magnetohydrodynamics Simulations**                      **Dinshaw Balsara, Daniel Spicer**                      **December 1997**

Higher order Godunov schemes for solving the equations of Magnetohydrodynamics (MHD) have recently become available. Because such schemes update the total energy, the pressure is a derived variable. In several problems in laboratory physics, magnetospheric physics and astrophysics the pressure can be several orders of magnitude smaller than either the kinetic energy or the magnetic energy. Thus small discretization errors in the total energy can produce situations where the gas pressure can become negative. In this paper we design a linearized Riemann solver that works directly on the entropy density equation. We also design switches that allow us to use such a Riemann solver safely in conjunction with a normal Riemann solver for MHD. This allows us to reduce the discretization errors in the evaluation of the pressure variable. As a result we formulate strategies that maintain the positivity of pressure in all circumstances. We also show via test problems that the strategies designed here work.

**TR-98-218**                      **A Staggered Mesh Algorithm Using High Order Godunov Fluxes to Ensure Solenoidal Magnetic Fields in Magnetohydrodynamic Simulations**                      **Dinshaw Balsara, Daniel Spicer**                      **December 1997**

The equations of Magnetohydrodynamics (MHD) have been formulated as a hyperbolic system of conser-

vation laws. In that form it becomes possible to use higher order Godunov schemes for their solution. This results in a robust and accurate solution strategy. However, the magnetic field also satisfies a constraint that requires its divergence to be zero at all times. This is a property that cannot be guaranteed in the zone centered discretizations that are favored in Godunov schemes without involving a divergence cleaning step. In this paper we present a staggered mesh strategy which directly uses the properly upwinded fluxes that are provided by a Godunov scheme. The process of directly using the upwinded fluxes relies on a duality that exists between the fluxes obtained from a higher order Godunov scheme and the electric fields in a plasma. By exploiting this duality we have been able to construct a higher order Godunov scheme that ensures that the magnetic field remains divergence free up to the computer's round-off error. Several stringent test problems have been devised to show that the scheme works robustly and accurately in all situations. In doing so it is shown that a scheme that involves a collocation of magnetic field variable that is different from the one traditionally favored in the design of higher order Godunov schemes can nevertheless offer the same robust and accurate performance of higher order Godunov schemes provided the properly upwinded fluxes from the Godunov methodology are used in the scheme's construction.

## Donald Becker, CESDIS

TR-98-214

**An Assessment of  
Beowulf-class, Computing  
for NASA Requirements:  
Initial Findings from the  
First NASA Workshop on  
Beowulf-class Clustered  
Computing**

**Donald Becker,  
Thomas Sterling,  
Mike Warren,  
Tom Cwik,  
John Salmon,  
Bill Nitzberg**

**January 1998**

The Beowulf class of parallel computing machine started as a small research project at NASA Goddard Space Flight Center's Center of Excellence in Space Data and Information Sciences (CESDIS). From that work evolved a new class of scalable machine comprised of mass market common off-the-shelf components (M<sup>2</sup>COTS) using a freely available operating system and industry-standard software packages. A Beowulf-class system provides extraordinary benefits in price-performance. Beowulf-class systems are in place and doing real work at several NASA research centers, are supporting NASA-funded academic research, and operating at DOE and NIH. The NASA user community conducted an intense two-day workshop in Pasadena, California on October 22-23, 1997. This first workshop on Beowulf-class systems consisted primarily of technical discussions to establish the scope of opportunities, challenges, current research activities, and directions for NASA computing employing Beowulf-class systems. The technical discussions ranged from application research to programming methodologies. This paper provides an overview of the findings and conclusions of the workshop. The workshop determined that Beowulf-class systems can deliver multi-Gflops performance at unprecedented price-performance but that software environments were not fully functional or robust, especially for larger "dreadnought" scale systems. It is recommended that the Beowulf community engage in an activity to integrate, port, or develop, where appropriate, necessary components of the software infrastructure to fully realize the potential of Beowulf-class computing to meet NASA and other agency computing requirements.

TR-98-219

**Achieving Ten Gflops  
on PC Clusters: A Case  
Study**

**Udaya Ranawake,  
John Dorband,  
Bruce Fryxell,  
Daniel Ridge,  
Erik Hendriks,  
Donald Becker,  
Phillip Merkey**

**May 1998**

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA

earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

### **Burton Edelson, George Washington University**

**TR-97-200**

**The Global Legal  
Information Network  
("GLIN")**

**Nabil Adam,  
Burt Edelson,  
Tarek El-Ghazawi,  
Milt Halem,  
Kostas Kalpakis,  
Nick Kosura,  
Rubens Medina,  
Yelena Yesha**

**December 1996**

The current globalization of the marketplace generates a greater need for cultures to learn more about one another so that decisions regarding international transactions or associations are based on trustworthy information. Additionally, many nations feel a sense of commonality not only with their immediate neighbors but also with distant trading or cultural partners. These expanding bonds help fuel the growth of common markets and greater cultural ties. Information, particularly legal information, is an essential element of these international ties because critical issues surrounding such relationships are resolved using this information. Legal researchers no longer can rely solely on the laws of a single nation to solve a legal problem; they must be able to access the law of several nations.

Fortunately, information technology has made possible faster, more accurate searches of larger and more current volumes of information. The result has been broader researching capabilities in the area of multinational comparative legal studies. Additionally, legal researchers appear to be expanding their language capabilities, as reflected in other nations. This technology may find application to worldwide databases within our lifetimes due to the great progress that has been made in machine translation.

### **Tarek El-Ghazawi, George Mason University**

**TR-97-200**

**The Global Legal  
Information Network  
("GLIN")**

**Nabil Adam,  
Burt Edelson,  
Tarek El-Ghazawi,  
Milt Halem,  
Kostas Kalpakis,  
Nick Kosura,  
Rubens Medina,  
Yelena Yesha**

**December 1996**

The current globalization of the marketplace generates a greater need for cultures to learn more about one another so that decisions regarding international transactions or associations are based on trustworthy information. Additionally, many nations feel a sense of commonality not only with their immediate neighbors but also with distant trading or cultural partners. These expanding bonds help fuel the growth of common markets and greater cultural ties. Information, particularly legal information, is an essential element of these international ties because critical issues surrounding such relationships are resolved using this infor-

mation. Legal researchers no longer can rely solely on the laws of a single nation to solve a legal problem; they must be able to access the law of several nations.

Fortunately, information technology has made possible faster, more accurate searches of larger and more current volumes of information. The result has been broader researching capabilities in the area of multi-national comparative legal studies. Additionally, legal researchers appear to be expanding their language capabilities, as reflected in other nations. This technology may find application to worldwide databases within our lifetimes due to the great progress that has been made in machine translation.

<b>TR-97-203</b>	<b>Wavelet-Based Image Registration on Parallel Computers</b>	<b>Tarek El-Ghazawi, Prachya Chalermwat, Jacqueline Le Moigne</b>	<b>November 1997</b>
------------------	-----------------------------------------------------------------------	---------------------------------------------------------------------------	----------------------

Digital image registration is very important in many applications, such as medical imagery, robotics, visual inspection, and remotely sensed data processing. In particular, NASA's Mission To Planet Earth (MTPE) program will be producing enormous Earth global change data, reaching hundreds of Gigabytes per day, that are collected from different spacecraft's and different perspectives using many sensors with diverse resolution and characteristics. The analysis of such data requires integration, therefore, accurate registration of these data. Image registration is defined as the process which determines the most accurate relative orientation between two or more images, acquired at the same or different times by different or identical sensors. Registration can also provide the absolute orientation between an image and a map.

### **Erik Hendriks, CESDIS**

<b>TR-98-219</b>	<b>Achieving Ten Gflops on PC Clusters: A Case Study</b>	<b>Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey</b>	<b>May 1998</b>
------------------	------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------	-----------------

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

## Konstantinos Kalpakis, University of Maryland Baltimore County

<b>TR-97-200</b>	<b>The Global Legal Information Network ("GLIN")</b>	<b>Nabil Adam, Burt Edelson, Tarek El-Ghazawi, Milt Halem, Kostas Kalpakis, Nick Kosura, Rubens Medina, Yelena Yesha</b>	<b>December 1996</b>
------------------	------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------

The current globalization of the marketplace generates a greater need for cultures to learn more about one another so that decisions regarding international transactions or associations are based on trustworthy information. Additionally, many nations feel a sense of commonality not only with their immediate neighbors but also with distant trading or cultural partners. These expanding bonds help fuel the growth of common markets and greater cultural ties. Information, particularly legal information, is an essential element of these international ties because critical issues surrounding such relationships are resolved using this information. Legal researchers no longer can rely solely on the laws of a single nation to solve a legal problem; they must be able to access the law of several nations.

Fortunately, information technology has made possible faster, more accurate searches of larger and more current volumes of information. The result has been broader researching capabilities in the area of multinational comparative legal studies. Additionally, legal researchers appear to be expanding their language capabilities, as reflected in other nations. This technology may find application to worldwide databases within our lifetimes due to the great progress that has been made in machine translation.

## Jacqueline LeMoigne, CESDIS

<b>TR-97-203</b>	<b>Wavelet-Based Image Registration on Parallel Computers</b>	<b>Tarek El-Ghazawi, Prachya Chalermwat, Jacqueline Le Moigne</b>	<b>November 1997</b>
------------------	---------------------------------------------------------------	---------------------------------------------------------------------------	----------------------

Digital image registration is very important in many applications, such as medical imagery, robotics, visual inspection, and remotely sensed data processing. In particular, NASA's Mission To Planet Earth (MTPE) program will be producing enormous Earth global change data, reaching hundreds of Gigabytes per day, that are collected from different spacecraft's and different perspectives using many sensors with diverse resolution and characteristics. The analysis of such data requires integration, therefore, accurate registration of these data. Image registration is defined as the process which determines the most accurate relative orientation between two or more images, acquired at the same or different times by different or identical sensors. Registration can also provide the absolute orientation between an image and a map.

<b>TR-97-206</b>	<b>Proceedings of the Image Registration Workshop</b>	<b>Jacqueline Le Moigne</b>	<b>November 1997</b>
------------------	-------------------------------------------------------	-----------------------------	----------------------

Automatic image registration has often been considered as a preliminary step for higher-level processing, such as object recognition or data fusion, but with the unprecedented amounts of data which are being and will continue to be generated by newly developed sensors. The very topic of automatic image registration has become an important research topic. The Image Registration Workshop (IRW '97), which was held at NASA/Goddard Space Flight Center on November 20-21, was one of the first to concentrate on the issue of automatic image registration. These workshop proceedings present a collection of very high quality work which has been grouped into four main areas: (1) theoretical aspects of image registration, (2) applications to satellite imagery, (3) applications to medical imagery, (4) image registration for computer vision research.



TR-97-198

**A Maximum Entropy  
Method with a Priori  
Maximum Likelihood**

**Rick Lyon,  
Jan M. Hollis,  
John E. Dorband**

April 1997

Implementations of the maximum entropy method for data reconstruction have almost universally used the approach of maximizing the statistic  $S - \lambda\chi^2$  where  $S$  is the Shannon entropy of the reconstructed distribution and  $\chi^2$  is the usual statistical measure associated with agreement between certain properties of the reconstructed distribution and the data. We develop here an alternative approach which maximizes the entropy subject to the set of constraints the  $\chi^2$  be at a minimum with respect to the reconstructed distribution. This in turn modifies the fitting statistics to be  $S - \lambda \cdot \nabla\chi^2$  where  $\lambda$  is now a vector. This new method provided a unique solution to both the well-posed and ill-posed problem, provides a natural convergence criterion which has previously been lacking in other implementations of maximum entropy, and provides the most conservative (least informative) data reconstruction result consistent with both maximum entropy and maximum likelihood methods, thereby mitigating against over-interpretation of reconstruction results. A spectroscopic example is shown as a demonstration.

### **Daniel Menascé, George Mason University**

TR-97-188

**Pythia and Pythia/ WK:  
Tools for the Performance  
Analysis of Mass Storage  
Systems**

**Odysseas I. Pentakalos,  
Daniel A. Menascé,  
Yelena Yesha**

January 1997

The constant growth on the demands imposed on hierarchical mass storage systems creates a need for frequent reconfiguration and upgrading to ensure that the response times and other performance metrics are within the desired service levels. This paper describes the design and operation of two tools, Pythia and Pythia/WK, that assist system managers and integrators in making cost-effective procurement decisions. Pythia automatically builds and solves an analytic model of a mass storage system based on a graphical description of the architecture of the system and on a description of the workload imposed the system. The use of a modeling wizard to perform this conversion unique among analytic performance tools. Pythia/WK uses clustering algorithms to characterize the workload from the log files of the mass storage system. The resulting workload characterization is used as input to Pythia.

TR-97-192

**Automated Clustering-  
Based Workload  
Characterization**

**Odysseas I. Pentakalos,  
Daniel A. Menascé,  
Yelena Yesha**

August 1996

The demands placed on the mass storage systems at various federal agencies and national laboratories are continuously increasing in intensity. This forces system managers to constantly monitor the system, evaluate the demand placed on it, and tune it appropriately using either heuristics based on experience or analytic models. Performance models require an accurate workload characterization. This can be a laborious and time consuming process. In previous studies [1,2], the authors used  $k$ -means clustering algorithms to characterize the workload imposed on a mass storage system. The result of the analysis was used as input to a performance prediction tool developed by the authors to carry out capacity planning studies of hierarchical mass storage systems [3]. It became evident from our experience that a tool is necessary to automate the workload characterization process.

<b>TR-97-202</b>	<b>Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems</b>	<b>Odysseas Pentakalos, Daniel Menascé, Yelena Yesha</b>	<b>July 1997</b>
------------------	----------------------------------------------------------------------------	------------------------------------------------------------------	------------------

Hierarchical mass storage systems are becoming more complex each day and there are many possible ways of configuring them. The options range from the type and number of devices to be used to their connectivity. An extensible object-oriented performance analyzer, called Pythia, was designed and implemented to allow users to easily investigate the most cost-effective configurations for a given workload. One of the most important reasons to build such a tool is to provide a simple way through which queuing analytic models can be used for performance prediction and system sizing of mass storage systems. The tool incorporated a modeling wizard component that is capable of automatically building a queuing network model from a mass storage system representation defined through a graphic editor. Thus, the user of the tool does not need to know queuing network modeling techniques to use it.

### **Phillip Merkey, CESDIS**

<b>TR-97-193</b>	<b>An Empirical Evaluation of the Convex SPP-1000 Hierarchical Shared Memory System</b>	<b>Thomas Sterling, Phillip Merkey, Daniel Savarese, Kevin Olson</b>	<b>August 1996</b>
------------------	-----------------------------------------------------------------------------------------	----------------------------------------------------------------------------------	--------------------

Cache coherency in a scalable parallel computer architecture requires mechanisms beyond the conventional common bus based snooping approaches which are limited to about 16 processors. The new Convex SPP-1000 achieves cache coherency across 128 processors through a two-level shared memory. NUMA structure employing directory based and SCI protocol mechanisms. While hardware support for managing a common global name space minimizes overhead costs and simplifies programming, latency considerations for remote accesses may still dominate and can under unfavorable conditions constrain scalability. This paper provides the first published evaluation of the SP-1000 hierarchical cache coherency mechanisms from the perspective of measured latency and its impact on basic global flow control mechanisms, scaling of a parallel science code, and sensitivity of cache miss rates to system scale. It is shown that global remote access latency is only a factor of seven greater than that of local cache miss penalty and the scaling of a challenging scientific application is not severely degraded by the hierarchical structure for achieving consistency across the system processor caches.

<b>TR-98-219</b>	<b>Achieving Ten Gflops on PAC Clusters: A Case Study</b>	<b>Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey</b>	<b>May 1998</b>
------------------	-----------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------	-----------------

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

## **Udaya Ranawake, University of Maryland Baltimore County**

<b>TR-98-219</b>	<b>Achieving Ten Gflops on PC Clusters: A Case Study</b>	<b>Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey</b>	<b>May 1998</b>
------------------	------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------	-----------------

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

## **Daniel Ridge, University of Maryland College Park**

<b>TR-98-219</b>	<b>Achieving Ten Gflops on PC Clusters: A Case Study</b>	<b>Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey</b>	<b>May 1998</b>
------------------	------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------	-----------------

The Beowulf project is a NASA Initiative to harness the parallelism of PC clusters built from commodity microprocessors and networking hardware and to develop the technology to apply these systems to NASA earth and space science computational needs. In this paper, we describe a case study using an important space science application that achieves more than 10 Gflops on 199 processors of a Beowulf class PC cluster. This represents nearly a ten fold increase in performance for this class of computer systems within one year. We describe the methodologies used to achieve this breakthrough and discuss the results from benchmarking runs that compare the performance of these systems with high end supercomputers such as the Cray T3E and the Convex SPP 2000.

Key words: Beowulf project, PC clusters, benchmarks, performance evaluation, scalability.

## **Yelena Yesha, CESDIS and University of Maryland Baltimore County**

<b>TR-97-188</b>	<b>Pythia and Pythia/ WK: Tools for the Performance Analysis of Mass Storage Systems</b>	<b>Odysseas I. Pentakalos, Daniel A. Menascé, Yelena Yesha</b>	<b>January 1997</b>
------------------	------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------	---------------------

The constant growth on the demands imposed on hierarchical mass storage systems creates a need for frequent reconfiguration and upgrading to ensure that the response times and other performance metrics are within the desired service levels. This paper describes the design and operation of two tools, Pythia

and Pythia/WK, that assist system managers and integrators in making cost-effective procurement decisions. Pythia automatically builds and solves an analytic model of a mass storage system based on a graphical description of the architecture of the system and on a description of the workload imposed on the system. The use of a modeling wizard to perform this conversion unique among analytic performance tools. Pythia/WK uses clustering algorithms to characterize the workload from the log files of the mass storage system. The resulting workload characterization is used as input to Pythia.

TR-97-190

**Electronic Commerce  
and Digital Libraries:  
Towards a Digital Agora**

**Nabil Adam,  
Yelena Yesha**

January 1997

Electronic commerce (EC) and digital libraries (DL) are two increasingly important areas of computer and information sciences with different user requirements but similar infrastructure requirements. In exploring strategic directions, we examine both requirements of the global information infrastructure that are necessary prerequisite for EC and DL [2], and specific requirements of EC and DL within the global infrastructure.

Both EC and DL are concerned with systems that support the creation of information sources and with the movement of information across global networks. EC supports effective and efficient business interactions and transactions that take place on behalf of consumers, sellers, intermediaries, and producers, while DL supports effective and efficient interaction among knowledge seekers. A digital library may require the transactional aspects of EC to manage the purchasing and distribution of its content while a digital library can be used as a resource in electronic commerce to manage products, services, providers and consumers. EC and DL share a common infrastructure in the networking, security, searching and advertising, negotiating and matchmaking, contracting and ordering, billing, payment, production, distribution, accounting, and customer service mechanisms that support such distributed information systems [31].

In a generic EC/DL model, providers (information providers, merchants, retailers, wholesalers) make multimedia objects available to consumers (customers, information seekers, users) in exchange for payment. An EC/DL system itself is characterized as a collection of distributed autonomous sites (servers) that work together to give the consumer the appearance of a single cohesive collection. Each site may store a large number of multimedia objects (documents, images, video, audio, software, structured data). This content may be stored in a variety of formats and on a variety of media such as disk, tape or CD-ROM and typically originates from a variety of providers who may wish to control its use (retrieval or modification) or to add value. Consumers are assumed to have a wide variety of domain expertise and computer proficiency which must be taken into account by designers of EC/DL systems.

Section 2 examines EC and DL research requirements in six key subareas, which section 3 provides case studies that describe three electronic commerce research projects (USC-ISI, CommerceNet, First Virtual) and six digital libraries projects sponsored by an NSF/ARPA/NASA initiatives.

TR-97-192

**Automated Clustering-  
Based Workload  
Characterization**

**Odysseas I. Pentakalos,  
Daniel A. Menascé,  
Yelena Yesha**

August 1996

The demands placed on the mass storage systems at various federal agencies and national laboratories are continuously increasing in intensity. This forces system managers to constantly monitor the system, evaluate the demand placed on it, and tune it appropriately using either heuristics based on experience or analytic models. Performance models require an accurate workload characterization. This can be a laborious and time consuming process. In previous studies [1,2], the authors used *k*-means clustering algorithms to characterize the workload imposed on a mass storage system. The result of the analysis was used as input to a performance prediction tool developed by the authors to carry out capacity planning studies of hierarchical mass storage systems [3]. It became evident from our experience that a tool is necessary to automate the workload characterization process.

<b>TR-97-194</b>	<b>Globalizing Business, Education, Culture Through the Internet</b>	<b>Nabil Adam, Baruch Awerbuch, Jacob Slonim, Peter Wegner, Yelena Yesha</b>	<b>February 1997</b>
------------------	------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------	----------------------

Globalization occurs at both the national and international levels. Infrastructure is initially developed and regulated at the national level, since most utilization of the telecommunication infrastructure is within rather than among nations. Many of the technical and social questions arising at the national level are relevant to international globalization, while some issues such as interoperability among heterogeneous multilingual components occur primarily at the international level.

The technology of globalization is being driven by commercial incentives for improving the efficiency of business enterprises as well as societal concerns with improving the quality of life. We examine electronic commerce to illustrate business enterprises and education to illustrate the impact of globalization on the quality of life.

Underlying globalization is a set of technologies for human-computer interaction, finding and filtering information, security, negotiating and matchmaking, integration and interoperability, and networking. We discuss a few of these technologies.

<b>TR-97-200</b>	<b>The Global Legal Information Network ("GLIN")</b>	<b>Nabil Adam, Burt Edelson, Tarek El-Ghazawi, Milt Halem, Kostas Kalpakis, Nick Kosura, Rubens Medina, Yelena Yesha</b>	<b>December 1996</b>
------------------	--------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------

The current globalization of the marketplace generates a greater need for cultures to learn more about one another so that decisions regarding international transactions or associations are based on trustworthy information. Additionally, many nations feel a sense of commonality not only with their immediate neighbors but also with distant trading or cultural partners. These expanding bonds help fuel the growth of common markets and greater cultural ties. Information, particularly legal information, is an essential element of these international ties because critical issues surrounding such relationships are resolved using this information. Legal researchers no longer can rely solely on the laws of a single nation to solve a legal problem; they must be able to access the law of several nations.

Fortunately, information technology has made possible faster, more accurate searches of larger and more current volumes of information. The result has been broader researching capabilities in the area of multinational comparative legal studies. Additionally, legal researchers appear to be expanding their language capabilities, as reflected in other nations. This technology may find application to worldwide databases within our lifetimes due to the great progress that has been made in machine translation.

<b>TR-97-202</b>	<b>Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems</b>	<b>Odysseas Pentakalos, Daniel Menascé, Yelena Yesha</b>	<b>July 1997</b>
------------------	------------------------------------------------------------------------------------	------------------------------------------------------------------	------------------

Hierarchical mass storage systems are becoming more complex each day and there are many possible ways of configuring them. The options range from the type and number of devices to be used to their connectivity. An extensible object-oriented performance analyzer, called Pythia, was designed and implemented to allow users to easily investigate the most cost-effective configurations for a given workload. One of the most important reasons to build such a tool is to provide a simple way through which queuing

analytic models can be used for performance prediction and system sizing of mass storage systems. The tool incorporated a modeling wizard component that is capable of automatically building a queuing network model from a mass storage system representation defined through a graphic editor. Thus, the user of the tool does not need to know queuing network modeling techniques to use it.



**CENTER OF EXCELLENCE IN SPACE DATA AND INFORMATION SCIENCES  
CODE 930.5  
NASA GODDARD SPACE FLIGHT CENTER  
GREENBELT, MD 20771**

**TECHNICAL REPORT SERIES ORDER FORM**

301-286-4403

Internet: [cas@cesdis1.gsfc.nasa.gov](mailto:cas@cesdis1.gsfc.nasa.gov)

Number of Copies Requested	Report Number	Title
_____	TR-90-01	Analyzing a CSMA/CD Protocol through a Systems of Communicating Machines Specification (Raymond E. Miller)
_____	TR-90-02	Altruistic Locking (Kenneth Salem)
_____	TR-90-03	Modeling the Logical Structure of Flexible Manufacturing Systems with Petri-Nets (P. David Stotts)
_____	TR-90-04	On the Bit-Complexity of Discrete Solutions of PDEs: Compact Multigrid (John Reif)
_____	TR-90-05	Rules and Principles of Scientific Data Visualization (Hikmet Senay)
_____	TR-90-06	Changes in Connectivity in Active Contour Models (Ramin Samadani)
_____	TR-90-07	Designing C++ Libraries (James M. Coggins)
_____	TR-90-08	Stabilization and Pseudo-Stabilization (Raymond E. Miller)
_____	TR-90-09	Coordinating Multi-Transaction Activities (Kenneth Salem)
_____	TR-90-10	Bounding Procedure Execution Times in a Synchronous (P. David Stotts)
_____	TR-90-11	VISTA: Visualization Tool Assistant for Viewing Scientific Data (Hikmet Senay)
_____	TR-90-12	Model-Driven Image Analysis to Augment Databases (Ramin Samadani)

---

_____	TR-90-13	Interfacing Image Processing and Computer Graphics Systems Using an Artificial Visual System (James M. Coggins)
_____	TR-90-14	Protocol Verification: The First Ten Years, The Next Ten Years; Some Personal Observations (Raymond E. Miller)
_____	TR-90-15	Coverability Graphs for a Class of Synchronously Executed Unbounded Petri Net (P. David Stotts)
_____	TR-90-16	Compositional Analysis and Synthesis of Scientific Data Visualization Techniques (Hikmet Senay)
_____	TR-90-17	Evaluation of an Elastic Curve Technique for Automatically Finding the Auroral Oval from Satellite Images (Ramin Samadani)
_____	TR-90-18	Anticipated Methodologies in Computer Vision (James M. Coggins)
_____	TR-90-19	Synthesizing a Protocol Converter from Executable Protocol Traces (Raymond E. Miller)
_____	TR-90-20	YTRACC: An Interactive Debugger for YACC Grammars (David P. Stotts)
_____	TR-90-21	Finding Curvilinear Features in Speckled Images (Ramin Samadani)
_____	TR-90-22	Multiscale Geometric Image Descriptions for Interactive Object Definition (James M. Coggins)
_____	TR-90-23	Testing Protocol Implementations Based on a Formal Specification (Raymond E. Miller)
_____	TR-90-24	A Mills-Style Iteration Theorem for Nondeterministic Concurrent Program (P. David Stotts)
_____	TR-90-25	A Computer Vision System for Automatically Finding the Auroral Oval from Satellite Images (Ramin Samadani)
_____	TR-90-26	Multiscale Vector Fields for Image Pattern Recognition (James M. Coggins)
_____	TR-90-28	Generalizing Hypertext (P. David Stotts)
_____	TR-90-29	Evaluation of an Elastic Curve Technique for Automatically Finding the Auroral Oval from Satellite Images (Ramin Samadani)
_____	TR-90-30	Interactive Object Definition in Medical Images Using Multiscale, Geometric Image Descriptions (James M. Coggins)
_____	TR-90-31	Two New Approaches to Conformance Testing of Communication Protocols (Raymond E. Miller)
_____	TR-90-32	Increasing the Power of Hypertext Search with Relational Queries (P. David Stotts)

---

_____	TR-90-33	A Multiscale Description of Image Structure for Segmentation of Biomedical Images (James M. Coggins)
_____	TR-90-34	Temporal Hyperprogramming (P. David Stotts)
_____	TR-90-35	Biomedical Image Segmentation Using Multiscale Orientation Fields (James M. Coggins)
_____	TR-90-36	Programmable Browsing Semantics in Trellis (P. David Stotts)
_____	TR-90-37	Image Structure Analysis Supporting Interactive Object Definition (James M. Coggins)
_____	TR-90-38	Separating Hypertext Content from Structure in Trellis (P. David Stotts)
_____	TR-90-39	Hierarchy, Composition, Scripting Languages, and Translators for Structured Hypertext (P. David Stotts)
_____	TR-90-40	Browsing Parallel Process Networks (P. David Stotts)
_____	TR-90-41	aTrellis: A System for Writing and Browsing Petri-Net-Based Hypertext (P. David Stotts)
_____	TR-91-42	Generating Test Sequences with Guaranteed Fault Coverage for Conformance Testing of Communication Protocols (Raymond E. Miller)
_____	TR-91-43	Specification and Analysis of a Data Transfer Protocol Using Systems of Communicating Machines (Raymond E. Miller)
_____	TR-91-44	An Exact Algorithm for Kinodynamic Planning in the Plane (John Reif)
_____	TR-91-45	Place/Transition Nets with Debit Arcs (P. David Stotts, Parke Godfrey)
_____	TR-91-46	Adaptive Prefetching for Disk Buffers (Kenneth Salem)
_____	TR-91-48	Adaptive Control of Parameters for Active Contour Models (Ramin Samadani) Visual System (James M. Coggins)
_____	TR-91-50	Structured Dynamic Behavior in Hypertext (David Stotts)
_____	TR-91-51	BLITZEN: A Highly Integrated Massively Parallel Machine (John Reif)
_____	TR-91-52	Efficient Parallel Algorithms for Optical Computing with the DFT Primitive (John Reif)
_____	TR-91-53	This Technical Report has been superceded by TR-92-87 A Minimization-Pruning Algorithm for Finding Elliptical Boundaries in Images with Non-Constant Background and with Missing Data (Ramin Samadani)
_____	TR-91-54	A Functional Meta-Structure for Hypertext Models and Systems (P. David Stotts)
_____	TR-91-55	An Optimal Parallel Algorithm for Graph Planarity (John Reif)

---

_____	TR-91-56	A Randomized EREW Parallel Algorithm for Finding Connected Components in a Graph (Hillel Gazit and John Reif)
_____	TR-91-57	Study of Six Linear Least Square Fits (Eric Feigelson)
_____	TR-91-58	Fast Computations of Vector Quantization Algorithms (John Reif)
_____	TR-91-59	Probabilistic Diagnosis of Hot Spots (Kenneth Salem)
_____	TR-91-60	Multi-Media Interaction with Virtual Worlds (Hikmet Senay)
_____	TR-91-61	Image Compression Methods with Distortion Controlled Capabilities (John Reif)
_____	TR-91-62	Management of Partially-Safe Buffers (Kenneth Salem)
_____	TR-91-63	Non-Deterministic Queue Operations (Kenneth Salem)
_____	TR-91-64	Dynamic Adaptation of Hypertext Structure (P. David Stotts)
_____	TR-91-65	Scientific Data Visualization Software: Trends and Directions (James Foley)
_____	TR-91-66	Planar Separators and the Euclidean Norm (Hillel Gazit)
_____	TR-91-67	A Deterministic Parallel Algorithm for Planar Graphs Isomorphism (Hillel Gazit)
_____	TR-91-68	A Deterministic Parallel Algorithm for Finding a Separator in Planar Graphs (Hillel Gazit)
_____	TR-91-69	An Algorithm for Finding a Separator in Planar Graphs (Hillel Gazit)
_____	TR-91-70	Optimal EREW Parallel Algorithms for Connectivity Ear Decomposition and st- Numbering of Planar Graphs (Hillel Gazit)
_____	TR-91-71	An Optimal Randomized Parallel Algorithm for Finding Connecting Components in a Graph (Hillel Gazit)
_____	TR-91-72	Modified Version of Generating Minimal Length Test Sequences for Conformance Testing of Communication Protocols (Raymond Miller)
_____	TR-91-73	Adaptive Image Segmentation Applied to Extracting the Auroral Oval from Satellite Images (Ramin Samadani)
_____	TR-91-74	Parallel Programming on the Silicon Graphics Workstation Using the Multiprocessing Library (Cynthia Starr)
_____	TR-91-75	Placing Replicated Data to Reduce Seek Delays (Kenneth Salem)
_____	TR-92-76	CESDIS Annual Report; Year 3
_____	TR-92-77	Generating Test Sequences with Guaranteed Fault Coverage for Conformance Testing of Communication Protocols (Raymond E. Miller)

---

_____	TR-92-78	On the Generation of Minimal Length Conformance Tests for Communication Protocols (Raymond E. Miller)
_____	TR-92-79	A Knowledge Based System for Scientific Data Visualization (Hikmet Senay)
_____	TR-92-80	On Generating Test Sequences for Combined Control and Data Flow for Conformance Testing of Communication Protocols (Raymond E. Miller)
_____	TR-92-81	MR-CDF: Managing Multi-Resolution Scientific Data (Kenneth Salem)
_____	TR-92-82	Adaptive Block Rearrangement (Kenneth Salem)
_____	TR-92-83	A Markov Field/Accumulator Sampler Approach to the Atmospheric Temperature Inversion Problem (Noah Friedland)
_____	TR-92-84	Adaptive Snakes: Control of Damping and Material Parameters (Ramin Samadani)
_____	TR-92-85	Faults, Errors and Convergence in Conformance Testing of Communication Protocols (Raymond E. Miller, Sanjoy Paul)
_____	TR-92-86	Research Issues for Communication Protocols (Raymond E. Miller)
_____	TR-92-87	This Technical Report Supercedes TR-91-53 A Minimization-Pruning Algorithm for Finding Elliptical Boundaries in Images with Non-Constant Background and with Missing Data (Ramin Samadani)
_____	TR-92-88	Summary Report of the CESDIS Workshop on Scientific Database Management (Kenneth Salem)
_____	TR-92-89	Structural Analysis of a Protocol Specification and Generation of a Maximal Fault Coverage Conformance Test Sequence (Raymond E. Miller)
_____	TR-92-90	Kernel-Control Parallel Versus Data Parallel: A Technical Comparison (Terrence Pratt)
_____	TR-92-91	Efficient Synchronization with Minimal Hardware Support (James H. Anderson)
_____	TR-92-92	A Fine-Grained Solution to the Mutual Exclusion Problem (James H. Anderson)
_____	TR-92-93	On the Granularity of Conditional Operations (James H. Anderson, Mohamed G. Gouda)
_____	TR-92-94	CESDIS Annual Report; Year 4
_____	TR-93-95	Image Analysis by Integration of Disparate Information (Jacqueline Le Moigne)
_____	TR-93-96	A Virtual Machine for High Performance Image Processing (Douglas Smith)

---

_____	TR-93-97	Generating Maximal Fault Coverage Conformance Test Sequences of Reduced Length for Communication Protocols (Raymond E. Miller)
_____	TR-93-98	Bounding the Performance of FDDI (Raymond E. Miller)
_____	TR-93-99	Report on the Workshop on Data and Image Compression Needs and Uses in Scientific Community (Stephen R. Tate)
_____	TR-93-100	Fine Grain Dataflow Computation without Tokens for Balanced Execution (Thomas Sterling)
_____	TR-93-101	Implementing Extended Transaction Models Using Transaction Groups (Kenneth Salem)
_____	TR-93-102	Adaptive Block Rearrangement Under UNIX (Kenneth Salem)
_____	TR-93-103	The Realities of High Performance Computing and Dataflow's Role in It: Lessons from the NASA HPCC Program (Thomas Sterling)
_____	TR-93-104	Summary Report of the CESDIS Seminar Series on Earth Remote Sensing (Jacqueline Le Moigne)
_____	TR-93-105	Space-Efficient Hot Spot Estimation (Kenneth Salem)
_____	TR-93-106	DQDB Performance and Fairness as Related to Transmission Capacity (Raymond E. Miller)
_____	TR-93-107	Deadlock Detection for Cyclic Protocols Using Generalized Fair Reachability Analysis (Raymond E. Miller)
_____	TR-94-108	Summary Report of the CESDIS Seminar Series on Future Earth Remote Sensing Missions (Jacqueline Le Moigne)
_____	TR-94-109	Generalized Fair Reachability Analysis for Cyclic Protocols: Part 1 (Raymond E. Miller)
_____	TR-94-110	CESDIS Annual Report; Year 5
_____	TR-94-111	This Technical Report has been superceded by TR-94-129. I/O Performance of the MasPar MP-1 Testbed (Tarek A. El-Ghazawi)
_____	TR-94-112	Parallel Registration of Multi-Sensor Remotely Sensed Imagery Using Wavelet Coefficients (Jacqueline Le Moigne)
_____	TR-94-113	Paradise - A Parallel Geographic Information System (David De Witt)
_____	TR-94-114	Computer Assisted Analysis of Auroral Images Obtained from High Altitude Polar Satellites (Ramin Samadani)
_____	TR-94-115	2Q: A Low Overhead High Performance Buffer Management Replacement Algorithm (Theodore Johnson)
_____	TR-94-116	Sensitivity Analysis of Frequency Counting (Theodore Johnson)

_____	TR-94-117	Client-Server Paradise (David De Witt)
_____	TR-94-118	Performance Characteristics of a 100 MegaByte/second Disk Array (Matthew T. O'Keefe)
_____	TR-94-119	Compiler and Runtime Support for Out-of-Core HPF Programs (Alok Choudhary)
_____	TR-94-120	Use of Subband Decomposition for Management of Scientific Image Databases (Kathleen G. Perez-Lopez)
_____	TR-94-121	Client-Server Paradise (David DeWitt)
_____	TR-94-122	Multi-Resolution Wavelet Decomposition on the MasPar Massively Parallel System (Jacqueline LeMoigne and Tarek A. El-Ghazawi)
_____	TR-94-123	An Initial Evaluation of the Convex SPP-1000 for Earth and Space Science Applications (Thomas Sterling, Phillip Merkey)
_____	TR-94-124	Runtime Support for Parallel I/O in PASSION (Alok Choudhary)
_____	TR-94-125	The Performance Impact of Data Placement for Wavelet Decomposition of Two Dimensional Image Data on SIMD Machines (Jacqueline LeMoigne and Tarek El-Ghazawi)
_____	TR-94-126	Planet Photo-Topography Using Shading and Stereo (Charles XiaoJian Yan)
_____	TR-94-127	Highly Scalable Data Balanced Distributed B-trees (Theodore Johnson)
_____	TR-94-128	Index Replication in a Distributed B-tree (Theodore Johnson)
_____	TR-94-129	Characteristics of the MasPar Parallel I/O System (Tarek El-Ghazawi)
_____	TR-94-130	PASSION: Parallel and Scalable Software for Input-Output (Alok Choudhary)
_____	TR-94-131	Development of a Data Reduction Expert Assistant (Glenn Miller)
_____	TR-94-132	Multivariate Statistical Analysis Software Technologies for Astrophysical Research Involving Large Data Sets (S. G. Djorgovski)
_____	TR-94-133	The Grid Analysis and Display System (GrADS) (James Kinter)
_____	TR-94-134	An Interactive Environment for the Analysis of Large Earth Observation and Model Data Sets (Kenneth Bowman and Robert Wilhelmson)
_____	TR-94-135	A Distributed Analysis and Visualization System for Model and Observational Data (Robert Wilhelmson)
_____	TR-94-136	VIEWCACHE: An Incremental Database Access Method for Autonomous Interoperable Databases (Nicholas Roussopoulos)
_____	TR-94-137	Topography from Shading and Stereo (Berthold Horn)

---

_____	TR-94-138	Experimenter's Laboratory for Visualized Interactive Science (Elaine Hansen)
_____	TR-94-139	A Land-Surface Testbed for EOSDIS (William Emery)
_____	TR-94-140	High Performance Compression of Science Data (James Storer)
_____	TR-94-141	SAVS: A Space and Atmospheric Visualization Science System (Edward P. Szuszcwicz)
_____	TR-94-142	Interactive Interface for NCAR Graphics (Bill Buzbee)
_____	TR-94-143	McIDAS-eXplorer: A Tool for Analysis of Planetary Data (Sanjay Limaye)
_____	TR-94-144	Software-based Fault Tolerance (Jonathan Bright)
_____	TR-95-145	AstroNet: A Tool Set for Simultaneous, Multi-Site Observations of Astronomical Objects (Supriya Chakrabarti)
_____	TR-95-146	Refining Image Segmentation by Integration of Edge and Region Data (Jacqueline Le Moigne and James Tilton)
_____	TR-95-147	An Approximate Performance Model of a Unitree Mass Storage System (Odysseas I. Pentakalos, Daniel A. Menasce, Milt Halem and Yelena Yesha)
_____	TR-95-148	Unsupervised, Robust Estimation-based Clustering of Remotely Sensed Images (Nathan S. Netanyahu, James C. Tilton and J. Anthony Gualtieri)
_____	TR-95-149	Knowledge Discovery from Structural Data (Diane J. Cook, Lawrence B. Holder and Surnjani Djoko)
_____	TR-95-150	Online Data Compression in a Mass Storage File System (Odysseas I. Pentakalos and Yelena Yesha)
_____	TR-95-151	A User's Guide to Pablo® I/O Instrumentation (Ruth A. Aydt)
_____	TR-95-152	Input/Output Characteristics of Scalable Parallel Applications (Phyllis E. Crandall, Ruth A. Aydt, Andrew A. Chien, Daniel A. Reed)
_____	TR-95-153	Towards a Parallel Registration of Multiple Resolution Remote Sensing Data (Jacqueline Le Moigne)
_____	TR-95-154	PPFS: A High Performance Portable Parallel File System (James V. Huber Jr., Christopher L. Elford, Daniel A. Reed, Andrew A. Chien, David S. Blumenthal)
_____	TR-95-155	An Approximate Performance Model of a Unitree Mass Storage System (Odysseas I. Pentakalos, Daniel A. Menasce, Milt Halem, Yelena Yesha)
_____	TR-95-156	Communication Strategies for Out-of-Core Programs on Distributed Memory Machines (Rajesh Bordawekar and Alok Choudhary)

---

_____	TR-95-157	Optimal Allocation for Partially Replicated Database Systems on Ring Networks (A. B. Stephens, Yelena Yesha and Keith Humenik)
_____	TR-95-158	Minimizing Message Complexity of Partially Replicated Data on Hypercubes (Keith Humenik, Peter Matthews, A. B. Stephens and Yelena Yesha)
_____	TR-95-159	Two Approaches for High Concurrency in Multicast-Based Object Replication (Theodore Johnson and Lionel Maugis)
_____	TR-95-160	Designing Distributed Search Structures with Lazy Updates (Theodore Johnson and Padmashree Krishna)
_____	TR-95-161	The Proceedings of The Petaflops Frontier Workshop-February 6, 1995 (Thomas Sterling and Michael J. MacDonald)
_____	TR-95-162	Findings of the Second Pasadena Workshop on System Software and Tools for High Performance Computing Environments (Thomas Sterling, Paul Messina and Jim Pool)
_____	TR-95-163	An Experimental Study of Input/Output Characteristics of NASA Earth and Space Sciences Applications (Michael R. Berry and Tarek El-Ghazawi)
_____	TR-95-164	An Analytic Model of Hierarchical Mass Storage Systems with Network-Attached Storage Devices (Daniel A. Menasce, Odysseas I. Pentakalos and Yelena Yesha)
_____	TR-95-165	Analytical Performance Modeling of Hierarchical Mass Storage Systems (Odysseas I. Pentakalos, Daniel Menasce, Milt Halem and Yelena Yesha)
_____	TR-95-166	CESDIS Annual Report; Year 6
_____	TR-95-167	CESDIS Annual Report; Year 7
_____	TR-96-168	Evaluation of Segmented Ethernet Interconnect Topologies for the Beowulf Parallel Workstation (Chance Reschke, Thomas Sterling, Donald J. Becker, Daniel Ridge, Phillip Merkey, Odysseas Pentakalos and Michael R. Berry)
_____	TR-96-169	The Performance of Earth and Space Science Applications on the Convex Exemplar Scalable Shared Memory Multiprocessor (Thomas Sterling, Phillip Merkey and Daniel Savarese)
_____	TR-96-170	Parallel Input/Output Issues in Sparse Matrix Computations (Sorin G. Nastea, Tarek El-Ghazawi and Ophir Frieder)
_____	TR-96-171	The Use of Wavelets for Remote Sensing Image Registration and Fusion (Jacqueline Le Moigne and Robert F. Crompt)
_____	TR-96-172	I/O, Performance Analysis, and Performance Data Immersion (Daniel A. Reed, Tara Madhyastha, Ruth A. Ayd, Christopher L. Elford, Will H. Scullin, Evgenia Smirni)

_____	TR-96-173	Optimal Allocation of Replicated Data in Tree Networks (A. B. Stephens, David M. Lazoff and Yelena Yesha)
_____	TR-96-174	Analytical Performance Modeling of Hierarchical Mass Storage Systems (Odysseas I. Pentakalos, Daniel A. Menascé, Milt Halem, Yelena Yesha)
_____	TR-96-175	Analytical Modeling of Distributed Hierarchical Mass Storage Systems with Network-Attached Storage Devices (Odysseas I. Pentakalos, Daniel A. Menascé, and Yelena Yesha)
_____	TR-96-176	MIPi: Multi-level Instrumentation of Parallel Input/Output (Michael R. Berry and Tarek A. El-Ghazawi)
_____	TR-96-177	Tuning the Performance of I/O-Intensive Parallel Applications (Anurag Acharya, Mustafa Uysal, Robert Bennett, Assaf Mendelson, Michael Beynon, Jeff Hollingsworth, Joel Saltz, Alan Sussman)
_____	TR-96-178	Interactive Smooth-Motion Animation of High Resolution Ocean Circulation Calculations (Aaron Sawdey, Derek Lee, Thomas Ruwart, Paul Woodward, Matthew O'Keefe, Rainer Bleck)
_____	TR-96-179	A comparison of data-parallel and message-passing versions of the Miami Isopycnic Coordinate Ocean Model (MICOM) (Rainer Bleck, Sumner Dean, Matthew O'Keefe, Aaron Sawdey)
_____	TR-96-180	Instrumenting a Unix Kernel for Event Tracing (Steven R. Soltis, Matthew T. O'Keefe, Thomas M. Ruwart)
_____	TR-96-181	An Object Oriented Performance Analyzer of Hierarchical Mass Storage Systems (Odysseas I. Pentakalos, Daniel A. Menascé, Yelena Yesha)
_____	TR-96-182	An Automated Parallel Image Registration Technique of Multiple Source Remote Sensing Data (Jacqueline LeMoigne, William J. Campbell, Robert F. Crompt)
_____	TR-96-183	Using High Performance Fortran for Earth and Space Science Applications (Terrence W. Pratt)
_____	TR-96-184	Performance Evaluation of Piecewise Parabolic Method on Convex Exemplar SPP1000 (Udaya A. Ranawake)
_____	TR-96-185	Accessing Sections of Out-of-Core Arrays Using an Extended Two-Phase Method (Alok Choudhary, Rajeev Thakur)
_____	TR-96-186	A Visual Database System for Image Analysis on Parallel Computers and its Application to the EOS Amazon Project (Linda Shapiro, Steven Tanimoto, James Ahrens)
_____	TR-96-187	Image Categorization Using Texture Features (Aya Soffer)
_____	TR-97-188	Pythia and Pythia/WK: Tools for the Performance Analysis of Mass Storage Systems (Odysseas Pentakalos, Daniel Menascé, Yelena Yesha)

---

_____	TR-97-189	Analytical Modeling of Robotic Tape Libraries Using Stochastic Automata (Tugrul Dayar, Odysseas Pentakalos, A. B. Stephens)
_____	TR-97-190	Electronic Commerce and Digital Libraries: Towards a Digital Agora (Nabil Adam, Yelena Yesha)
_____	TR-97-192	Automated Clustering-Based Workload Characterization (Odysseas Pentakalos, Daniel Menascé, Yelena Yesha)
_____	TR-97-193	An Empirical Evaluation of the Convex SPP-1000 Hierarchical Shared Memory System (Thomas Sterling, Daniel Savarese, Phillip Merkey, Kevin Olson)
_____	TR-97-194	Globalizing Business, Education, Culture Through the Internet (Nabil Adam, Baruch Awerbuch, Jacob Slonim, Peter Wegner, Yelena Yesha)
_____	TR-97-195	CESDIS Annual Report; Year 8
_____	TR-97-196	Hubble Space Telescope Faint Object Camera Calculated Point-Spread Functions (Rick Lyon, Jan M. Hollis, John Dorband)
_____	TR-97-197	Motion of the Ultraviolet R Aquarii Jet (Rick Lyon, Jan M. Hollis, John Dorband, W.A. Feibelman)
_____	TR-97-198	A Maximum Entropy Method with a Priori Maximum Likelihood Constraints (Rick Lyon, Jan M. Hollis, John Dorband)
_____	TR-97-199	Information Extraction Based Multiple-Category Document Classification for the Global Legal Information Network (Nabil Adam, Richard Holowczak)
_____	TR-97-200	The Global Legal Information Network (GLIN) (Nabil Adam, Burt Edelson, Tarek El-Ghazawi, Milt Halem, Kostas Kalpakis, Nick Kozura, Rubens Medina, Yelena Yesha)
_____	TR-97-201	Modeling and Analysis of Workflows Using Petri Nets (Nabil Adam, Vijayalakshmi Atluri, Wei-Kuang Huang)
_____	TR-97-202	Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems (Odysseas Pentakalos, Daniel Menascé, Yelena Yesha)
_____	TR-97-203	Wavelet-Based Image Registration on Parallel Computers (Tarek El-Ghazawi, Prachya Chalermwat, Jacqueline Le Moigne)
_____	TR-97-204	An Architecture-Independent Workload Characterization Model for Parallel Computer Architectures (Abdullah Ibrahim Meajil) Dissertation
_____	TR-97-205	Project Management - Industrial Prospective: Focusing on Industrial Software Engineering; Best Practice for Developing High (Jacob Slonim)
_____	TR-97-206	Proceedings of the Image Registration Workshop (Jacqueline Le Moigne)
_____	TR-97-207	Satellite Imaging and Sensing (Jacqueline LeMoigne, Robert Crompt)

_____	TR-97-208	Accessing Sections of Out-of-Core Arrays Using an Extended Two-Phase Method (Alok Choudhary, Rajev Thakur)
_____	TR-97-209	Efficient Compilation of Out-of-Core Data Parallel Programs (Alok Choudhary, Rajesh Bordawekar, Rajeev Thakur)
_____	TR-97-210	The Design of VIP-FS: A Virtual, Parallel File System for High Performance Parallel and Distributed Computing (Alok Choudhary, Juan Miguel del Rosario, Michael Harry)
_____	TR-97-211	Runtime Support for Parallel I/O in PASSION (Alok Choudhary, Rajeev Thakur, Rajesh Bordawekar, Ravi Ponnusamy)
_____	TR-97-212	Architecture-Independent Locality-Improving Transformations of Computational Graphs (Sanjay Ranka, Chao-Wei, Manoj Gunwani)
_____	TR-97-213	SPRINT: Scalable Partitioning, Refinement, and Incremental Partitioning Techniques (Sanjay Ranka, Chao-Wei Ou)
_____	TR-98-214	An Assessment of Beowulf-class Computing for NASA Requirements: Initial Findings from the First NASA Workshop on Beowulf-class Clustered Computing (Don Becker, Thomas Sterling, Mike Warren, Tom Cwik, John Salmon, Bill Nitzberg)
_____	TR-98-215	CESDIS Annual Report; Year 9
_____	TR-98-216	Analysis of the Eigenstructure of the Chew, Goldberger and Low System of Equations (Dinshaw Balsara, Daniel Spicer)
_____	TR-98-217	Maintaining Pressure Positivity in Magnetohydrodynamic Simulations (Dinshaw Balsara, Daniel Spicer)
_____	TR-98-218	A Staggered Mesh Algorithm Using High Order Godunov Fluxes to Ensure Solenoidal Magnetic Fields in Magnetohydrodynamic Simulations (Dinshaw Balsara, Daniel Spicer)
_____	TR-98-219	Achieving Ten Gflops on PC Clusters: A Case Study (Udaya Ranawake, John Dorband, Bruce Fryxell, Daniel Ridge, Erik Hendriks, Donald Becker, Phillip Merkey)

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

PHONE \_\_\_\_\_

E-MAIL \_\_\_\_\_

AREAS OF RESEARCH INTEREST \_\_\_\_\_

---

---

# ***APPENDIX E***

---

## **CESDIS Personnel and Associates**

## CESDIS ADMINISTRATIVE OFFICE

301-286-4403 fax: 301-286-1777

Individual extensions will roll over to another number or phonemail, if the party called does not answer. Please allow sufficient rings for this to happen.

### U. S. Mail Address

CESDIS  
Code 930.5  
NASA Goddard Space Flight Center  
Greenbelt, MD 20771

### Federal Express/UPS Address

CESDIS  
Building 28, Room W223  
NASA Goddard Space Flight Center  
Greenbelt, MD 20771

### DIRECTOR

Yelena Yesha

CESDIS: 301-286-4108  
UMBC: 410-455-3542

yesha@cesdis.edu  
yeyesha@cs.umbc.edu

### SENIOR AND STAFF SCIENTISTS

Donald Becker  
Jacqueline Le Moigne  
Les Meredith  
Phillip Merkey  
Terry Pratt

301-286-0882  
301-286-8723  
301-286-8830  
301-286-3805  
301-286-0880

becker@cesdis.edu  
lemoigne@nibbles.gsfc.nasa.gov  
les@usra.edu  
merk@cesdis.edu  
pratt@cesdis.edu

### TECHNICAL PERSONNEL

Chang-Hong Chien  
Erik Hendriks

301-286-0881  
301-286-0065

cchien@cesdis.edu  
hendriks@cesdis.edu

### ADMINISTRATIVE PERSONNEL

Nancy Campbell  
Georgia Flanagan  
Michele Meyett  
L'Tanya Pierce  
Dawn Segura

301-286-4099  
301-286-2080  
301-286-8755  
301-286-8951  
301-286-0913

campbell@cesdis.usra.edu  
georgia@cesdis.usra.edu  
shelly@cesdis.usra.edu  
pierce@cesdis.usra.edu  
dawn@cesdis.usra.edu





**University of Maryland Baltimore County**  
 Department of Computer Science and Electrical Engineering  
 5401 Wilkens Avenue  
 Baltimore, MD 21228-5398

David Ebert	410-455-3541	ebert@cs.umbc.edu
Susan Hoban	301-286-7980	shoban@pop900.gsfc.nasa.gov
Kostas Kalpakis	410-455-3143	kalpakis@cs.umbc.edu
Richard Lyon	301-286-4302	lyon@jansky.gsfc.nasa.gov
Tim Murphy	301-286-9805	murphy@albert.gsfc.nasa.gov
Udaya Ranawake	301-286-3046	udaya@neumann.gsfc.nas.gov
Miodrag (Misha) Rancic	301-286-2439	mrancic@ciga.gsfc.nasa.gov
Sushel Unninayar	301-286-2757	sushel@cesdis.edu

**University of Maryland College Park**

Center for Automation Research  
 Computer Vision Laboratory  
 College Park, MD 20742-3275

Nathan Netanyahu	301-286-4652	nathan@nibbles.gsfc.nasa.gov
------------------	--------------	------------------------------

**University of Maryland College Park**

Department of Computer Science  
 College Park, MD 20742

Santiago Egido Artega		arteaga@cs.umd.edu
-----------------------	--	--------------------

**University of Rochester**

Department of Physics and Astronomy  
 Bausch and Lomb Building  
 Rochester, NY 14627-0171

Adam Frank	716-275-1717	afrank@alethea.pas.rochester.edu
------------	--------------	----------------------------------

**University of Washington**

Department of Astronomy  
 FM-20  
 Seattle, WA 98195

George Lake	206-543-7106	lake@astro.washington.edu
-------------	--------------	---------------------------

**CESDIS ASSOCIATES**

**Brandeis University**

Computer Science Department  
Waltham, MA 02254-9110

James Storer

617-736-2714

storer@cs.brandeis.edu

**University of California**

Computer Engineering Department  
Room 225, Applied Sciences  
Santa Cruz, CA 95064

Glen Langdon

408-459-2212

langdon@cse.ucsc.edu

**University of Maryland College Park**

Department of Computer Science  
College Park, MD 20742

Joel Saltz

301-405-2684

saltz@cs.umd.edu

**University of North Carolina**

Department of Computer Science  
Sitterson Hall  
Chapel Hill, NC 27599-3175

David Stotts

919-962-1833

stotts@cs.unc.edu

**University of Waterloo**

Department of Computer Science  
Waterloo, Ontario N2L 3G1, Canada

Kenneth Salem

519-888-4567, ext 3485

kmsalem@zonker.uwaterloo.ca

## ALPHABETICAL DIRECTORY OF CESDIS PERSONNEL

Adam, Nabil	973-353-5239	adam@adam.rutgers.edu
Akyildiz, Ian	404-894-5141	ian@ee.gatech.edu
Amir, Yair	410-516-4803	yairamir@cs.jhu.edu
Balsara, Dinshaw	217-244-1481	u10596@ncsa.uiuc.edu
Becker, Don	301-286-0882	becker@cesdis.usra.edu
Campbell, Nancy	301-286-4099	campbell@cesdis.usra.edu
Chien, Chang-Hong	301-286-0881	cchien@cesdis.usra.edu
Ebert, David	410-455-3541	ebert@cs.umbc.edu
Edelson, Burt	202-994-5509	edelson@seas.gwu.edu
Egido Arteaga, Santiago		arteaga@cs.umd.edu
El-Ghazawi, Tarek	703-993-3610	tarek@science.gmu.edu
Felsher, Murray	301-428-0557	felsher@tmn.com
Flanagan, Georgia	301-286-2080	georgia@cesdis.usra.edu
Frank, Adam	716-275-1717	afrank@alethea.pas.rochester.edu
Hendriks, Erik	301-286-0065	hendriks@cesdis.edu
Hoban, Susan	301-286-7980	shoban@pop900.gsfc.nasa.gov
Kalpakis, Kostas	410-455-3143	kalpakis@cs.umbc.edu
Kouatchou, Jules		kouatcho@math.gwu.edu
Lake, George	206-543-7106	lake@astro.washington.edu
Le Moigne, Jacqueline	301-286-8723	lemoigne@nibbles.gsfc.nasa.gov
Lyon, Richard	301-286-4302	lyon@jansky.gsfc.nasa.gov
Menascé, Daniel	703-993-1537	menasce@cs.gmu.edu
Meredith, Les	301-286-8830	les@usra.edu
Merkey, Phillip	301-286-3805	merk@cesdis.usra.edu
Meyett, Michele	301-286-4403	shelly@cesdis.usra.edu
Murphy, Tim	301-286-9805	murphy@albert.gsfc.nasa.gov
Netanyahu, Nathan	301-286-4652	nathan@nibbles.gsfc.nasa.gov
Pierce, L'Tanya	301-286-8951	pierce@cesdis.usra.edu
Pratt, Terry	301-286-0880	pratt@cesdis.usra.edu
Ranawake, Udaya	301-286-3046	udaya@neumann.gsfc.nasa.gov
Rancic, Miodrag (Misha)	301-286-2439	mrancic@ciga.gsfc.nasa.gov
Segura, Dawn	301-286-0913	dawn@cesdis.usra.edu
Singh, Lisa	847-491-7141	lisa-singh@nwu.edu
Singhal, Mukesh	614-292-5839	singhal@cis.ohio-state.edu
Somerville, Richard	619-534-4644	rsomerville@ucsd.edu
Stetina, Fran	301-286-0769	fran@suzieq.gsfc.nasa.gov
Unninayar, Sushel	301-286-2757	sushel@cesdis.usra.edu
Yesha, Yelena	CESDIS: 301-286-4108 UMBC: 410-455-3542	yesha@cesdis.usra.edu yeyesha@cs.umbc.edu