NAS5-32337

CESDIS ANNUAL REPORT
Year 6
July 1993 - June 1994

Dr. Terrence W. Pratt
Acting Director

Center of Excellence in Space Data and Information Sciences

Operated by Universities Space Research Association in cooperation with the National Aeronautics and Space Administration.

A CESDIS OVERVIEW

CESDIS, the Center of Excellence in Space Data and Information Sciences was developed jointly by NASA, Universities Space Research Association (USRA), and the University of Maryland in 1988 to focus on the design of advanced computing techniques and data systems to support NASA Earth and space science research programs. CESDIS is operated by USRA under contract to NASA. The Director, Associate Director, Staff Scientists, and administrative staff are located on-site at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

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,
- Data base and intelligent data management systems for parallel computers,
- Image processing,
- Digital libraries, 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 research tasks of shorter duration for computer science research requested by NASA Goddard scientists.

The on-site staff currently consists of the Director, Associate Director, four Staff Scientists, the Senior Administrator, and three Administrative Assistants. This group provides program direction and liaison among the funded project personnel, NASA scientific and administrative personnel, and USRA accounting and contract management personnel.

CESDIS also provides programmatic and logistical expertise in the form of two Program Coordinators who work with NASA personnel as they implement NASA's High Performance Computing and Communications (HPCC) program.

INTRODUCTION FROM THE ACTING DIRECTOR

The 1993-94 CESDIS year included a broad range of computer science research applied to NASA problems. This report provides an overview of these research projects and programs as well as a summary of the various other activities of CESDIS in support of NASA and the university research community. We have had an exciting and challenging year.

Several major new initiatives funded by the NASA High Performance Computing and Communications (HPCC) Program form the basis for most of our new research projects. The most prominent new program this year is the CESDIS HPCC Basic Research Program in Parallel Computing, a program of peer-reviewed research awards to universities, with a general focus on research in input-output systems, data base systems, and intelligent data management for parallel computers. We made the announcement of this program in December 1992, collected and reviewed proposals during early 1993, and announced awards in May 1993. Ten three-year awards were made. Work actually got underway in Fall 1993 and we began to see the first results during this CESDIS year. This is an exciting set of projects with potential impact beyond the HPCC program. Already we have seen connections established between PI's in our program and the NASA EOSDIS project, where issues of input/output and database management on parallel machines are expected to be important.

We have also been able to build a more substantial in-house research staff during this CESDIS year, doubling the number of CESDIS research staff from two to four through HPCC support. Dr. Thomas Sterling, Senior Scientist, has been particularly effective in leading the CESDIS in-house HPCC research efforts. Dr. Sterling developed the concepts for the JNNIE, Beowulf, and ESS Parallel Benchmarks projects and worked to find the funding for these projects. In addition, he has played a major role in the national HPCC program, including a lead in the Petaflops Workshop held in February 1994. These activities are detailed in the sections that follow.

The CESDIS environment received a major upgrade this year. Our office space at Goddard was completely renovated and enlarged (although we have already outgrown it!). And most of our computer equipment was replaced or upgraded. We now are well-equipped, primarily with Sun workstations for researchers and high-end Macintoshes for the rest of the staff. Our thanks to the Goddard folks who made these changes possible.

A number of research projects were completed this year. The most prominent of these was the Duke University project in data compression, headed by Professor John Reif, which ended in September, 1993. This was one of the first peer-reviewed projects funded by CESDIS, and it produced a steady stream of significant research publications during its five-year term. CESDIS also managed fourteen projects for the NASA Applied Information Systems Research Program for two years, most of which ended in July 1994.

A seminar series during Fall 1993 on *Future Earth Remote Sensing Missions* was ably developed and run by Dr. Jacqueline Le Moigne of CESDIS, to the benefit of Goddard scientists. During Spring 1994, we brought the New York University Center for Digital Multimedia to Goddard for a series of ten hands-on seminars on the development of multimedia presentations for Earth and space science. A number of Goddard science groups participated and developed prototype presentations.

Finally, I note with regret the retirement of Dr. Ray Miller as CESDIS Director at the end of September 1993. CESDIS benefited greatly from Ray's leadership during its first five years of operation. A Search Committee, headed by Dr. Harold Stone of the CESDIS Science Council, is leading the search for a permanent CESDIS Director.

Terrence W. Pratt
Acting Director
July 1994

TABLE OF CONTENTS

USRA AND THE CESDIS SCIENCE COUNCIL

The Universities Space Research Association (USRA) operates CESDIS under a contract with NASA. USRA is a consortium of colleges and universities, currently numbering over 75, with graduate research programs in space sciences or related areas. USRA also operates research centers and programs at other NASA centers, including the Institute for Computer Applications in Science and Engineering (ICASE) at the NASA Langley Research Center and the Research Institute for Advanced Computer Science (RIACS) at the NASA Ames Research Center. USRA is governed by its member institutions through a Council of Institutions and a Board of Trustees and is led by its President, Dr. Paul J. Coleman, Professor of Space Physics at UCLA, and its Executive Director, Dr. W. David Cummings.

Each USRA institute or program is overseen by a Science Council that serves, in effect, as a scientific board of directors. Science Council members are appointed by the USRA Board of Trustees and typically serve three year terms. Members of the CESDIS Science Council during 1993-94 were:

- Dr. Lawrence Snyder (Convenor)
  University of Washington
- Dr. David DeWitt
  University of Wisconsin
- Dr. S. Lennart Johnsson
  Thinking Machines Corporation
- Dr. Michael O'Donnell
  University of Chicago
- Dr. Theodosios Pavlidis
  State University of New York at Stony Brook
- Dr. Patricia Selinger
  IBM Almaden Research Center
- Dr. Harold S. Stone
  IBM Thomas J. Watson Research Center
- Dr. Michael R. Stonebraker
  University of California at Berkeley
- Dr. Satish K. Tripathi
  University of Maryland

The CESDIS Science Council typically meets twice during each year at Goddard to review ongoing CESDIS research programs and future plans. The Science Council met this year on January 10-11 and on July 7, 1994.

(Tasks 1 and 26)

ACTING DIRECTOR/ASSOCIATE DIRECTOR ACTIVITIES

Terrence W. Pratt

With the retirement of Raymond Miller as CESDIS Director, Associate Director Terry Pratt assumed the role of Acting Director on October 1. He will serve in this capacity until a new Director is selected while continuing to perform as many of the functions of the Associate Director as possible.

Biographical Sketch

Dr. Pratt earned B.A., M.A., and Ph.D. degrees in mathematics and computer science at the University of Texas at Austin. Prior to joining CESDIS, he held teaching and research positions at Michigan State University in East Lansing, the University of Texas at Austin, and the University of Virginia in Charlottesville. 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 ICASE and NASA Langley scientists on the development of languages and environments for parallel computers. He is the author of two books: Programming Languages: Design and Implementation *(Prentice-Hall, 2d edition, 1984); and,* Pascal: A New Introduction to Computer Science *(Prentice-Hall, 1990).*

Dr. Pratt is a member of the Association of Computing Machinery, 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.

ACTING DIRECTOR/ASSOCIATE DIRECTOR ACTIVITIES

CESDIS was undergoing profound changes during 1993-94. Through September 1993, Dr. Raymond Miller continued as CESDIS Director, with Dr. Terrence Pratt serving as Associate Director. Dr. Miller retired from his CESDIS position at the end of September 1993. From October 1993 through June 1994 Dr. Pratt served as Acting Director while a search was made for a new permanent Director. This section was written by Dr. Pratt.

From July through September 1993, we were primarily concerned with two major CESDIS initiatives. The first was the startup of the ten new university projects in the areas of parallel input/output, parallel database management systems, and intelligent data management that had been chosen for awards under the CESDIS HPCC Basic Research Program in Parallel Computing. The first annual reports from these ten projects appear elsewhere in this report.

An important part of the Associate Director's job is to manage and facilitate the research collaborations between NASA researchers and the university projects funded by this program. To this end CESDIS hosted visits by most of the PI's of these projects to Goddard during the year and also visited some of the PI's at their home institutions.

The second major initiative during this period was the support by CESDIS of the peer review process needed by NASA for the hundreds of proposals submitted in response to the Applied Information Systems Research Program NRA 93-OSSA-09. We advised Dr. Glenn Mucklow of NASA Headquarters concerning the number and size of the review panels, the proposals to be assigned to each panel, and the membership of each panel. We also assisted in soliciting mail reviews and in tabulating the results. Awards were made by NASA under this program in late Fall 1993. Unlike the previous round of awards under this program, the administrative management of these new awards will not be handled by CESDIS.

During the period from October 1993 through April 1994, CESDIS saw rapid growth in its scientific and technical staff. We added two permanent Staff Scientists and a Program Coordinator, doubling the staff in these positions. A third temporary Staff Scientist position was also added from November 1993 through June 1994.

The 1993-94 CESDIS year saw the development of the Digital Libraries Technology program within the Space Data and Computing Division at Goddard. This program is part of the HPCC/IITA program at NASA. In preparation for a CESDIS role in this program as it develops over the next several years, we participated in numerous meetings and discussions during the year as to the shape and organization of this program.

During the period July 1993 through November 1993, we pursued the possibility of a project with Hughes Applied Information Systems, the holders of the master contract for the EOSDIS Core System, to build a program of collaborations between Hughes and university computer science researchers. Although ultimately unsuccessful in completing this

Publications

Pratt, T. W. Kernel-Control Parallel Versus Data Parallel: A Technical Comparison, presented at the Workshop on Languages, Compilers, and Run-Time Environments for Distributed Memory Machines in Boulder, Colorado, September 1992. SIGPLAN Notices, January 1993.

French, J. C., Pratt, T. W., and Das, M. Performance Measurement of the Concurrent File System of the Intel iPSC/2 Hypercube, Journal of Parallel and Distributed Computing, *February 1993.*

plan, we gained a significant number of new contacts within the EOS/EOSDIS missions at Goddard, which are expected to be important in future CESDIS work.

A major administrative activity involved the reorganization of the basic CESDIS funding mechanism. In previous years, the central management and administrative activities at CESDIS, plus a major part of the research program, were funded through "core funding" directly from NASA Headquarters. Unfortunately this funding mechanism was discontinued during 1993. All research at CESDIS is now supported through individual research "tasks" that are originated by various groups within Goddard or NASA. Also the central CESDIS management and administrative activities are now funded through a "tax" levied on these research tasks by the Space Data and Computing Division. An extensive restructuring of the CESDIS budget organization has been required to shift to these new funding mechanisms.

In addition, this year saw a complete renovation of all CESDIS office space and a complete overhaul of our computer equipment. At the end of June 1994, CESDIS was housed in newly renovated space and most of our research scientist desks held the latest in Sun, Mac, or SGI workstations.

The CESDIS Acting Director participates in numerous planning and review meetings during the year for NASA programs in which CESDIS plays a role. This was a particularly busy year for such meetings. Of major importance was the peer review of all the programs in the Goddard Space Data and Computing Division in December and several reviews of portions of the HPCC/ESS program during the year.

Research Activities

Only a modest amount of research was possible during 1993-94 due to the press of activities related to the management and administration of CESDIS. A number of avenues were explored for future research initiatives, and a small research project in the evaluation of the High Performance Fortran (HPF) language was begun.

The HPF evaluation project concerns the potential effectiveness of HPF for programming parallel machines for the types of Earth and space science (ESS) applications codes that are critical to the success of the HPCC/ESS program. HPF is a data parallel extension to Fortran 90 that includes directives for distributing large arrays on parallel machines, among other features. The evaluation of HPF concerns both (1) the usability of HPF and its data parallel features for ESS algorithms, and (2) the performance of HPF codes on the testbed parallel machines being used by the HPCC/ESS project.

The HPF evaluation project began in March 1994. During the period from March through June, the Acting Director was involved primarily in tracking the current state of the HPF language definition and its implementations for parallel machines. The governing body for HPF is the HPF Forum, a loosely organized group of researchers headed by Professor Ken Kennedy of Rice University that meets periodically to

clarify or revise the language specification and to track activity by the compiler vendors in building HPF implementations. Two meetings were held during this period, in April and June, with the primary goal of making minor clarifications and extensions to the language and of defining requirements for larger extensions that might be needed to allow HPF to apply to a broader range of scientific algorithms. Two more meetings are planned later in 1994. The Acting Director played an active role in these meetings.

Other meetings to explore possibilities for research collaborations were held with various groups from the space science laboratories at Goddard (Code 600), from the EOSDIS project, from JPL, and from ICASE. In addition, the Acting Director attended several conferences and workshops, including Supercomputing '93 in Portland during November, the Goddard Conference on Mass Storage Systems during October (including serving on the Program Committee for this conference), the HPF Forum Kickoff Meeting in Houston during January, and the Applied Information Systems Research Program Workshop in Boulder during August.

In April, the Acting Director attended the 1994 Simulation Multiconference in San Diego and made a short presentation entitled "High Performance I/O Requirements of NASA Earth Science Applications" that described some of the types of data products generated by NASA missions such as EOS and SeaWiFS, including their computational and I/O requirements. The potential for use of parallel computers for data product generation was also explored.

The Acting Director also tracked the progress of the Scalable I/O Initiative, a national group of researchers interested in scalable parallel input/output systems for high performance computers. This group is led by Paul Messina of Cal Tech and Rick Stevens of Argonne. The group was formed to pursue a large grant for a national collaboration between university, national lab, and industry researchers to address research issues in scalable I/O. The Acting Director participated in several meetings of this group during the year.

RESEARCH ACTIVITIES

CESDIS Research Staff

The role of the CESDIS research staff is to interact with Goddard Earth and space scientists to develop research projects of long term interest to NASA. The first three CESDIS staff scientists were assistant professors at the University of Maryland who devoted a portion of their time to CESDIS activities in addition to their teaching responsibilities.

There are currently four full-time staff scientists involved in three research projects which are reported in this section.

- Dr. Jacqueline Le Moigne's work on image analysis complements research within the Space Data and Computing Division's Information Science and Technology Office involving the development of data management systems which can handle the archiving and querying of data produced by Earth and space missions. Dr. Le Moigne's task involves: investigating fast, parallel methods for computing tree representations and examining search speeds that result from a variety of tree structures; investigating parallel data fusion; and, investigating image segmentation techniques for data registration, feature extraction, use of fractals, and/or wavelet transforms.

- Dr. Thomas Sterling directs high performance computing systems evaluation in support of the Earth and Space Science Project within Goddard's HPCC program. He works in collaboration with the Investigator Teams to conduct evaluations of the testbeds across applications and architectures leading to the selection of the next generation scalable teraFLOPS testbed. He is supported in this work by Dr. Philip Merkey who joined CESDIS as a staff scientist in February 1994 and Professor Tarek El-Ghazawi of George Washington University.

- Donald Becker joined CESDIS in April 1994 to explore the potential of a high performance parallel workstation built from inexpensive hardware and software. In particular he will utilize the Linux version of the Unix operating system.

A project to evaluate the new High Performance Fortran language is just beginning, under the leadership of Dr. Terrence Pratt, Acting Director of CESDIS.

Image Analysis Using the Wavelet Theory

Jacqueline Le Moigne

Biographical Sketch

Dr. Le Moigne received her Ph.D. in computer vision from the University of Paris VI, Paris, France, in 1983. Upon completing two years as a National Academy of Sciences-National Research Council Senior Resident Research Associate within the Space Data and Computing Division (930), Dr. Le Moigne joined CESDIS in October 1992.

While working with Code 930, she was involved in parallel processing on the MasPar-MP1. Projects included fusion of regions and edges for segmentation of satellite images and pose determination for space robotics. Both analyzed visual data and utilized specific data structures, such as parallel trees .

Her current work involves the multi-sensor registration and analysis of remotely sensed data. This research is of main interest for many Earth Science applications, one of them being the assessment of forested areas utilizing AVHRR and Landsat-TM data.

Dr. Le Moigne is also involved in the development of parallel algorithms developed on the MasPar MP2. Her recent work involved the development of such algorithms to assist in

Research efforts this year have centered on the use of the wavelet transformation to perform registration of multi-sensor imagery. Wavelet Transforms which had been studied last year, were further analyzed in order to extract features which are significant for image registration. An overall scheme for multi-sensor data registration was designed, and preliminary testings were made on AVHRR (Advanced Very High Resolution Radiometer) and Landsat-TM (Thematic Mapper) data. For the registration application, as well as others such as data compression, the speed of wavelet processing is a major issue, and another area of research has been to study several parallel implementations and their respective timings.

Image Registration Utilizing Wavelet Coefficients

Research performed by Jacqueline Le Moigne

Among the data which emerge as being the most important for global change research, the data for documenting and monitoring global change, such as land cover data have been identified as essential, and one example of the most important land transformations is the change in the areal extent of tropical forests. Previous studies show that the analysis of such type of data will require extrapolation among several scales (spatial, radiometric, and temporal). So, for this application, we will consider a data set including data from several sources, mainly NOAA/AVHRR-LAC and GAC (Low and Global Area Coverage, respectively), and Landsat/TM and MSS data. Other sensors that might be considered later include the SPOT-HRV and the future MODIS sensors. Before integrating this multi-sensor or multi-temporal information, it is very important to *correlate* or *register* this data taken at the same or different times by identical or different sensors. The most common approach to image registration is to choose in both input image and reference image some well defined *ground control points* (GCP's), which are usually characteristic features of the images, and then to compute the parameters of a *deformation model* which defines the transformation between reference image and input image. The main difficulty lies in the choice of the GCPs. Most commercial systems assume some interactive choice, and are not well suited for the automatic processing of a large number of data. We propose a registration method based on the use of Wavelet Transforms which performs an automatic extraction of the GCPs.

The Wavelet Transform, similarly to the Fourier Transform, is very useful in performing signal analysis and reconstruction. The Wavelet Transform provides a better localization than the Fourier Transform as well as a better division of the time-frequency plane than a Windowed Fourier Transform. With Wavelets, the window function is called the "mother wavelet", and the original signal is observed through the transla-

the study of the Wavelet Transform.

From 1983-1987, Dr. Le Moigne served as a research scientist in the University of Maryland's 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.

During 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.

Publications

Le Moigne, J. "Summary Report of the CESDIS Seminar Series on Future Earth Remote Sensing Missions." TR- 94-108.

Le Moigne, J. "Parallel Registration of Multi-Sensor Remotely Sensed Imagery Using Wavelet Coefficients." TR-94-112, also in Proceedings of SPIE's OE/Aerospace Sensing, Wavelets Applications conference, Orlando, April 5-8, 1994.

El-Ghazawi, T.A., and Le Moigne, J. "Multi-Resolution Wavelet Decomposition on the MasPar Massively Parallel System." Accepted for publication in the International Journal on Computers and their Applications.

tions as well as the dilations of the "mother wavelet". Two-dimensional (2-D) wavelets, which are continuous or discrete, are very useful for analyzing 2-D images. We only studied discrete transforms, which are more adapted to discrete signals; in this case, reconstruction is obtained by building an orthonormal basis or a "frame" (i.e., an "overcomplete" basis). As of now, we only considered orthonormal bases. One way to perform wavelet decomposition and reconstruction is to use a multi-resolution scheme (known as the "Mallat algorithm"), which involves the convolution of the original image and its successive compressed images by a low-pass and a high-pass filters. Table 1 summarizes graphically this multi-resolution decomposition process.

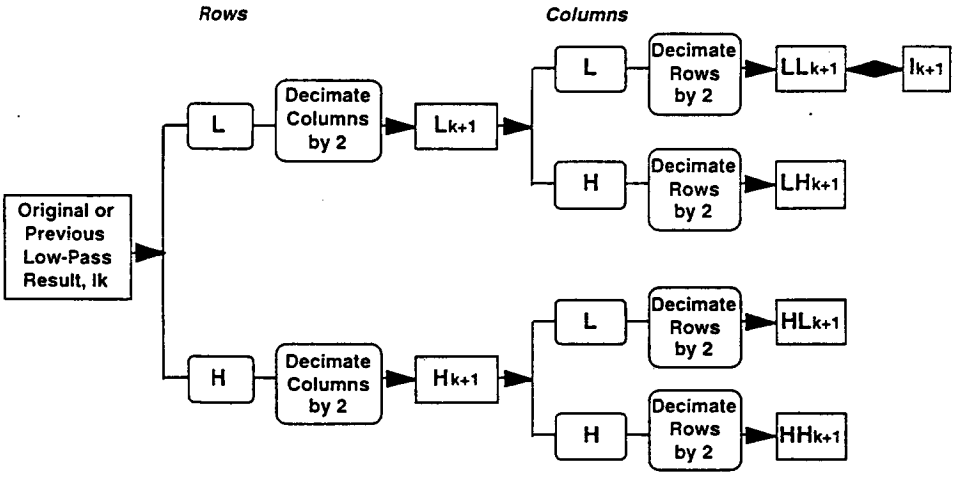


Table 1. Wavelet Multi-Resolution Decomposition (one level)

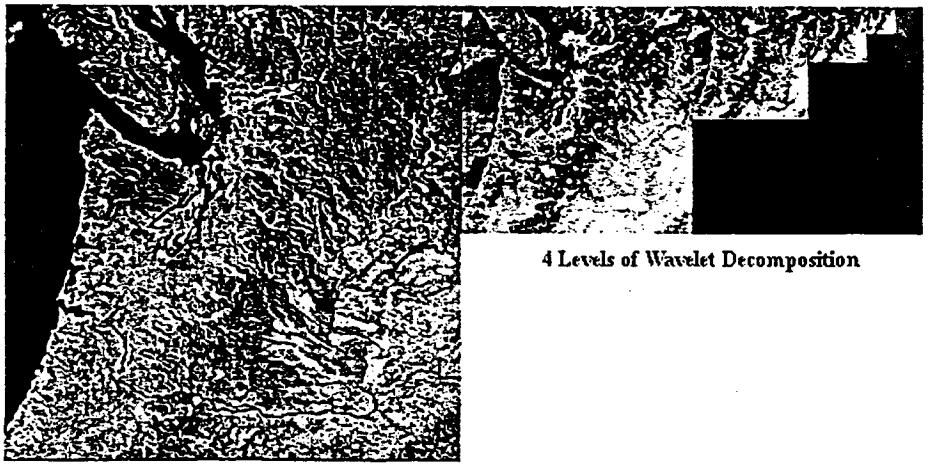


Figure 1a. Four Levels Wavelet Decomposition - AVHRR Data, Pacific NorthWest - Low/Low Compressed Images

Chan, A.K., Chui, C., Le Moigne, J., Lee, H.J., Liu, J.C., and El-Ghazawi, T.A. "The Performance Impact of Data Placement for Wavelet Decomposition of Two Dimensional Image Data on SIMD Machines." To be published in proceedings of to Frontiers'95, Fifth Symposium on the Frontiers of Massively Parallel Computation, McLean, VA, February 6-9, 1995.

Figures 1 a and b show an AVHRR image and its four levels of wavelet decomposition. For image registration purposes, only the Low/High (LH) and the High/Low (HL) coefficients, which correspond mainly to edge features and seem to be the most significant, will be utilized. Maxima of these multi-resolution wavelet coefficients are utilized as GCPs and are extracted in both reference and input data. Then, instead of matching these maxima (or GCP's) one to one, the registration is performed by searching for the best deformation model which transforms globally the set of maxima in the reference image into the set of maxima in the input image. Following the multiresolution scheme provided by the wavelet decomposition allows to reduce the search space at each level, and the registration is performed iteratively at increasing spatial resolution; at the lowest resolution, a rough approximation of the transformation is found, and it is refined iteratively going from low to high resolution. We developed and implemented such a method which searches for a rigid or affine transformation by correlating the wavelet maxima of the input image with the successive transformed maxima of the reference image. The algorithm, implemented on a massively parallel computer, the MasPar MP-1, has been tested on similar-resolution images. Figures 2 a and b show the results of this algorithm when the AVHRR image of Figure 1 has been rotated by 32 degrees. These figures show for each level, the maxima of the LH wavelet coefficients and the successive "best rotations" which are found by correlation.

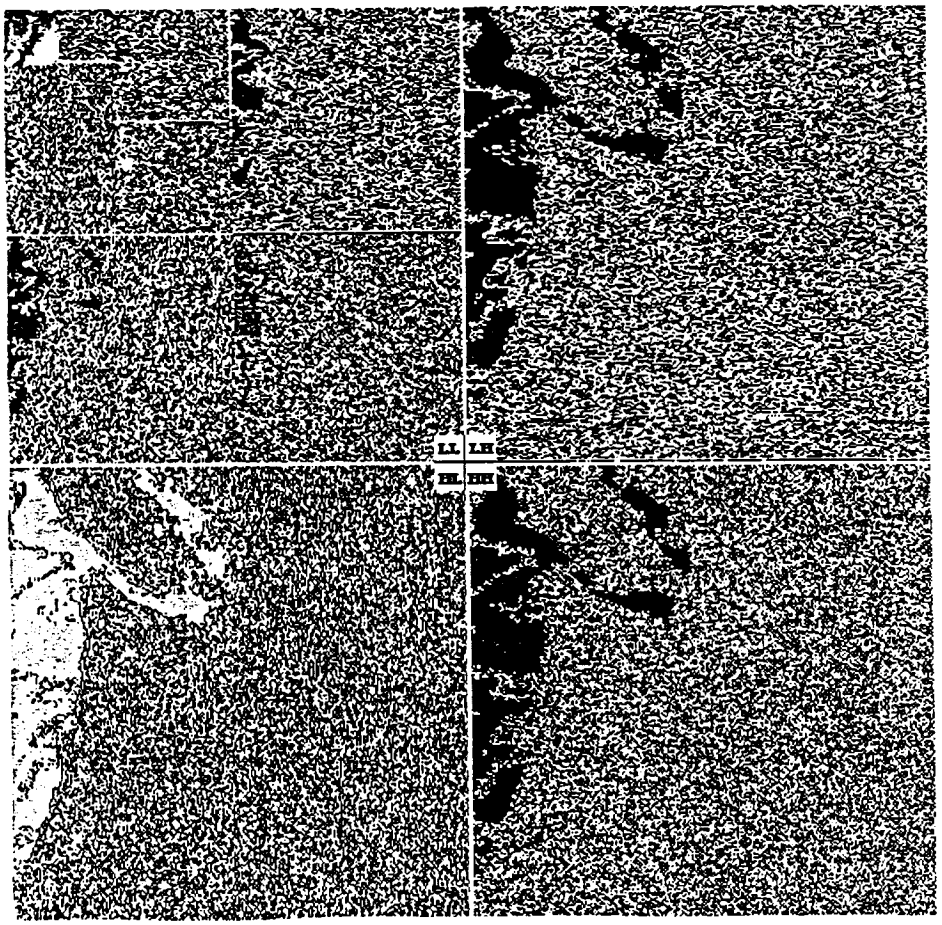


Figure 1b. Four Levels Wavelet Decomposition - AVHRR Data, Pacific NorthWest - Coefficient Images

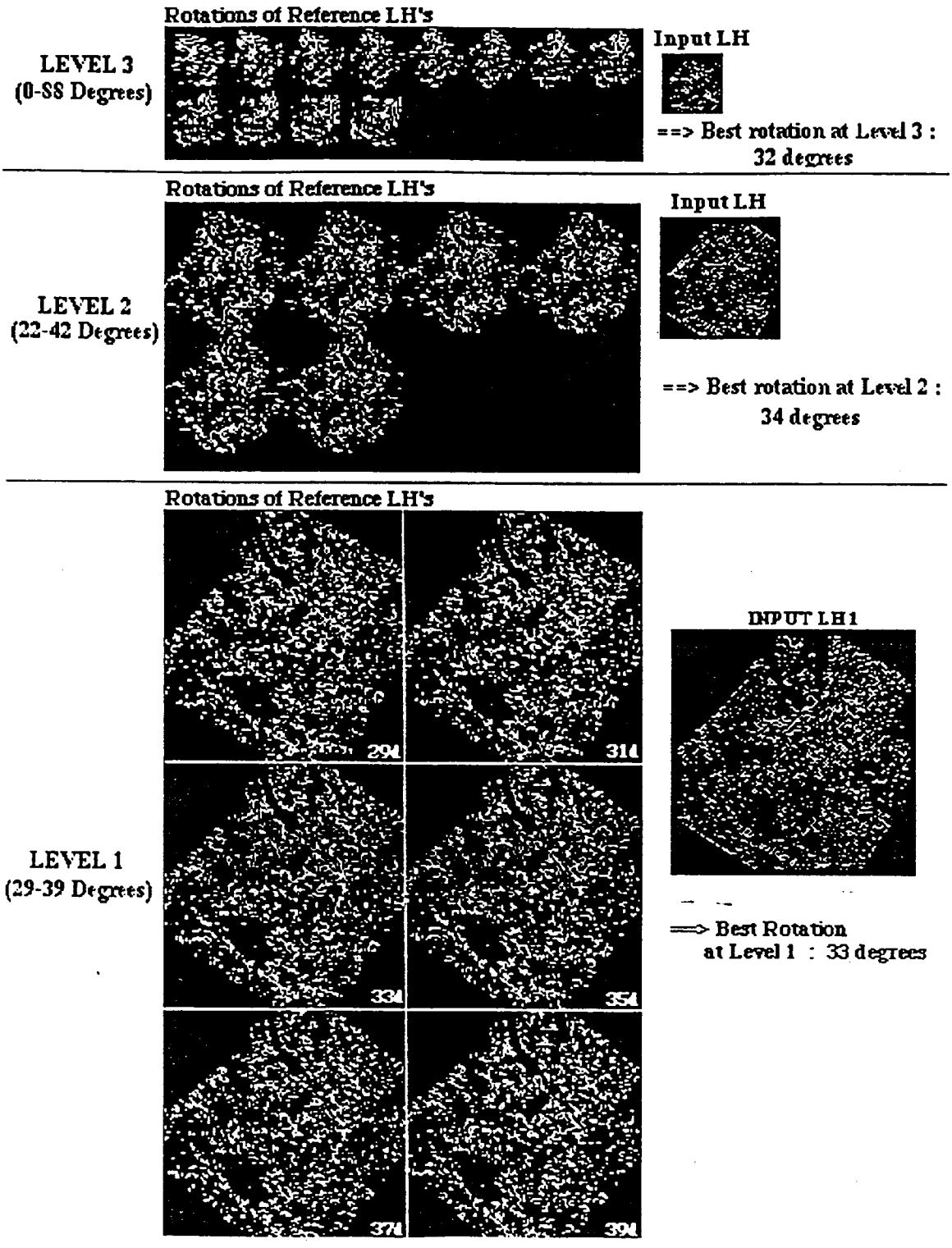


Figure 2a. Wavelet Maxima Correlation for Levels 3, 2, and 1

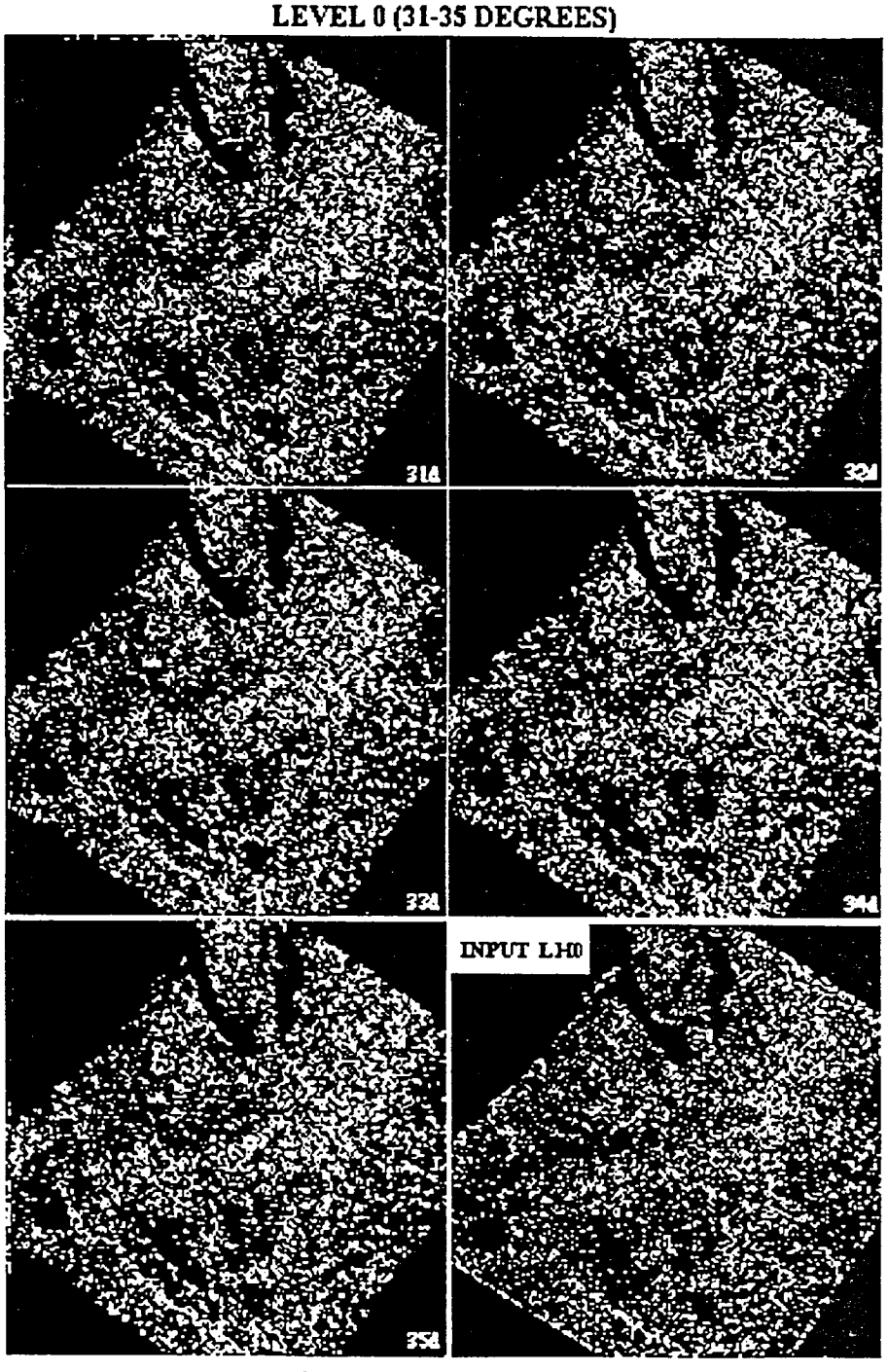


Figure 2b. Wavelet Maxima Correlation for Level 0

The algorithm is now being generalized to data of very different resolution, e.g. AVHRR and Landsat-TM data. In this case, we propose to first reduce the resolution of the high-resolution data by a wavelet decomposition, thus bringing both data to the same coarser resolution. Then, since the number of levels of wavelet decomposition is limited by the sizes of the original images, only 1 or 2 levels of decomposition can be used for the previously described "maxima matching". In order to refine even further the registration, a last iteration will be added which consists in extracting edges in the low-resolution data (e.g. AVHRR data) and then matching these edges to the wavelet coefficients maxima of the corresponding resolution in the decomposition of the high-resolution data (e.g. Landsat-TM data).

Parallel Implementation of the Wavelet Decomposition

Research performed by Jacqueline Le Moigne and Tarek El-Ghazawi (George Washington University) in collaboration with J. C. Liu (Texas A&M University)

In this study, we have investigated the parallel implementation of the Mallat algorithm (described by Table 1) on a mesh-connected, massively parallel architecture. Experimental results from mappings onto the MasPar computers, MP-1 and MP-2, and from two different virtualizations, are presented and analyzed; in particular, these algorithms, tested on a Landsat-Thematic Mapper (TM) image, emphasize the importance of communication bandwidth in achieving reasonable performance with a Massively Parallel Machine.

MasPar Computer Corporation currently produces two families of massively parallel-processor computers, namely the MP-1 and the MP-2. Both systems are essentially similar, except that the second generation (MP-2) uses 32-bit RISC processors instead of the 4-bit processors used in MP-1. The MasPar MP-1 (MP-2) is a fine-grained, massively parallel computer with Single Instruction Multiple Data (SIMD) architecture. The MasPar has up to 16,384 parallel processing elements (P.E.s) arranged in a 128x128 array, operating under the control of a central array control unit (ACU). When the image data is larger than this basic size of the parallel array, a "virtualization" of the P.E. array has to be defined. Two different techniques can be utilized for this purpose; a "Cut and Stack" technique, which divides the image into blocks and then processes it block by block, and a "Hierarchical" virtualization which stores a local neighborhood of pixels into each P.E.. We studied implementations using both types of mappings onto the parallel array.

According to the description given in Table 1, the wavelet algorithm can be defined as a combination of successive *filterings* and *decimations*. We also studied two different implementations, one where filtering and decimation are performed independently, and another one where both operations are performed simultaneously.

Table 2 summarizes the results which have been obtained on a 512X512 Landsat-TM of the Pacific Northwest with varying filter sizes and different numbers of decomposition levels.

- *Sequential* refers to an implementation on a DEC 5000 model 240 workstation.
- *CS0 Parall* is a straight-forward parallelization, using a Cut and Stack virtualization and the X-Net for filtering and decimation.
- *CS1 Parall* refers to an algorithm using a Cut and Stack virtualization, a modified "snake sweeping" algorithm for filtering, and decimation on the X-Net compacting the image in a pipelined fashion.
- *CS2 Parall* refers to an algorithm using a Cut and Stack virtualization, a modified "snake sweeping" algorithm for filtering, and the global router for decimation.
- *Hierch Parall* refers to a technique which uses a hierarchical mapping of the data, standard MasPar filtering routines, and decimation within each PE.
- *Dilution Parall* refers to an algorithm using also a hierarchical virtualization, but performing simultaneously filtering and decimation.

Filter Size	Levels of Decomp.	Sequential (sec.)	Cut and Stack				Hierarchical	
			CS0 Parall		CS1 Parall	CS2 Parall	Hierch Parall	Dilution Parall
			MP-1	MP-2				
2	1	3.13	-	2.58	2.52	0.14	0.0100	0.0109
	2	3.61	-	3.82	3.77	0.25	0.0139	0.0114
	3	3.91	-	4.45	4.40	0.37	0.0170	0.0119
	4	4.11	5.52	4.77	4.71	0.48	0.0201	0.0123
4	1	3.99	3.00	2.58	2.53	0.16	0.0140	0.0128
	2	4.54	4.54	3.88	3.81	0.30	0.0185	0.0138
	3	4.88	5.47	4.53	4.46	0.43	0.0219	0.0146
	4	4.87	5.75	4.86	4.80	0.57	0.0254	0.0154
6	1	4.50	-	2.59	2.56	0.18	0.0174	0.0148
	2	5.91	-	3.91	3.85	0.34	0.0230	0.0163
	3	6.07	-	4.59	4.52	0.49	0.0267	0.0174
	4	5.98	5.71	4.95	4.88	0.65	0.0304	0.0186
8	1	5.47	-	2.63	2.57	0.20	0.0209	0.0169
	2	7.12	-	3.96	3.87	0.38	0.0274	0.0188
	3	7.55	-	4.66	4.59	0.56	0.0314	0.0203
	4	7.55	5.84	5.05	4.97	0.73	0.0355	0.0218
10	1	6.57	-	2.65	2.59	0.22	0.0245	0.0188
	2	7.90	-	4.01	3.93	0.42	0.0319	0.0212
	3	8.86	-	4.75	4.65	0.62	0.0362	0.0231
	4	8.92	5.94	5.17	5.05	0.82	0.0404	0.0250
20	1	17.43	-	2.80	2.71	0.33	0.0410	0.0286
	2	21.21	-	4.31	4.16	0.63	0.0530	0.0333
	3	21.99	-	5.18	4.97	0.93	0.0587	0.0369
	4	23.31	6.61	5.73	5.47	1.24	0.0641	0.0407

Table 2. Results of Different Parallel Implementations of the Wavelet Decomposition (execution times in seconds)

The first observation which can be made from these results concerns the slight increase in running time for the same algorithm, *CS0*, timed on the MP-1 and the MP-2; the improvement ratio varies between 12% and 16%, although the processors of the MasPar MP-2 are eight times faster than for the MP-1. The explanation of these results lies in the percentage of the time in the algorithm spent for computations in the filtering step compared to the time spent for communication in the decimation step (since decimation is done over the X-net, which has not changed by moving to MP-2 from MP-1); from measurements made for 4 levels of decomposition and a filter of size 4, the filtering represents about 3% of the time, while the decimation represents about 97 % of the time. The same remark explains a speed improvement of about 1.2 times between the two algorithms *CS0* and *CS1*, which differ only by the modified sweeping snake method used for filtering in *CS1;* these two algorithms perform decimation over the X-net. However, with *CS2* the communication for the decimation is reduced drastically by using the global router. In the algorithm *CS2,* for 4 decomposition levels and a filter size 4, the respective time percentages for filtering and decimation are about 27% and 73%. The speed improvement of *CS2* over *CS1* is in the range of 4 to 18 times, while the improvement ratio for *CS2* over the sequential version varies from 9 to 53 times.

These results also show that, although a Cut and Stack mapping can be quite effective for such applications as a tree search problem, in the case of wavelet decomposition, a hierarchical virtualization gives much

better speed improvements. This is explained by the measurements given previously and showing that a large amount of the computing time is spent in the decimation step. Most of the decimation time can be saved if data points are organized differently. In the decimation step, only half of the output pixels is retained for further processing. If these adjacent pixels are stored in the same processor, there is no need for inter-processors routing, at least as long as the size of the image is larger than the size of the basic array (128X128 in this case). With a hierarchical mapping of the data, a speed improvement in the range of 8 to 24 times can be observed (between *CS2 Parall* and *Hierch Parall*). Compared to a sequential implementation, the parallel hierarchical *Hierch Parall* implementation offers a speed improvement in the range of 200 to nearly 600 times. And this last result is improved by combining filtering and decimation in the parallel dilution algorithm, *Dilution Parall*. Further speed improvement provided by the parallel dilution algorithm is in the range of 8 percent to 39 percent over the basic parallel hierarchical scheme, and from 99.6 percent to 99.8 percent over a sequential implementation.

Future research goals include the generalization of our algorithm for the registration of any type of multi-sensor data and its application for the analysis of data in various remote sensing studies.

High Performance Computing (HPC) Systems Evaluation for the ESS Project

Thomas Sterling and Philip Merkey

The NASA HPCC program Earth and Space Sciences (ESS) project includes an evaluation activity as a major component of its overall program objectives. The purpose of the evaluation activity is to provide an in-depth understanding of the behavior of HPC systems performing ESS-related Grand Challenge problems. One important product of this activity will be the development of the ESS Parallel Benchmarks (EPB) to do for the ESS community what the NAS Parallel Benchmarks have so successfully done for the computational aerosciences community. Of more long term significance is the research performed under this activity to analyze the sources and degree of performance degradation and their implications for scalability. These results will be useful to NASA scientists in determining the best methods for applying these new systems to their applications and for predicting the level of useful performance they are likely to experience. These results will also be useful to NASA management by providing a basis for procurement decisions and setting research goals. Lastly, these results will be useful to the hardware and software vendors to provide feedback concerning the effectiveness of their systems on real world problems.

CESDIS is supporting the ESS project goal for evaluation by providing two research scientists to coordinate the evaluation activities within NASA and among the science investigators associated with the ESS project. Closely associated with this work is an inter-agency effort being initiated in the area of evaluation. The National Science Foundation (NSF) and NASA are engaged in the Joint NSF-NASA Initiative in Evaluation (JNNIE) to perform near term studies using a shared set of metrics, sharing computational resources between the two agencies, and benefiting from this large pool of talent. The following reports document activities in each of these areas.

An Innovative Approach to Benchmarking Scalable Parallel Computers for the Earth and Space Sciences Problem Domain

Thomas Sterling and Philip Merkey with Steven Zalesak (NASA Goddard Space Flight Center, Space Data and Computing Division, High Performance Computing Branch), and Tarek El-Ghazawi (George Washington University)

Biographical Sketch

Dr. Sterling received his Ph.D in 1984 from MIT, supported through a Hertz Fellowship. Today he is a Staff Scientist at the USRA Center of Excellence in Space Data and Information Sciences where he serves in the capacity of Director of HPC Systems Evaluation for the NASA HPCC Earth and Space Science program. Prior to joining CESDIS in

With the wide array of scalable parallel processing (SPP) architectures emerging on the high performance computing (HPC) marketplace, it is becoming increasingly important to be able to assess the attributes and verify the suitability of a particular machine prior to procurement. Perhaps more importantly, at this critical juncture in the evolution of HPC system architecture during the transition from vector supercomputers to SPP systems comprising hundreds or thousands of microprocessors, it is important that the direction of computer system design be influenced by detailed quantitative information about the operational behavior and the principal factors contributing to it. Benchmarks are playing an increasingly visible role in comparative studies of new architectures. They are a valuable way of enhancing our understanding of the relative merits of different architectures. But they have a reputation for being abused and their results are often the source of contention rather than revelation. There is a new benchmarking development project underway at the NASA Goddard Space Flight Center that employs an innovative methodology.

1991, Dr. Sterling was a Research Scientist at the IDA SRC investigating multi-threading architectures and performance modeling and evaluation. He has published two dozen papers on related topics and holds six patents in parallel computer architecture and instrumentation. Dr. Sterling is a member of IEEE, ACM, and Sigma Xi.

Publications

Hamilton Island workshop paper on variable granularity optimization to be published by IEEE Press.

Paper accepted for publication for special issue on dataflow computing in Journal of Parallel and Distributed Computing on ATD architecture with Jeff Arnold.

Sterling, T.L. et al, System Software and Tools for High Performance Computing Environments, JPL California Institute of Technology, JPL Publication 93-15, April 1993.

Sterling, T.L., and MacDonald, M.J., The Realities of High Performance Computing and Dataflow's Role in It: Lessons from the NASA HPCC Program, Proceedings of the IFIP WG10.3 Working Conference on Architectures and Compilation Techniques for Fine and Medium Grain Parallelism, Horth-Holland, 1993, pgs. 165-176.

Thistle, M.R., Sterling, T.L., Kuehn, J.T., Anastasio, T.A., The Effectiveness of Random Mapping on Fine-Grain

The Earth and Space Sciences project of the NASA HPCC program includes broad research objectives and challenging computational demands. This community is actively in pursuit of methods to effectively harness SPP technology to advance its scientific goals. Means for evaluating, comparing, and contrasting alternative architectures is of critical importance within the framework of this computational context. The ESS project is in the formative stages of a three year initiative to establish a set of benchmarks that can contribute to SPP system assessment. The ESS Parallel Benchmark (EPB) set is being derived from key core codes reflecting diversity in science, algorithms, and computational techniques. Beyond the problems represented within EPB, an innovative methodology and structure is being developed to offset some of the traditional problems encountered in the use of conventional benchmarks. This report discusses the primary problems and presents key aspects of the EPB philosophy and implementation. The report is offered to the HPC community early in the evolution of the EPB suite with the object of eliciting constructive comment and critique from colleagues in the field that can be used to enhance the quality and utility of the final product.

The Earth and space science community encompasses physical dynamical phenomena from mesoscale meteorological formations to cosmic scale structures. Its applications span the range of such fundamental questions as the creation and evolution of the universe to such timely issues as climatic warming and the ozone hole. While many of the computational techniques utilized by the ESS community are of the more traditional variety with non-dynamic data structures, such as high order Godunov techniques on fixed meshes, spectral techniques (both Galerkin and collocation), and multigrid techniques on a fixed hierarchy of meshes, a good portion of the community is responding to the time varying and non-uniform structures often occurring in their computations by advancing a class of techniques that are runtime adaptive for better load balancing and superior utilization. These will rely on architectural features for data migration, context switching, and task synchronization that may not be well served by some of the extant SPP architectures. Gravitational problems using tree codes or PIC codes exhibit strong dynamic and non-uniform structures, again relying on underlying architectural support mechanisms that could significantly impact the overall system performance. A completely different class of computational requirements occurs for data assimilation problems that may have to acquire, process, and store up to a Terabyte of data a day. These observations have led to the conviction that the ESS community is involved in an array of computational problems and methods that, while overlapping other fields in important ways, presents a unique mix and therefore can only be represented by a new benchmark set. The EPB suite is conceived to fill this requirement.

A benchmark set, to be an effective tool for comparative analysis, must exhibit two attributes that are often in conflict. The validity of a comparative experiment rests on the equivalence of the workloads to which the target architectures are applied. Unless the respective workloads can be guaranteed to be, in some repeatable sense, the same, there is no basis for drawing meaningful conclusions from the outcome. For sequential architectures, benchmarks such as LINPACK and SPECmarks achieve this equivalence through fixed source code representation. Then, both the architectures and their respective opti-

MIMD Architectures, Proceedings of the 1992 International Conference on Parallel Processing, CRC Press, Aug. 1992, pgs. II-41 - II-48.

Sterling, T.L. Findings from the Pasadena Workshop on HPC Software Technology, Proceedings of the Thirty-Eighth IEEE Computer Society International Conference, IEEE Computer Society Press, February 1993, pg. 200-204.

Sterling, T.L., Hardware Support for Multiprocessor Instrumentation: Lessons from Six Examples, 1993 Workshop on Parallel Computing Systems, April 1993.

mizing compilers are being tested. While the actual number of primitive operations may vary somewhat between architectures, the basic level, structure, and functionality of the two workloads is retained, providing for credible results.

Portability of source code is to a large extent a solved problem on sequential machines (special system function calls, and language enhancements aside). The situation is quite different for current generation SPP architectures. Not only are languages not compatible, but even the programming paradigm itself may vary drastically among systems and programmers. Message passing, data parallel, concurrent threads, object oriented and other techniques are currently employed in a continuing quest for an effective and general programming methodology. Porting of parallel programs can be an arduous task with little certainty that the investment in development of a particular code will extend beyond the immediate target hardware/software system. Yet portability of a parallel benchmark is more than a convenience, it is essential if the benchmark is to provide the controlled working set against which two or more disparate parallel systems are to be compared. The NAS Parallel Benchmarks (NPB) embodies a particularly successful approach to this challenge by adopting a higher level abstraction than source code as the specification for the workload. Each of eight kernels are specified by a description of the algorithm to be implemented rather than the precise source code for it. This circumvents the problem of compatibility of parallel programming languages or execution paradigms while retaining the essential equivalence of the benchmarks at the functionality level. It also enhances programmer flexibility in implementation details to achieve best performance thus permitting specific capabilities unique to a particular architecture to be brought to bear on the problem where appropriate.

The ESS Parallel Benchmarks take a third approach that attempts to capture the best of both previous approaches while enhancing portability and control at the same time. A subtle but important issue is the question of what exactly is being measured. Benchmarks like SPEC measure source code workload. Benchmarks like NAS measure performance as it relates to functionality. Both kinds of comparisons are meaningful but they mean different things. As described below, the EPB takes a dual path to achieve needed control and portability on the one hand, and enable flexibility on the other. The key contribution is that source code specification of useful work is retained ensuring source code workload level equivalence. At the same time, portability is achieved through agent templates. Finally, algorithmic generality is attained through a second broad level of problem driven specifications.

The EPB approach embraces the dual track approach just introduced. The Base Level Series permits control of the benchmarks' useful workload while enabling portability. The second track is called the Challenge Series and provides extreme flexibility that goes beyond limitations of specific parallel algorithms. First, we describe the Base Level Series. Each benchmark of the Base Level Series consists of a set of concurrent tasks, each of which is a sequential thread of Fortran or C code. Together, the work specified in the source code threads comprises all of the useful work necessary to be performed to achieve the functionality of the problem and algorithm. The only exception is in the information exchange among tasks. For a pure uniprocessor imple-

Biographical Sketch

Dr. Merkey joined CESDIS in February of 1994 after working with the Supercomputing Research Center as a Research Staff Member. He was at SRC from 1986-1994 in a classified work environment.

His primary professional interests are parallel computer systems architecture including: performance evaluation, modeling and analysis of parallel computers. He is also interested in applied mathematics including information theory, combinatorics, and algorithms.

Dr. Merkey obtained his Ph.D. in mathematics from the University of Illinois in 1986.

mentation, these would simply be read/writes to global memory. But for parallel implementation, these interactions are performed by more general functions or agents. While the semantics of the agents are defined by the Base Level Series, their implementation is left to the programmer and can be adapted to the target parallel architecture. These communication/synchronization agents are considered as overhead work, to be differentiated from the useful work of the specified threads. The agents will vary in number of operations performed according to architecture and control strategy. Other agents will define load balancing mechanisms. These can in fact be implemented as no-op's if the programmer desires, saving on overhead at the possible risk of poor work distribution. Lastly, specific agents will support some instrumentation primitives. This is described in more detail below.

The EPB Base Series is employing a method of program specification referred to as Templates or Skeletons. Templates are being offered by the HPC community as an exciting approach to writing portable parallel codes. The ESS project takes Template one step further as a basis for portable parallel benchmarks. It must be understood that there is no expectation that the benchmark algorithm is in some sense optimized for a particular architecture. The fixed sequential tasks limit the flexibility that the programmer has in this optimization. However, for parallel programming, the most important aspect from the performance standpoint is in the global control and communication methods as well as data partitioning and allocation. These are allowed to be crafted to best fit the host parallel architecture allowing controlled comparative experiments.

The Challenge Series is devised to emphasize the other important aspect of the NASA HPCC ESS project: the end science. Here, there is no specification of the algorithm, let alone the actual code to be used. Rather, the Challenge Series is a set of science problems that represent important computational challenges to the ESS community. The purpose of the Challenge Series is to provide a basis for tracking the medium and long term advances of the ESS community in achieving its computational goals. The input and output data sets are specified along with acceptable error bounds. Certain other limitations are imposed to avoid intentional or unintentional cheating. Within these constraints, the programmer can apply whatever algorithm best suits a target architecture including special purpose or domain specific systems. The only thing that is being measured here is how the industry is advancing in its support for the ESS problem domain. It can not be used to compare two architectures other than to indicate the best performance recorded in each case. This does provide at least an upper lower bound on the performance for the given benchmark problem. Needless to say, the Challenge Series is both flexible and portable. But the only control is at the (declarative) problem level.

Most benchmark activities represent their results as execution time or some projection of the measured performance onto some value or short set of values. For example, the synthetic Whetstone benchmark provides performance measure in "Whetstones" which is a weighted product of different types of floating point operations. The NPB delivers execution time which is the final arbiter of quality. The EPB suite embraces the NPB philosophy but extends it to express detail about the factors contributing to the observed system behavior and execution times. This includes useful work, overhead, and idle cycles due to

latency and starvation (insufficiency of program parallelism or poor load balancing). As alluded to above, one class of agents is dedicated to measurement of event timings. To distinguish between overhead work and latency caused delays, an ancillary set of benchmarks called the Shadow Series provides skeletons of the Base Level Series without any of the useful work but supporting the infrastructure of agent communication. The Shadow series allows the intrusive nature of the timing measurement agents to be accounted for, although certain subtle event ordering pathologies can not be necessarily masked. The key value of this advanced measurement technique is that the Base Level Series timings corrected by the Shadow Series measurements will provide values of gross execution time and execution budget time. The latter will permit comparisons between architectures in terms of their performance degradation.

An initial release of the EPB 1.0 is scheduled for release 4th Quarter 1994.

The following report discusses work in progress on a subcontract with George Washington University for research to provide an in depth understanding of performance and workload characteristics for parallel computer architectures used in NASA Earth and space sciences. A set of ESS Grand Challenge applications is being used to conduct the study.

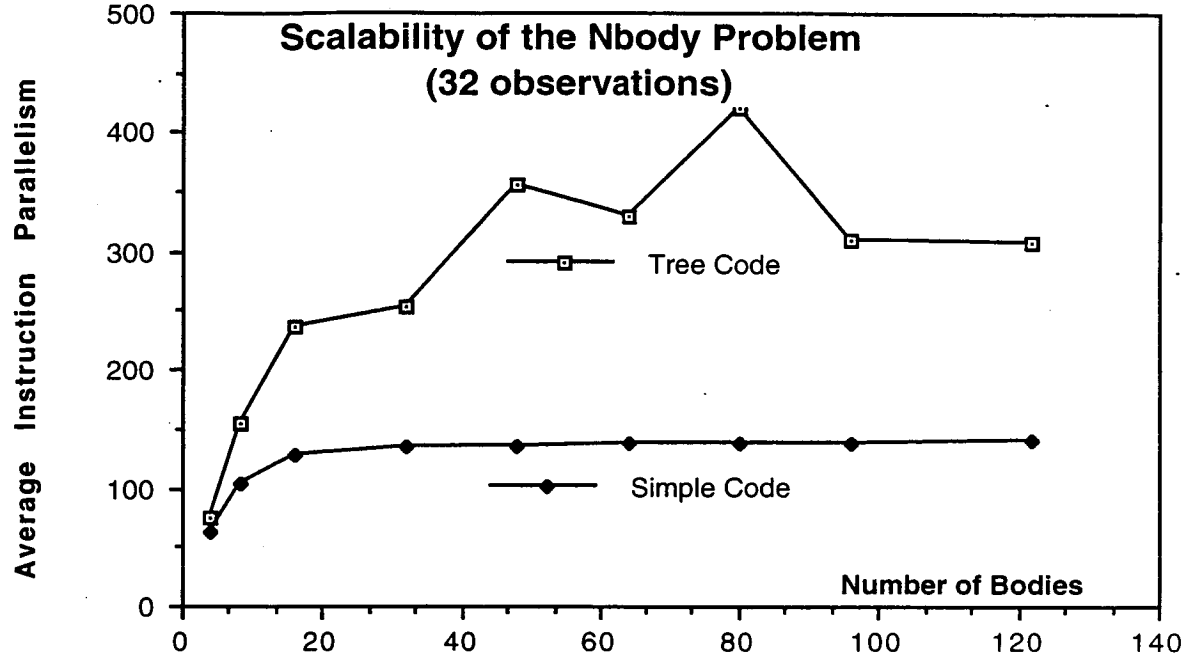
Workload Characterization for Scalable Computer Architectures

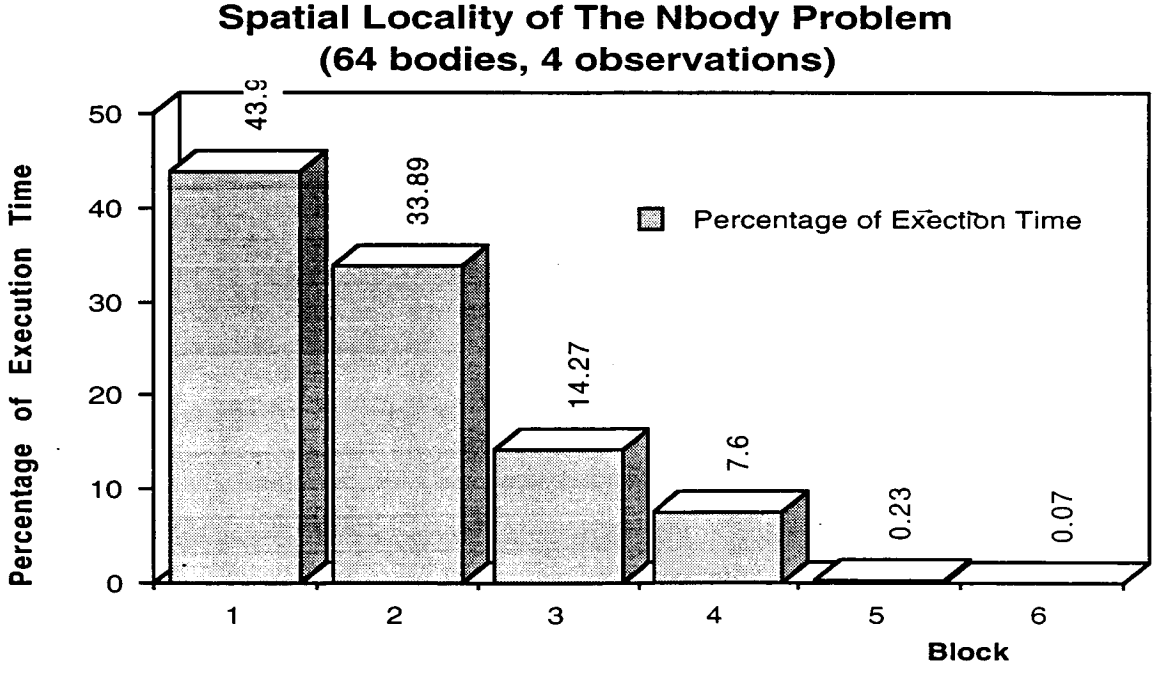
Tarek El-Ghazawi, Principal Investigator (George Washington University), Abdullah Meajil and Armagan Ozkaya, Research Assistants

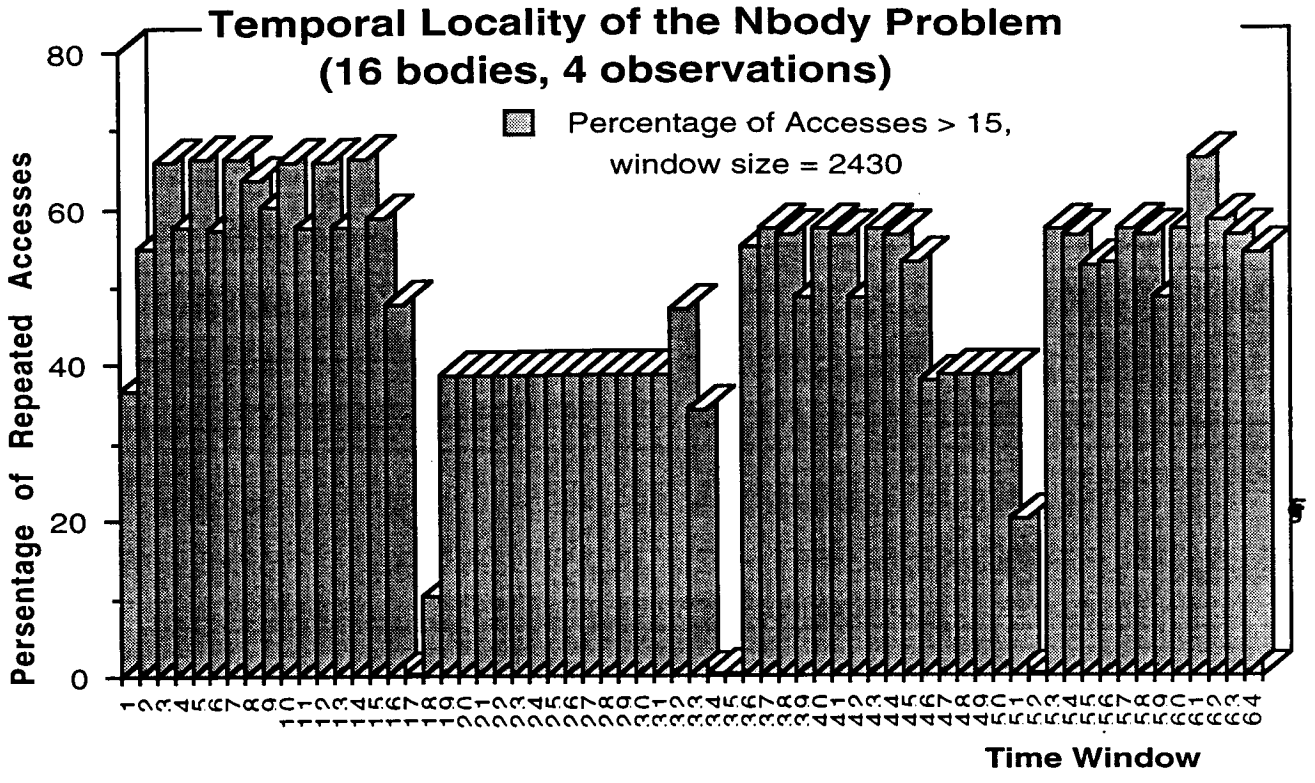
Workload characterization has been addressed in many different contexts and perceived differently in each case. Based on the goals of the underlying study, the parameters of the workload model were selected. At one end, proposed parameters represented the work being done by a specific machine. At the other end, workload parameters represented the work generated by an application regardless of the underlying system characteristics. The latter is of special interest to us and we refer to it as the architecture-invariant workload characterization. This architecture-invariant characterization allows the understanding of the inherent characteristics of the workload, using attributes selected to represent how target architectures will be potentially stressed by such an application. Such characterization can help in parallel benchmark design, experimental analysis of architectures based on benchmark measurements, identification of resource requirements, performance prediction and performance tuning. To do so, the workload model/parameters can be selected to characterize the application behavior with respect to attributes such as average degree of task parallelism, parallelism variability, average degree of data parallelism, data parallelism variability, spatial and temporal locality of references, and dynamic instruction count and mix.

Our approach employs the experimental analysis tool, SITA, developed at McGill University, after augmenting it with additional utilities to measure locality of reference workload parameters. SITA uses a trace of the application under study and performs extensive data dependency analysis. The trace data acquired includes dynamic instruction profiling, procedure and system calls profiling, and memory address tracing. The analysis provides instruction mixes, higher-level memory access profile, and program dependency information.

A detailed characterization of the ESS workload is being performed with special emphasis on quantitative evaluation of resource requirements and scaling properties. Results from this work are being used in conjunction with NSF in the Joint NSF-NASA Initiative in Evaluation (JNNIE). These combined studies apply a common set of metrics to applications of importance both to NASA and NSF. They are being used to analyze the degree of workload specificity of the ESS test suite with respect to the more general NSF science domain. Some initial observations from our current workload characterization effort are given here. Those initial results focused on two ESS applications, N-body and the Backpropagation Neural Networks. Based on these results we have observed that in these ESS applications temporal locality is higher for data, and locality properties are invariant to problem size. In backpropagation spatial locality is much higher and the independence on working on the training sets resulted in a linear scalability. Examples of these results are shown on the next page.


Scalability of the Nbody Problem
(32 observations)
Average Instruction Parallelism
500
400
300
200
100
0
Tree Code
Simple Code
Number of Bodies
0
20
40
60
80
100
120
140


Spatial Locality of The Nbody Problem
(64 bodies, 4 observations)
Percentage of Execution Time
50
40
30
20
10
0
43.9
33.89
14.27
7.6
0.23
0.07
Percentage of Execution Time
1
2
3
4
5
6
Block



JNNIE: Evaluation of Scalable Parallel Processing Architecture Performance for Major Scientific and Engineering Applications

Steven Hotovy (Cornell Theory Center) in collaboration with Thomas Sterling

Major scientific and engineering applications, at one time beyond the reach of even the fastest supercomputers, are now at the threshold of being realized as peak performance of scalable parallel processing (SPP) architectures surpass peak performance of 100 GigaFLOPS and main memory capacities between 10 and 100 GigaBytes. In spite of this extraordinary potential, it is an urgent and open question as to whether SPP systems can be of general and effective use or whether they will remain difficult-to-program special purpose devices of limited applicability for real-world scientific and engineering work.

The United States Federal HPCC program has established a joint initiative between the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA) to resolve this question. Drawing on the combined resources, computational testbeds, talents, and applications for the communities the two agencies represent, the Joint NSF-NASA Initiative in Evaluation (JNNIE) is engaged in a one and half year task to develop detailed assessments of the major commercial SPP systems. The critical areas of evaluation are effectiveness of execution, scalability, and ease-of-use. Each agency is providing approximately 10 significant application programs that reflect important computationally demanding problems within their respective communities. The final result of this major study will be a definitive characterization of contemporary SPPs in terms of their performance, the factors contributing to their behavior, the potential for scaling to TeraFLOPS performance, and the ease with which their capabilities can be brought to bear on existing problems. This report presents the objectives, methodology, and early results of the JNNIE study.

The SPP family of architectures is based on the premise that a new generation of high performance computers can be synthesized from very large arrays of workstation-class processors integrated by high speed interconnect networks. While a limited number of successful application ports have been documented, a number of factors may combine to make this approach generally infeasible. Specifically, the added complexity of managing parallel activities and resources over a space that is physically distributed incurs performance degradation through latency, contention, overhead, and starvation. Together, these influences can inhibit effective computation, limiting performance to unacceptable levels. Compounding these problems is that microprocessor architectures employing cache mechanisms for latency hiding are based on program behavior not necessarily typical of large scientific problems but, as would be anticipated, more appropriate for applications to be run on personal workstations. In particular, the assumptions of program temporal and spatial locality from which cache structures were derived appear less than valid for some important scientific problems. But with cache miss penalties approaching 50 cycles for the highest speed microprocessors and latencies far more extreme for distributed memories, poor locality can easily result in sustained performance an order of magnitude below published peak performance. The paper presents examples of this type of behavior to illustrate the basis for concern. The paper also shows cases of overhead costs resulting from the need to expend resources in their own management.

Execution time, the first of the JNNIE focus areas, is characterized as an execution budget; the time division of processor cycles and the way they are used. Scaling is the sensitivity of performance to perturbations in resource and problem size. This leads to an understanding of bounding limits to performance gain through simple augmentation of resources. Usability, the second JNNIE focus area, is difficult to quantize but is critical to determining the ultimate success of SPP systems as general high performance computing systems. This paper will present for the first time the detailed results of a survey conducted during the Fall of 1993 among computational scientists for both the NSF and NASA. The survey provides an important examination of the user community's perceptions of the factors that make the current generation SPP systems both easy and hard to harness.

The execution budget characterization used by JNNIE is a new hierarchical and recursive delineation of processor cycle usage. The paper presents this methodology in detail with examples. At each level (node of an abstract execution tree) the cycles are demarcated in terms of useful work, overhead work, and idle (or starvation) cycles. At a given level, useful work is defined as operations being performed that lead directly to the computational goal. One way to consider it is the work that would be performed on a uniprocessor to achieve the same end. Overhead is defined as work that is performed to manage parallelism. It is found that if one looks at the top level, overhead itself can be divided at a more detailed level into these three parts as well. This simple abstraction provides guidance in identifying salient parameters of system operation and defining metrics for experimentation.

One of the most important aspects of the JNNIE study is its base of applications. The applications selected cover a wide range of physical and computational domains, and because of this diversity, it is likely that most types of computational behaviors will be investigated. The conclusions drawn from these studies, then, will have broad application. All source codes will be made public so that the experiments can be repeated. However, this is not to be construed as a benchmarking activity with its own implications and limitations. Efforts will be made to provide an architecture independent characterization of the applications themselves. Using a variety of means to be discussed in the paper, certain properties intrinsic to the applications will be quantized to estimate the "algorithm scaling opportunity", that is, the maximum scalability of an architecture running a particular algorithm embedded within a larger problem. An example will be given in the paper.

This joint study involving two major agencies is likely to produce some of the most detailed and complete data in the near future for evaluating scalable parallel processing systems. These results are important to applications programmers, system software designers, computer architects, system vendors, and the engineering/scientific community at large. This will be the first paper to begin to reveal the results of this important study.

The following report discusses work performed on a subcontract with the San Diego Supercomputer Center to evaluate the performance and usability of scalable parallel computers. The UCLA Atmospheric General Circulation Model was selected as a representative application for evaluation on local scalable parallel computers.

JNNIE Phase 1: The Parallel UCLA AGCM

Robert H. Leary (San Diego Supercomputer Center)

1.0 Introduction and Summary

This report provides a background summary of the scientific and computational aspects of the parallel distributed memory MIMD implementation of the UCLA Atmospheric General Circulation Model (AGCM) originally developed by Arakawa [1,2], as well as micro-and macroperformance data for its implementation on an 8-processor cluster of DEC Alpha workstations interconnected by an FDDI/Gigaswitch network at the San Diego Supercomputer Center (SDSC). A performance model for a particular class of domain decompositions is developed and fit to performance data on the Alpha cluster to obtain estimates of the parallel inefficiencies due to load imbalance and communication overheads. For this class, each type of parallel inefficiency is shown to increase approximately linearly with the number of processors, with communication accounting for approximately 60% and load imbalance 40% of the total parallel inefficiency.

2.0 Background Summary

The Parallel UCLA AGCM is a collaborative effort of researchers at the Lawrence Livermore National Laboratory and the Department of Atmospheric Sciences at UCLA. The model is being used in a comprehensive series of studies on the dynamics of the atmosphere and of the coupled atmosphere-ocean system, under the direction of C. Roberto Mechoso. For a more complete description, see [3], which formed the basis for this background summary.

The Parallel UCLA AGCM has been selected as one of the codes to be evaluated at SDSC for the JNNIE project. The target parallel platforms for the evaluation are the above-mentioned DEC Alpha cluster, as well as the 400-node Intel Paragon supercomputer at SDSC. The Parallel UCLA AGCM is particularly well suited for evaluation on multiple MIMD platforms as it has been specifically engineered for portability in such environments. For the Phase I portion of the program, the evaluation was performed on the Alpha cluster.

2.1 The Scientific Challenge

The growth in human population and level of industrial activity since the Industrial Revolution has led to a large-scale and largely uncontrolled experiment on global environmental systems. Atmospheric concentrations of greenhouse gases such as carbon dioxide, methane, nitrous oxide, and chloroflourocarbons are measurably rising. These greenhouse gases, along with water vapor, are major factors in the radiation budget of the atmosphere and hence in the Earth's climate control system.

While it is likely that continued increases in the concentration of greenhouse gases over the next several decades will eventually affect global climate, we are currently far from being able to predict specifics of magnitude, timing, and geographic pattern of such changes. The changes themselves, as well as attempts to alter patterns of energy production and consumption to counteract a possible runaway global temperature rise, clearly will have severe social and economic costs. Thus there is a pressing need to improve the scientific basis for predicting the response of the Earth's environmental systems to both natural and human-induced perturbations.

2.2 The Computational Challenge

Uncertainties in the climate historical record and the magnitude, mechanisms, and effects of the possible perturbing influences, as well as the impracticality of comprehensive laboratory climate experiments, have created a vital role for computer-based climate simulation experiments. Such experiments potentially will allow quantitative tests of specific climate models against observational data sets, and thus provide a basis for increased understanding of the Earth's climate system and for predicting future changes.

However, the complexity of the physical processes that determine climate, as well as the need to accommodate a wide range of spatial and temporal scales, necessarily result in complex models with enormous computational requirements. Even typical current atmospheric models with limited chemical reactions, simplified oceans, and neglect of marine and terrestrial biosystems, overtax conventional vector supercomputers. Future models with increased resolution and better representation of underlying processes will increase the gap between required and available computational resources.

Clearly both faster algorithms and computers are necessary to attempt such advanced simulations. Parallel processing has the potential for providing the required performance increase, but the basic algorithms must be implemented in parallel form. The Parallel UCLA AGCM is part of a long range plan to develop a comprehensive climate systems modeling capability.

2.3 Parallelization and Portability Strategy

2.3.1 The UCLA AGCM

The underlying UCLA AGCM, as is typical of most AGCMs, consists of two parts. The first part is the solution of the equations of hydrodynamics derived from the Navier-Stokes equations suitably adapted to the regimes encountered by the atmosphere. The equations are implemented based on a grid point finite difference model, with a horizontal Arakawa C-mesh in latitude/longitude and vertical staggering of velocity and thermodynamic variables. The vertical differencing scheme uses the hydrostatic approximation to reduce the three-dimensional equations to sets of coupled two-dimensional equations. Time differencing is explicit, with time steps chosen to be consistent with the Courant-Friedrich-Levy (CFL) stability condition at low latitudes. This step size can be chosen large enough to violate the CFL condition near polar cells through the use of a Fourier filtering scheme that dampens unstable modes in such regions.

The second part of the UCLA AGCM consists of a collection of parameterized physical process models collectively referred to as the "column physics" modules, as they are typically computed for a vertical column over a given horizontal mesh cell. These include energy, momentum, and moisture transport (long and short wave radiation, cumulus convection, surface fluxes), cloud instability, simplified ozone photochemistry, and ground temperature evolution calculations, among others. The column physics time step is determined essentially by the vertical structure of the model and does not need to be decreased with increased horizontal resolution.

2.3.2 Parallel Decomposition Strategy

The Parallel UCLA AGCM is designed for a distributed memory (message-passing) MIMD computing environment. The basic domain decomposition is two-dimensional over latitude and longitude, with each subdomain thus consisting of a number of contiguous vertical columns. This choice is based on the strong coupling of the column physics processes within cells along the column, as well as the fact that the number of meshpoints along the vertical direction is usually small. Column physics coupling between horizontal domains is slight and imposes very small communication requirements.

Each subdomain is logically rectangular with north-south latitudinal and east-west longitudinal borders, with subdomains assigned to processors in a deterministic manner. Each subdomain consists of an inner rectangle of cells to be calculated by the assigned processor, and a bordering frame of cells to which data will be

passed via messages from the neighboring subdomains. Due to the explicit time differencing, communication of any given quantity for the difference approximation of horizontal derivatives in the hydrodynamic calculations is necessary only once per time step, and involves only messages that need to be sent to each of a processor's four nearest neighbors. Careful attention has been given to uniform structuring of the subdomains, so for example, the equation differencing routines require no knowledge of the physical location of the subdomains. Moreover, the same code can be used for multiple and single subdomain decompositions, thus allowing execution on traditional uniprocessors for debugging and comparison.

In addition to the simple point-to-point communication between subdomains required for the solution of the difference equations, the filtering scheme (involving longitudinal convolution of vectors) introduces the requirement for substantial global communication among the subdomains. Moreover, the filters involve more computational work at higher latitudes and thus introduce a load balancing problem. Together, the global communication requirements and the load balancing problem can pose a serious challenge to achieving good parallel efficiency.

2.3.3 Portability Considerations

The Parallel UCLA AGCM is a large code (about 17,000 lines) implemented in Fortran 77 that is designed for use on a wide variety of MIMD platforms. The main portability considerations involve dynamic memory management and message passing, both of which are generally platform-dependent.

Dynamic memory management, which is characterized by the allocation and deallocation of memory at run time rather than compile time, allows more efficient memory usage in addition to making it much easier to resize domains dynamically for load balancing. Unlike C, Fortran language dynamic memory management constructs (when available) are non-standard. The Parallel UCLA AGCM handles this situation by the use of the M4 macro preprocessor to expand memory allocation macros to machine-dependent source code.

A similar strategy is used to handle message passing. Fixed message passing macros and conditional CPP precompile directives appear in the source code, which are then preprocessed by CPP to generate specific message passing calls prior to compilation. Various target message passing systems are supported, including PVM 2.4, PVM 3.2, and P4 as well as the native message passing systems for the Intel Paragon (NX) and the TMC CM-5 (CMMD 2.0).

3.0 DEC Alpha Cluster Implementation

Previously, the Parallel UCLA AGCM was successfully implemented on a variety of workstation clusters (including Sun and IBM RS6000 clusters connected by ethernet) and MIMD parallel computers (including the BBN-TC2000). As part of the ongoing Parallel UCLA AGCM development project, a new implementation was initiated on a cluster of eight DEC Alpha processors at SDSC connected by an FDDI/Gigaswitch network. In principle, this network supports a point to point maximum bandwidth of 12.5 Mbytes/sec with maximum aggregate bandwidth of 450 Mbytes/sec. Each Alpha processor has a maximum floating point speed of 133 Mflops, for an aggregate 1.07 Gflop capability.

The original message passing system selected for implementation was PVM 3.2.1. Subsequently, network timing tests revealed a possible communication inefficiency due to a bug in the message packetization algorithm. The bug has been corrected in PVM 3.2.5 and later versions, and the timing results presented below are based on a second implementation with PVM 3.2.6.

3.1 Network Performance

A timing study was undertaken to determine the basic performance of the FDDI/Gigaswitch network in terms of latency and point-to-point bandwidth. The TIMING/TIMING_SLAVE pair of programs distributed with

PVM 3.2 for this purpose was run. This test consists of a master program running on one processor sending a message of a given length to a slave program on a second processor, and the slave program returning a short acknowledgment to the master. The total time from the master programming initiating the send to the master program receiving the acknowledgment is measured. Thus, with the usual linear timing model

$$\text{message time} = a + b\,n$$

to send a message of length n bytes, where a is a latency and 1/b a bandwidth, the total time to send a message of length n bytes and receive a short (essentially zero-length) acknowledgment is 2a + bn. Figure 1 shows the measured times for a series of messages ranging in size from n = 0 to n = 52,000 bytes, along with a linear least squares fit to the data. The linear model is seen to provide an excellent fit to the data with latency a = 537 msec and inverse bandwidth b = 0.1872 msec/byte (or equivalently, a bandwidth 1/b of 5.2 Mbyte per second).

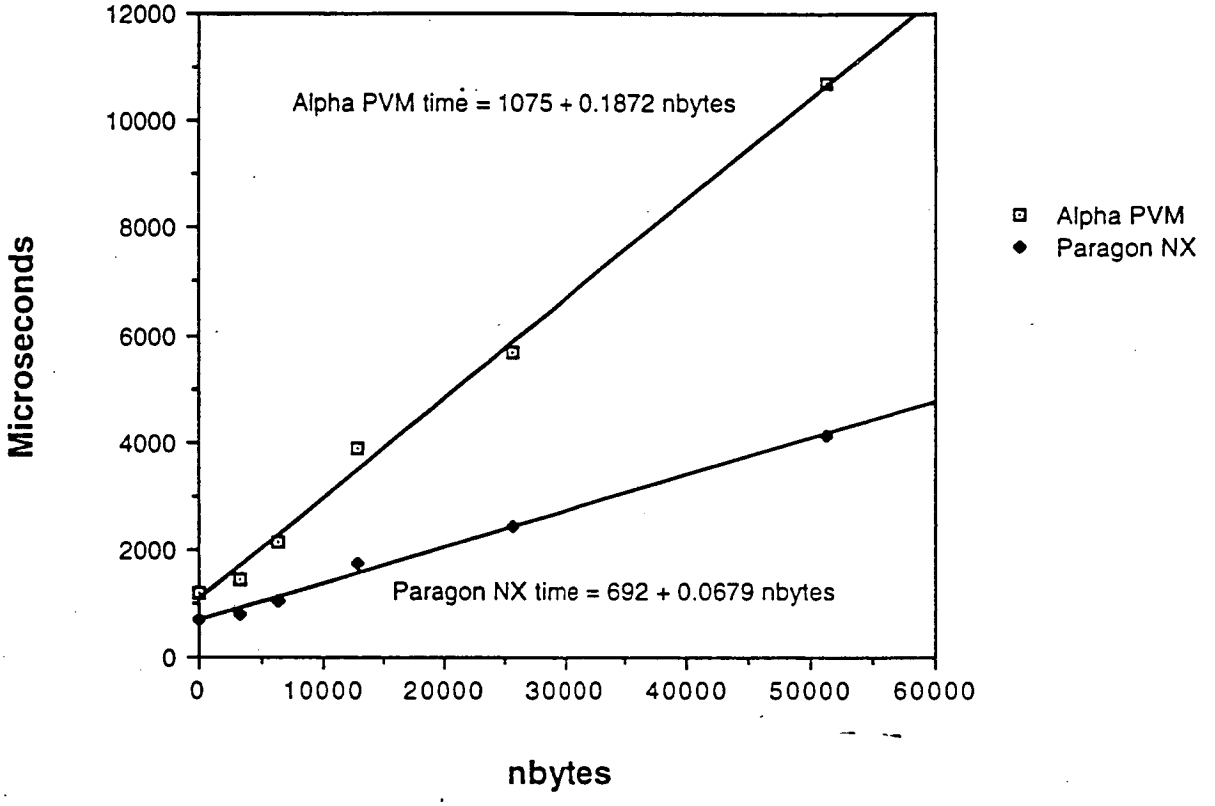


Figure 1. Alpha PVM 3.2.6 vs Paragon NX Network Performance

By way of comparison, Figure 1 also shows a similar plot for the Intel Paragon for a functionally identical master and slave program pair communicating via the native NX message passing system (OS version R1.1.3). Here, as expected on this more closely coupled system, both the latency (346 msec) and the bandwidth (15.2 Mbytes/sec) are significantly improved relative to the Alpha cluster.

3.2 AGCM Performance on the Alpha Cluster

The Parallel UCLA AGCM code was implemented on the SDSC Alpha cluster and tested on a data set over a global 45 x 72 x 9 grid (4 degree latitude and 5 degree longitude spacing with 9 vertical levels) . Runs were made for a total of 3 simulated days, with a minimum of I/O (only reads of two initial history and boundary condition files, and a final write of the updated history file were performed) for all values of p (number of

processors) from 1 to 8. For these runs the p processors were mapped onto a logical 1 x p decomposition of the grid in longitude and latitude hence each subdomain consisted of a strip encircling the globe between two given latitudes. On each run, the total CPU times (user + system) t_1, t_2, ..., t_p were measured on the p processors and the maximum value t_{max} selected to represent the overall system performance (this value represents wall clock time to completion on a dedicated system). Since the computational load (and hence the t_i values) can vary significantly from subdomain to subdomain, an attempt was made to quantify this load imbalance in the form of an imbalance fraction

$$\text{imbalance} = (t_{max} - t_{av})/t_{av}$$

$$\text{where } t_{av} = (t_1 + t_2 + \ldots + t_p)/p.$$

The various measured values are listed in Table 1 for runs with only hydrodynamics ("hydrodynamics" column) enabled and runs with both hydrodynamics and column physics enabled ("total" column). Also listed are times for "column physics" obtained by subtracting on a processor by processor basis the "hydrodynamics" time from the "total" time and taking the maximum over all processors. Note that processor requiring the maximum time to complete the hydrodynamics portion is generally not the same as the processor requiring the maximum time on column physics - hence the column physics and hydrodynamics times may sum to more than the total time. Also listed is a speedup with a parenthetical parallel efficiency. The speedup is calculated as the ratio T/t_{max} of the total time T for one processor to the time t_{max} for p processors operating in parallel, and the parallel efficiency is the ratio $T/(pt_{max})$ of the observed speedup to p, the linear speedup for p processors.

Table 1. ALPHA CLUSTER TIMINGS
(secs/simulated day)
4 deg lat. X 5 deg long., 9 vertical levels
with minimal I/O (history written once at end of 3-day run)

# of processors	hydrodynamics	column physics	total	imbalance	speedup
1p	622(11.1Mflops)	676(12.5)	1298(11.9)	0.000	1.0(1.00)
2p (IX2)	325 (21.2)	359 (23.5)	684 (22.5)	0.007	1.9 (0.95)
3p (IX3)	231 (29.9)	262 (32.2)	485 (31.7)	0.042	2.7 (0.89)
4p (IX4)	179 (38.6)	203 (41.6)	380 (40.4)	0.048	3.4 (0.85)
5p (IX5)	156 (44.5)	175 (48.2)	326 (47.1)	0.087	4.0 (0.80)
6p (IX6)	141 (49.2)	140 (60.2)	281 (54.7)	0.083	4.6 (0.77)
7p (IX7)	129 (53.8)	123 (68.5)	252 (61.0)	0.106	5.2 (0.74)
8p (IX8)	118 (58.9)	116 (72.8)	229 (66.7)	0.113	5.7 (0.71)

The times in Table 1 are also converted to an equivalent Mflop rate. This was done based on the total number of flops required on a single processor of a Cray Y-MP for execution of the code with p=l, as measured by the Cray hardware performance monitor. Here the overall Cray Y-MP time is 164 cpu seconds per simulated day (100 seconds column physics, 64 seconds hydrodynamics). The corresponding Mflop rates as measured by the Cray hardware performance monitor are 94 Mflops total, 80 Mflops column physics, and 109 Mflops hydrodynamics. Thus the overall 8-processor Alpha cluster speed of 67 Mflops represents 72% of the performance of a single Cray Y-MP processor on this problem.

4.0 A Simple Performance Model

We present a simple performance model of the cluster for the 1 x p decomposition. Here we assume that primary sources of parallel inefficiency are communication overhead and load imbalance. The time to completion t_i for processor i in a p-processor parallel computation then takes the form

$$t_i = (T/p)\ (1+ I_i(p)) + C(p,i)$$

where T is the time to completion for a single processor employed on the entire job, $I_i(p)$ is a load imbalance representing fractional deviation of the computational load assigned to processor i from the average load over all processors, and C(p,i) is the communications overhead for processor i. (Note that we are making the implicit assumption that the computation has a negligible serial component).

We argue below and present direct timing and message counting evidence that C(p,i) = C for i = 2, 3, ..., p-1 and C(p,i) = C/2 for i = 1 and i = p, where C is a constant independent of p and i. Thus the observed maximum time is given by

$$t_{max} = (T/p)\ (1 + I_{max}(p)) + C$$

assuming the maximum occurs on a processor assigned to an interior latitude strip (this is consistently observed to be the case in practice). Then the parallel efficiency is given by

$$T/(pt_{max}) = 1/(1 + I_{max}(p) + Cp/T)$$

Assuming that the communication overhead C is small relative to T (later we develop the estimate C=29.8 secs/day while T=1298 secs/day) and that $I_{max}(p)$ is small, the right hand side can be expanded in a Taylor series. Keeping only first order terms, we have

$$1- T/(pt_{max}) = I_{max}(p) + Cp/T$$

which indicates that the parallel inefficiency $1 - T/(pt_{max})$ is equal to the sum of the maximum imbalance fraction $I_{max}(p)$ and a communication overhead term which is linear in the number of processors. Moreover, as seen from the plot in Figure 2, the observed function $I_{max}(p)$ is well approximated by a linear function $I_{max}(p) = a\ p$, where a = 0.0146. The overall parallel inefficiency is also well approximated by $1 - T/(pt_{max}) = b\ p$, with b = 0.0376 . This results in the estimate C/T = b - a = .0230, or C = 29.8 secs/day. Thus for p > 2, approximately 40% of the parallel inefficiency is due to load imbalance, and 60% is due to communications overhead.

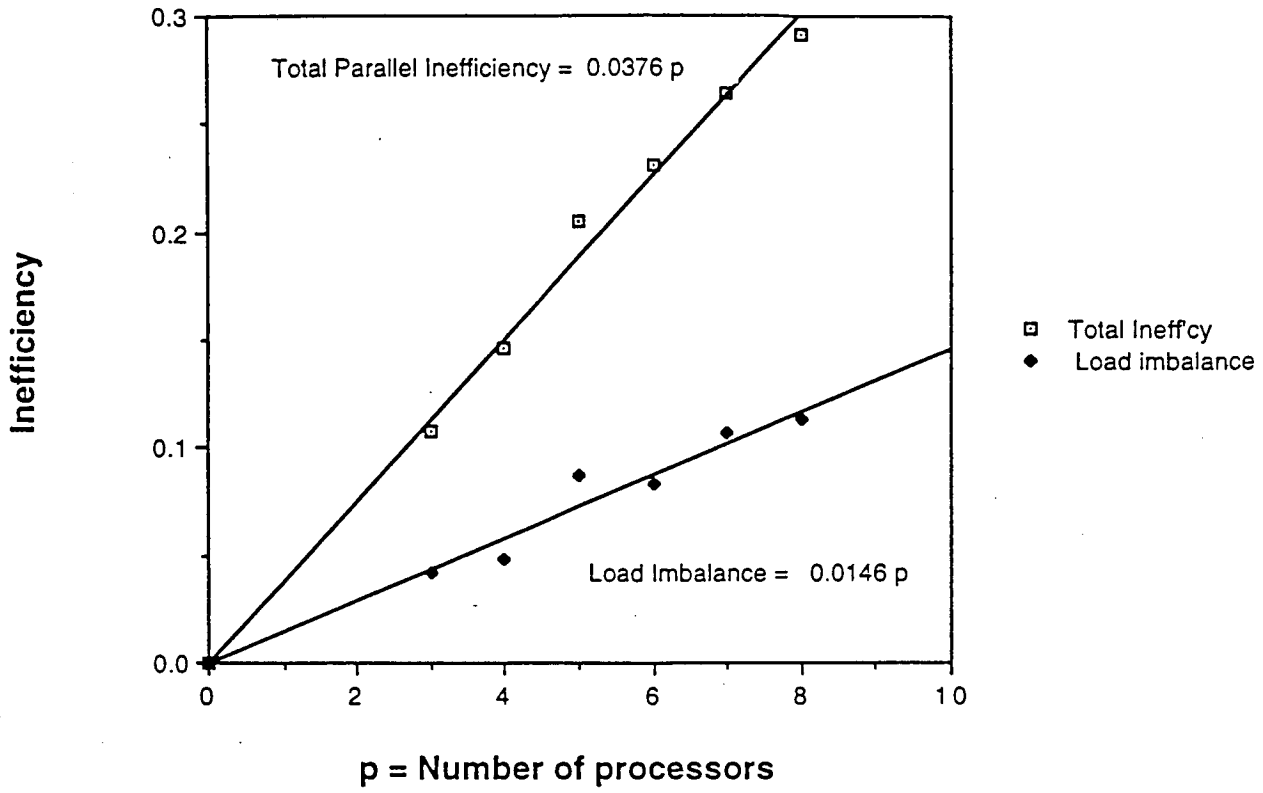


Figure 2. Parallel Inefficiency

For the 1 by p decomposition, we expect communication overhead C(p,i) to be approximately constant for the interior strips i = 2, 3, .., p-l and independent of p for p > 2. This follows from the fact that each such strip has two borders of identical size which is independent of p. Similarly, strips 1 and p containing the poles have only one border and thus should require half the communications overhead of the interior strips. This is supported by the sample data in Table 2, which lists the number of messages sent and the associated system CPU time for a 1 x p decomposition for p = 2, 4, 6, and 8. The number of messages sent in all cases approximately 54,000 for interior strips and 27,000 for the polar strips. Similarly, the total system time averages about 44 seconds in the interior and 22 seconds at the poles. (Note that the communication overhead consists of both system and user time. The system time is directly available from code instrumentation; here we make the assumption that the total communication overhead is proportional to system time).

Table 2

COMMUNICATION PERFORMANCE
(totals for 3 simulated days)

	processor number	message sends	system time (secs)
p=2	1	26969	22.8
	2	26897	22.1
p=4	1	26897	22.6
	2	53566	42.1
	3	53563	43.5
	4	26822	23.3
p=6	1	26881	22.5
	2	53550	41.4
	3	53535	43.6
	4	53535	41.6
	5	53535	46.7
	6	26794	24.1
p=8	1	26865	22.0
	2	53534	43.3
	3	53531	44.4
	4	53531	44.6
	5	53527	44.0
	6	53527	43.7
	7	53527	43.8
	8	26786	22.4

5.0 Conclusion

The Parallel UCLA AGCM has been implemented at SDSC on a cluster of eight DEC Alpha processors connected via an FDDI/Gigaswitch network. Overall performance is comparable to three fourths of a single processor of a Cray Y-MP, and parallel efficiency is 0.71 for all eight processors using a 1 x p decomposition. A performance model has been developed which predicts the parallel inefficiency for this decomposition as the sum of a load imbalance and a communication term, both of which are approximately linear in the number of processors. The model fits timing tests very well and leads to the conclusion that about 40% of the parallel inefficiency is due to load imbalance and 60% to communication overhead.

6.0 References and Acknowledgement

[l] A. Arakawa and V. Lamb, Methods in Comp. Phys. 17 (1977) 173-265.

[2] A. Arakawa and V. Lamb, Mon. Weath. Rev., 109 (1981) 18-36.

[3] M. F. Wehner, J. J. Ambrosiano, J. C. Brown, W. P. Dannevik, P. G. Eltgroth, A. A. Mirin and J. D. Farrara, C. C Ma, C. R. Mechoso, J. R. Spahr, "Toward a High Performance Distributed Memory Climate Model", Second International Symposium on High Performance Distributed Computing (HPDC-2, 1992), Spokane, WA, IEEE Computer Society, 16-25.

J. D. Farrara of UCLA performed the initial implementation of the Parallel UCLA AGCM on the Alpha cluster.

HPCC/ESS GOPS Workstation System Software Environment

Donald Becker

Biographical Sketch

Donald Becker joined CESDIS in April of 1994. Previously Mr. Becker worked at the Supercomputing Research Center. There he wrote a substantial portion of the low-level LINUX networking code, including over a dozen device drivers for network adaptors.

He also 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. Work was done in C and Lisp on Suns, a Cray-2, CM-2 and a Convex.

Between July 1987 and January 1990, Mr. Becker was employed as a senior engineer in the Advanced Technology Department at Harris Corporation.

He performed R & D work on the Concert Multiprocessor, including the design and construction of two performance monitoring boards, maintained and extended

the Concert C compiler (based on PCC) and libraries, and wrote network software.

The goal of the Earth and Space Science Project in the NASA High Performance Computing and Communications Program (HPCC/ESS) 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. Two broad areas of investigation include:

- System software and tools for scalable parallel processing environments to facilitate programming and enhance efficiency, and
- Scalable parallel processing testbed acquisition and evaluation towards TeraFLOPS performance.

This task will provide ESS with a workstation approaching a billion operations/second--a giga-ops/sec (GOPS) performance comprising multiple processors and a Unix-like operating system to integrate its resources. The objective of this project is to advance the state of the art in workstations for high performance computing through three complementing enhancements:

1. Extend the performance to GOPS capability,
2. Substantially expand the disk capacity and disk bandwidth, and
3. Employ MIMD parallelism by integrating multiple low cost PC-grade processors.

An important constraint on workstation implementation is cost, and this project addresses this through the use of commodity, off-the-shelf subsystems to keep hardware cost below $50,000.

The focus of the project is on the software environment which will be based on the Linux operating system, a Unix-like environment that is open, free, and available with complete sources. Linux will be initially extended to support PVM for a loosely coupled system, followed by a shared virtual paging system to support a global shared reference space. Workstation applications will include the NASA-sponsored FAST for visualization, the NASA-sponsored AIMS for performance monitoring, and a few scientific applications. The workstation will be heavily instrumented for detailed performance analysis and workload profiling. The GOPS workstation will be a target for an advanced architecture independent programming methodology.

Donald Becker joined CESDIS in mid-April 1994 to implement this task. Progress through June 1994 includes the following:

- The first workstation employing Linux 1.0 has been brought online for software development and is supported by a Sparc 10 workstation for filer server and ftp site.

Other projects included writing a user-level NFS server, working on the software specification and design of a transportable SHF terminal, and working on the hardware design and language for a static dataflow processor.

Mr. Becker received his B.S. in electrical engineering from the Massachusetts Institute of Technology.

- The first multiple Linux workstation has been defined and is in procurement. This software development station will provide the necessary environment for development of all system software to run on the 16 processor prototype when completed in the fourth quarter 1994. The software development station is a four processor cluster of x486 commodity CPU boards integrated by two Ethernet networks and one 100 Mbps network.

- The first version of the PCMCIA "point enablers" software has been developed to support 14.4 kbps modems with this new bus protocol.

- A new driver for a new HP Ethernet card has been written and debugged. It has since been integrated into the new Linux development kernel release.

- Patches and bug fixes for other Ethernet drivers have been implemented in response to feedback from a rapidly growing user community.

- A WWW server for CESDIS and for Linux information has been set up and populated with results of GSFC Parallel Linux activities. This installation is heavily hyperlinked for easy access and cross referencing.

(Task 43)

Evaluation of High Performance Fortran

Terrence Pratt

Task Objective

To evaluate the design and implementation of high performance Fortran (HPF) for use in programming ESS applications on massively parallel machines. Target machines include the ESS testbed computer systems under the NASA HPCC/ESS program.

HPF is a new Fortran dialect proposed as a standard across a wide range of hardware platforms including massively parallel systems, conventional supercomputers, and workstations. HPF promises both portability and high performance on these platforms. The language is currently being implemented by several vendors.

Issues to be explored include the expressive power of HPF for ESS algorithms and the performance of HPF coded algorithms on ESS testbed machines compared to other languages and programming techniques.

High Performance Fortran (HPF) provides a set of extensions to the Fortran 90 language that allow the expression of "data parallel" algorithms for the solution of scientific computing problems. The primary innovations in HPF involve features for the distribution and alignment of large arrays on distributed memory parallel computers. It is expected that for many scientific algorithms the ability to correctly distribute and align arrays to match the array referencing patterns within the source code will allow compilers to optimize the object code to achieve high performance.

As HPF compilers emerge during the next year, it is important to evaluate both the expressive power of HPF and the performance of HPF compilers on ESS algorithms and testbed machines. This project began in March 1994, but no major work has yet been undertaken due to the project scientist being involved in CESDIS management.

In order to track the development of both the HPF language and the HPF compilers, we have attended all meetings of the HPF Forum, led by Ken Kennedy of Rice University. The HPF Forum is the group responsible for the definition of the language. Four meetings are planned for 1994. Meetings attended during this reporting period were April 6-8 and June 1-3.

RESEARCH ACTIVITIES

Peer Reviewed Projects

In 1988 a call for proposals was drafted under the direction of Interim Director John Hopcroft. Eighty-six proposals were received and reviewed. Four projects were funded for an initial 3-year period with an additional 2-year option. Only John Reif at Duke University was funded for the entire five years. The final Duke project report appears on page 85 of this section.

In December 1992 CESDIS called for proposals in parallel computing to be sponsored with funding ($50K per year for two to three years) from the NASA High Performance Computing and Communications Program. Consistent with the primary CESDIS mission, the projects funded form a national group of researchers interested in collaborating to attack key problem areas in parallel computing that affect NASA's efforts to collect, manage, store, and process massive data sets. Ten proposals in three major areas were selected for funding.

Summary reports of first year activities among the following 10 principal investigators are included in this section.

(Task 14)

UNIVERSITY OF TEXAS AT ARLINGTON

Parallel Knowledge Discovery from Large Complex Databases

Investigators
Diane J. Cook, Lawrence B. Holder
Department of Computer Science Engineering

Task Objective

The goal of this project is to design and implement a system that takes raw data as input and efficiently discovers interesting concepts that can target areas for further investigation and can be used to compress the data. Our approach will provide an intelligent parallel data analysis system.

This effort will build upon two existing data discovery systems: the SUBDUE system used at the University of Texas at Arlington, and NASA's AutoClass system. AutoClass has been used to discover concepts in several large databases containing real or discrete valued data. Subdue, on the other hand, has been used to find interesting and repetitive structure in the data. Although both systems have been successful in a variety of domains, they are hampered by the computational complexity of the discovery task. The size and complexity of the databases expected from the Earth and Space Science (ESS) program will demand processing capabilities found in parallel machines.

We propose to combine the two approaches to concept discovery and speed up the discovery process by developing a parallel implementation of the systems. We will combine the two discovery systems by using Subdue to compress the data fed to AutoClass, and letting Subdue evaluate the interesting structures in the classes generated by AutoClass. The parallel implementation of the resulting AutoClass/Subdue system will be run on the Connection Machine 5, and will be tested for speedup in a number of large databases used by ESS.

This proposed effort will benefit both our group and the Automatic Classification group at NASA Ames. Our group will be able to make use of the code and the insights provided by researchers at NASA Ames. NASA's researchers will be able to observe the benefits of incorporating structural discovery into their approach. They will also be able to use the parallel system we develop, and will have access to the results of applying the AutoClass/Subdue code to several large ESS databases.

1.0 Goals Stated in the Proposal

A general goal of the proposed research was to develop a combined discovery system (SUBDUE + AutoClass) to be applied to a variety of Earth and space science databases. Our objective was to improve overall system speed, discovery power, and generality, by

1. Improving the existing SUBDUE discovery system for application to structural Earth and space databases;
2. Using the SUBDUE system as a pre- and post-processor for NASA's AutoClass discovery system; and
3. Speeding up the combined discovery process through improved algorithms and parallel implementations on MIMD machines such as the Connection Machine 5.

2.0 Recent Work and Planned Work

This section organizes the year's work around issues in machine discovery.

2.1 Improved substructure encoding for data compression

One goal of applying machine discovery methods to large complex databases is for the purpose of compressing the data. By re-describing a database in terms of discovered concepts, the amount of information that is necessary to describe the data can be substantially reduced. During the current report period, we have generated a minimum-description-length encoding that more accurately reflects the description length of a substructure and has yielded high amounts of data compression.

The minimum description length principle (MDLP) introduced by Rissanen states that the best theory to describe a set of data is that theory which minimizes the description length of the entire data set. The MDL principle has been used for decision tree induction, image processing, concept learning from relational data, and learning models of non-homogeneous engineering domains.

We demonstrate how the minimum description length principle can be used to discover substructures in complex data. In particular, a substructure is evaluated based on how well it can compress the entire dataset using the minimum description length. We define the minimum description length of a graph to be the number of bits necessary to completely describe the graph.

According to the minimum description length (MDL) principle, the theory that best accounts for a collection of data is the one that minimizes I(S) + I(G–S), where S is the discovered substructure, G is the input graph, I(S) is the number of bits required to encode the discovered substructure, and I(G–S) is the number of bits required to encode the input graph G with respect to S.

The graph connectivity can be represented by an adjacency matrix. Consider a graph that has n vertices, which are numbered 0, 1, .., n-1. An n x n adjacency matrix A can be formed with entry A[i,j] set to 0 or 1. If A[i,j]=0, then there is no connection from vertex i to vertex j. If A[i,j]=1, then there is at least one connection from vertex i to vertex j. Undirected edges are recorded in only one entry of the matrix.

The encoding of the graph consists of the following steps.

1. Determine the number of bits vbits needed to encode the vertex labels of the graph. First, we need lg(v) bits to encode the number of vertices v in the graph, and lg(l(v)) bits to encode the number of bits needed to specify a vertex label, where l(v) is the number of unique vertex labels in the graph. Then, encoding the labels of all v vertices requires v(lg(l(v))) bits. We assume the vertices are specified in the same order they appear in the adjacency matrix. The total number of bits to encode the vertex labels is

$$vbits = lgv + lgl_v + vlgl_v$$

$$= (v+1) + lgl_v + lgv$$

2. Determine the number of bits rbits needed to encode the rows of the adjacency matrix A. Typically, in large graphs, a single vertex has edges to only a small percentage of the vertices in the entire graph. Therefore, a typical row in the adjacency matrix will have much fewer than v 1s, where v is the total number of vertices in the graph. We apply a variant of the coding scheme used by Quinlan and Rivest to encode bit strings with length n consisting of k 1s and (n-k) 0s, where k << (n-k). In our case, row i ($1 \le i \le v$) can be represented as a bit string of length v containing k(i) 1s. If we let b = (max over i) k(i), then the ith row of the adjacency matrix can be encoded as follows.

- Encoding the value of k(i) requires lg(b+1) bits.
- Given that only k(i) 1s occur in the row bit string of length v, only C(v,k(i)) strings of 0s and 1s are possible. Since all of these strings have equal probability of occurrence, lg(C(v,k(i))) bits are needed to encode the positions of 1s in row i. The value of v is known from the vertex encoding.

  Finally, we need an additional lg(b+1) bits to encode the number of bits needed to specify the value of k(i) for each row. The total encoding length in bits for the adjacency matrix is

$$rbits = lg(b+1) + \sum_{i=1}^{v} lg(b+1) + lg\binom{v}{k_i}$$

$$= (v+1)\, lg(b+1) \sum_{i=1}^{v} lg\binom{v}{k_i}$$

3. Determine the number of bits (ebits) needed to encode the edges represented by the entries A[i,j]=1 of the adjacency matrix A. The number of bits needed to encode entry A[i,j] is lg(m) + e(i,j)[1 + lg(l(e))], where e(i,j) is the actual number of edges between vertex i and j in the graph, m = (max over i,j)e(i,j), and l(e) is the number of unique edge labels in the graph. The lg(m) bits are needed to encode the number of edges between vertex i and j, and [1 + lg(l(e))] bits are needed per edge to encode the edge label and whether the edge is directed or undirected. In addition to encoding the edges, we need to encode the number of bits lg(m) needed to specify the number of edges per entry and the number of bits lg(l(e)) needed to specify the edge label. The total encoding of the edges is

$$ebits = lgm + lgl_e + \sum_{i=1}^{v}\sum_{j=1}^{v} lgm + e(i,j)[1 + lgl_e]$$

$$= lgm + lgl_e + e(1 + lgl_e) + \sum_{i=1}^{v}\sum_{j=1}^{v} A[i,j]\, lgm$$

$$= e + (e+1)\, lgl_e + (K+1)\, lgm$$

where e is the number of edges in the graph, and K is the number of 1s in the adjacency matrix A.

Both the input graph and discovered substructure can be encoded using the above scheme. After a substructure is discovered, each instance of the substructure in the input graph is replaced by a single node representing the entire substructure. The discovered substructure is represented in I(S) bits, and the graph after the substructure replacement is represented in I(G–S) bits. SUBDUE searches for the substructure S in graph G minimizing I(S)+I(G–S).

Once this encoding scheme was developed, we measured the amount of compression that SUBDUE provides across a variety of databases. We applied the discovery algorithm to databases in the areas of chemical compound analysis, image analysis, CAD circuit analysis, and artificially-generated graphs. The table on the next page shows the description length (DL) of the original graph, the description length of the best substructure discovered by SUBDUE, and the value of compression for each one of these databases. Compression here is defined as $\frac{\text{DL of compressed graph}}{\text{DL of original graph}}$.

As can be seen from this table, SUBDUE was able to reduce the database to slightly larger than 1/4 of its original size in the best case. The average compression value over all of these domains (treating the artificial graphs as one value) is 0.62. The results of this experiment demonstrate that the substructure discovered by SUBDUE can significantly reduce the amount of data needed to represent an input graph. We expect that compressing the graph using combinations of substructures and hierarchies of substructures will realize even greater compression in some databases.

DATABASE	DL (orig)	Opt Threshold	DL (comp)	Compression
Rubber	371.78	0.1	95.20	0.26
Cortisone	355.03	0.3	173.25	0.49
DNA	2427.94	1.0	2211.87	0.91
Pencils	1592.33	1.0	769.18	0.48
CAD - M1	4095.73	0.7	2148.80	0.52
CAD - S1SegDec	1860.14	0.7	1149.29	0.62
CAD - S1DrvBlk	12715.12	0.7	9070.21	0.71
CAD - BlankSub	8606.69	0.7	6204.74	0.72
CAD - And2	427.73	0.1	324.52	0.76
Artificial (avg 96)	1636.25	0.0..1.0	1164.02	0.71

2.2 Improved discovery performance using a fast inexact graph match

SUBDUE's substructure discovery method is computationally very expensive. This is due to the fact that all variations on the best substructures are generated and evaluated. The computationally complexity is also due to the fact that each evaluation step involves many applications of a graph isomorphism test. Because all known graph isomorphism algorithms are exponential in complexity, the graph match can reduce the performance of the discovery system. SUBDUE calculates the amount of actual difference in the two graphs, which adds even more computation to the algorithm.

During this past year, we have generated a fast algorithm for inexact graph match. Our inexact match is based on an algorithm described by Bunke and Allermann. In this approach, the similarity of two graphs is defined as inversely proportional to the minimum cost of the mapping from one graph to the other. The cost of a mapping is computed as the number of transformations required to change one of the graphs into a structure that is isomorphic to the second graph, weighted by the user-defined cost of each transformation.

Given two graphs g1 and g2, and a set of distortion costs, the actual computation of s(g1, g2) can be performed by a tree search procedure. A state in the search tree corresponds to a partial match that maps a subset of the nodes of g1 to a subset of the nodes in g2. Initially, we start with an empty mapping at the root of the search tree. Expanding a state corresponds to adding a pair of nodes, one from g_1 and one from g_2, to the partial mapping constructed so far. A final state in the search tree is a match that maps all nodes of g1 to g2 or to {empty}. As we are eventually interested in the mapping with minimum cost, each state in the search tree gets assigned the cost of the partial mapping that it represents. Thus the goal state to be found by our tree search procedure is the final state with minimum cost among all final states.

Given graphs g1 with n nodes and g2 with m nodes, $m \geq n$, the complexity of the full inexact graph match is $O(nm+1)$. Because this routine is used heavily throughout the discovery and evaluation process, the complexity of the algorithm can significantly degrade the performance of the system.

To improve the performance of the inexact graph match algorithm, we apply a branch-and-bound search to the tree. The cost from the root of the tree to a given node is computed as described above. Nodes are considered for pairings in order from the most heavily connected node to the least connected, as this constrains the remaining match. Because branch-and-bound search guarantees an optimal solution, the search ends as soon as the first complete mapping is found.

In addition, the user can place a limit on the number of search nodes considered (defined as a function of the size of the input graphs). Once the number of nodes expanded in the search tree reaches the defined limit, the search prunes away all choices at each level except the choice with the least cost until a leaf node representing a complete mapping is reached. By defining such a limit, significant speedup can be realized with a decrease in the accuracy of the computed match cost.

Additional experiments were tried with the intent of improving the efficiency of the inexact graph match. The experiments and their results are listed below.

- Probabilistic search (random flip). This technique was shown useful in GSAT, another NP-hard problem. The technique does not work well for inexact graph match. Because the nodes are inter-related through their edges, there is no way to correctly determine which node-pair is most inaccurate. Without any effective local information to guide the random flip, the search process simply jumps randomly over the entire search space.

- Pre-sorting the order of nodes for consideration. Nodes from the first graph are sorted in descending order with respect to the node's connectivity. The intuitive reason is that by pairing nodes with larger connectivity first, more edges are included in the partial mapping so an early distinction can be made between a good match and a poor match.

- Search heuristics. Three heuristics were used to direct the search process. The first optimistically assesses future node matches using the similarity of the node labels (not considering the edges). The second heuristic optimistically assesses future node matches using the connectivity of the nodes. The third heuristic combines the previous two heuristics, and is the most effective in guiding the search.

- Consider mappings from smaller graph –> larger graph. Ordering the graphs in this manner reduces the number of non-leaf nodes that need to be considered in the search space.

- Use a hash table to sort queue of open states in linear time. By storing the queue of open nodes as a hash table, newly expanded nodes can be added to the queue in linear time, bypassing the O(log n) computation time required by most sorting algorithms.

- Take advantage of threshold value to prune states with cost larger than threshold value. Results from performing graph match on random graphs shows that the number of states searched increases as the two graphs become increasingly dissimilar. Using the threshold value, however, states with cost greater than the threshold value can be pruned. When a small threshold value is employed, the number of search states is effectively limited to the lower region of the exponential curve.

The results of applying this improved graph match to the discovery algorithm of several databases is shown in the table on the next page. This graph match algorithm is fully implemented and is now being used in the SUBDUE system.

DATABASE	CPU times (sec) Original	New
def	43097.2	41.9
sample.g	122.4	2.5
e1	9882.2	40.2
delaydec	???	250.7
pencil	159.2	17.6

During the current report period, we further improved the efficiency of SUBDUE by including an optional pruning mechanism. Typically, once the description length of an expanding substructure begins to increase, further expansion of the substructure will not yield a smaller description length. As a result, we have added to SUBDUE an optional pruning mechanism that eliminates substructure expansions from consideration when the description lengths for these expansions increases.

2.3 Parallel SUBDUE / AutoClass

The first step to parallelizing the SUBDUE / AutoClass discovery system is to ensure that both systems use the same programming language, and that a language be chosen that is supported by a majority of parallel systems. Although the SUBDUE system uses the C programming language, the current version of AutoClass only runs under Common Lisp. In order to parallelize the combined system, we are first porting AutoClass to C with the help of NASA scientists Will Taylor, John Stutz, and Peter Cheeseman. During the current report period, we have finished this port. We have already received several requests for the C version of AutoClass from Lockheed, NASA Ames, and the Jet Propulsion Laboratory.

Although the parallel implementation of SUBDUE / AutoClass does not yet exist in code, its design has already begun. By the end of the next reporting period, we expect to have a parallel version of SUBDUE and a parallel version of AutoClass running on the Connection Machine 5 at NCSA. One of the students funded by this project, Joe Potts, is spending the summer at NASA Ames to work on this parallelization and to gain ideas for use in combining SUBDUE and AutoClass and for future research in both of these areas.

These parallel implementations will reflect a hybrid parallel-window / distributed-space approach to discovery in which the set of processors will be divided into clusters. Each cluster will store a copy of the entire database, and will search for substructures or concepts of a distinct size. Within each cluster, the database will be divided among members of the cluster to search for substructures or concepts of the given size. The number of processors within each cluster will be dependent on the size of the concept assigned to the particular cluster.

Access to the Connection Machine 5 is made available through a grant from the National Center for Supercomputing Applications. Access to the Intel Hypercube and the Intel Paragon is made available through a grant from the NAS project at NASA Ames.

2.4 Guiding substructure discovery with background knowledge

During the current report period, we have studied methods of controlling the discovery process in SUBDUE using background knowledge. Although the principle of minimum description length is useful for discovering substructures that maximize compression of the data, scientists may realize more benefit from the discovery of substructures that exhibit other domain-specific and domain-independent characteristics.

To make SUBDUE more powerful across a wide variety of domains, we have added the ability to guide the discovery process with background knowledge. Although the minimum description length principle still drives the discovery process, the background knowledge can be used to input a bias toward certain types of substructures. This background knowledge is encoded in the form of rules for evaluating substructures, and can represent domain-independent or domain-dependent rules. Each time a substructure is evaluated, these input rules are used to determine the value of the substructure under consideration. Because only the most-favored substructures are kept and expanded, these rules bias the discovery process of the system.

Each background rule can be assigned a positive, zero, or negative weight, that biases the procedure toward a type of substructure, eliminates the use of the rule, or biases the procedure away from a type of substructure, respectively. The value of a substructure is defined as the description length of the input graph using the substructure multiplied by the weighted value of each background rule from a set of rules *R* applied to the substructure.

$$value(s) = DL(G,s) \times \prod_{r=1}^{|R|} rule_r(s)^{w_r}$$

Two of the domain-independent heuristics that have been incorporated as rules into the SUBDUE system are compactness and coverage. The first rule, compactness, is a generalization of Wertheimer's Factor of Closure, which states that human attention is drawn to closed structures. A closed substructure has at least as many edges as vertices, whereas a non-closed substructure has fewer edges than vertices. Compactness is thus defined as the ratio of the number of edges in the substructure to the number of vertices in the substructure.

The second rule, coverage, measures the fraction of structure in the input graph described by the substructure. The coverage rule is motivated from research in inductive learning and provides that concept descriptions describing more input examples are considered better. Although the MDL principle measures the amount of structure, the coverage rule includes the relevance of this savings with respect to the size of the entire input graph. Coverage is defined as the number of unique vertices and edges in the instances of the substructure divided by the total number of vertices and edges in the input graph.

Domain-dependent rules can also be used to guide the discovery process in a domain where scientists can contribute their expertise. For example, circuit components can be classified according to their passivity. A component is said to be passive if it never delivers a net amount of energy to the outside world. A component which is not passive is said to be active. The active component rule favors substructures containing an active component. Once the active components are selected by SUBDUE, they can be compressed and attention can be focused on the passive components. This rule could also be weighted negatively to exclude substructures with active components. Similarly, the loop analysis rule favors subcircuits containing loops. A loop is defined here as a closed path whose starting vertex is the same as its ending vertex.

The substructure affording the most compression will not always be the most interesting or important substructure in the database. However, the additional background rules can be used to increase the chance of finding interesting substructures in these domains. In the case of the cortisone compound, we might be interested in finding common closed structures such as benzene rings. Therefore, we give a strong weight (8.0) to the compactness background rule and use a match threshold of 0.2 to allow for deviations in the benzene ring instances. In the resulting output, SUBDUE finds the benzene ring as desired.

2.5 SUBDUE / AutoClass combined algorithm

One weakness of the SUBDUE system is the inability to incorporate models of non-structural attribute values. For example, if objects in the domain have a temperature that ranges from 0 to 100, SUBDUE considers the values 50.0001 and 50.0002 as different as 0 and 100. Knowledge about the distribution of attribute

values is necessary to appropriately match this type of data in the substructures. One possibility is to infer the parameters of a normal model for each attribute from the data and use this information to affect the match cost of two graphs. Another possibility is to use SUBDUE as a pre-processor of the structural component of the data in order to construct new non-structural attributes for addition to the set of existing, non-structural attributes. The new set of attributes can then be processed by a nonstructural discovery method, in which the discovered concepts will be biased by the inclusion of structural information.

As we stated in our original proposal, we are pursuing the second possibility by investigating the integration of SUBDUE with the AutoClass system. AutoClass is an unsupervised, non-structural discovery system that identifies a set of classes describing the data. Each attribute is described by a given model (e.g., normal), and each class is described by particular instantiations of the models for each attribute. AutoClass searches for the classification maximizing the conditional probability of the data given the classification.

In the combined algorithm, SUBDUE is first run on the structural component of the data to produce prevalent substructures in the data. For each discovered substructure, the vertices of the substructure (which correspond to individual objects or examples in the non-structural data component) become new attributes whose values reflect the degree with which that particular object or example participates as the vertex in the substructure. For example, if SUBDUE discovered a substructure consisting of two vertices and one edge, then two new, non-structural attributes would be added to the original attributes of each data object. The value of the first new attribute, corresponding to first vertex of the substructure, measures the degree to which the object occurs as this vertex in an instance of the substructure. Likewise, the second new attribute measures a similar value corresponding to the second vertex of the substructure. These values can be determined by the match cost between the substructure and the instance containing the object.

The new data, augmented with this non-structural information about the structural regularities in the data, can now be passed to the AutoClass system. The structural information will bias the classifications preferred by AutoClass towards those consistent with the structural regularities in the data. Implementation of the integrated discovery system will begin during the next report period. We plan to evaluate the integration of SUBDUE and AutoClass by comparing the results of AutoClass alone on data with a weak, non-structural classification and a strong, structural classification. We also plan to compare previous AutoClass results with the classifications obtained with the integrated discovery system using the same data augmented with natural structural information such as temporal and proximate ordering.

3.0 Educational Benefits of This Work

We are currently advising two Ph.D. students and one Master's students who are being directly supported by this NASA grant. One student, Surnjani Djoko, is receiving half-time support. During this report period Surnjani completed the new MDL encoding and developed an improved graph representation for our CAD circuit databases. Surnjani has passed the first two phases of her Ph.D. work (the diagnostic and comprehensive exams) and is working on her dissertation. Her thesis work will focus on combining the SUBDUE and Autoclass discovery systems and developing methods of defining domain-independent rules for input to the discovery systems.

A second student, Joe Potts, is using a portion of the NASA money to visit NASA Ames this summer. During this report period Joe has been working on the conversion of AutoClass from Lisp to C. During his visit to NASA Ames, he will finish this port and will work on the CM 5 and Intel Paragon version of the parallel discovery system. Joe has recently entered the Ph.D. program and passed his diagnostic exam this month.

A third student, Tom Lai, is not funded directly from this grant but is using the facilities supplied by the grant to complete his Master's project. For his Master's project, Tom designed and implemented the fast inexact graph match routine and tested the performance of the algorithm on natural and artificially-generated databases. Tom will be graduating this summer.

The research sponsored by this NASA grant has contributed to the course Dr. Cook taught last spring on Parallel Algorithms for Artificial Intelligence. Because the class was so well received, it will be taught again

next year. The results of the parallel discovery algorithms will be incorporated into the new syllabus so that additional students can benefit from ongoing research and contribute to the state-of-the-art in efficient machine discovery methods.

4.0 Publications

Publications directly supported.

- D. J. Cook and L. B. Holder, "Efficient Knowledge Discovery Applied to Structural Data", *Journal of Artificial Intelligent Research*, 1, pages 231-255, 1994.
- S. Djoko, "Guiding Substructure Discovery with Minimum Description Length and Background Knowledge, to appear in *AAAI* poster session, 1994.
- L. Holder, D. J. Cook, and S. Djoko, "Substructure Discovery in the Subdue System", to appear in *Proceedings of the AAAI Workshop on Knowledge Discovery in Databases*, 1994.
- D. J. Cook and L. B. Holder, "Knowledge Discovery from Structural Data", submitted to the *Journal of Intelligence and Information Sciences*.

Related publications submitted prior to this report period.

- L. Holder and D. J. Cook, "Discovery of Inexact Concepts from Structural Data", *IEEE Transactions on Knowledge and Data Engineering*, 5(6), pages 992-994, 1993.
- L. B. Holder, D. J. Cook, and H. Bunke, "Fuzzy Substructure Discovery", *Ninth International Machine Learning Conference*, Aberdeen, Scotland, pages 218-223, 1992.
- L. B. Holder. "Empirical substructure discovery". In *Proceedings of the Sixth International Workshop on Machine Learning*, pages 133-136, 1989.

(Task 15)

UNIVERSITY OF WISCONSIN AT MADISON

Paradise - A Parallel Information System for EOSDIS

David J. DeWitt
Department of Computer Science

Task Objective

The Paradise project is an effort to prototype a scalable, parallel database system for storing, browsing, and reprocessing geographic data sets, in particular those that will be produced by the EOSDIS project. Paradise provides an object-oriented data model and an extension of SQL to support ad-hoc queries on extents of persistent objects. To facilitate the application of CPU and I/O parallelism in both the data acquisition and reprocessing stages, extents of Paradise objects can be declustered across multiple processors and disks on a parallel processor. Paradise provides a graphical user interface as its user interface and uses the SHORE storage manager for storing and manipulating persistent objects.

1.0 Current Project Status

As the first step toward implementing a parallel version of Paradise, over the past year we implemented a client-server version of Paradise. This version of Paradise provides support for extents of objects that can be defined and queried using an extended version of SQL. In addition to the normal collection of data types (e.g., ints, floats, strings, ...), Paradise also provides four GIS-specific data types which include point, polygon, polyline, and raster.

The target hardware platform for Paradise is a 64 processor Intel Paragon. Each processor is configured with 16 megabytes of memory and a 1.2 gigabyte disk drive. Such "shared-nothing" architectures (see Figure 1) have been widely adopted by commercial parallel database products.

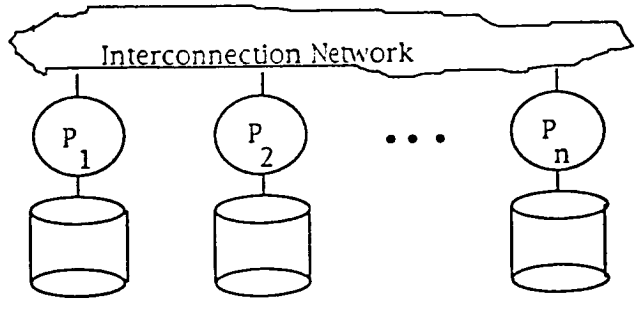


Figure 1. Shared-nothing Multiprocessor Configuration

Initially we used GEO as a user-interface to Paradise. However, in late 1993 we decided that GEO was no longer adequate and began implementing our own interface instead. There were several reasons for this change. First, GEO is implemented using the ET++ class library. This class library is extremely complex and not portable (it only runs on Sun workstations). Second, GEO was designed to use Postgres as its storage manager. We found ourselves constantly converting objects from their Paradise format on disk to a Postgres compatible format for display by GEO. For large collections of objects, this conversion proved very time consuming. Finally, GEO's display capabilities for spatial data are quite limited. In particular, it is not capable of handling objects that have more than one "displayable" attribute; for example, a raster attribute and a polygon attribute. Rather than attempt to fix GEO limitations, we decided that we would be better off starting from scratch. To minimize the effort involved, the approach we adopted was to clone GEO's "look and feel" but to reimplement it using Tk and Tcl instead.

All the Paradise ADTs provide methods that are intended to be invoked through our extended version of SQL. For example, the raster ADT provides a clip() method that clips an image by a bounding box.

We also improved the performance of the raster and polygon ADTs. In particular, since raster images from a satellite can be quite large (on the order typically of 8-10 megabytes), we modified the implementation of the raster ADT to divide raster images into smaller, fixed size chunks (typically each chunk is about 100 Kbytes in size). As an image is being loaded, it is divided into chunks which are then compressed using a lossless compression algorithm. These chunks are then stored as separate objects in SHORE along with a header that acts as a directory for the chunks. With the Sequoia benchmark images (which are from a NOAA satellite) we achieve about a 3:1 compression factor. In addition, we modified the raster ADT algorithm so that it uses the raster header to avoid reading chunks from secondary (and eventually tertiary storage) unless absolutely necessary. For example, if a raster image is clipped by a bounding box, the raster software will use the directory structure that keeps track of the chunks forming a raster image to read only those chunks that are contained within the bounding box into memory. On some of the Sequoia benchmark queries the combination of chunking, compression, and intelligent clipping provides a factor of 10 improvement in performance over the naive strategy of reading the entire image into memory, uncompressing it, and then applying the clip operation.

During the last quarter of year 1 our activities included:

1) Completed a first version of the Paradise query optimizer. In addition to normal relational database optimization techniques, the Paradise query optimizer now incorporates cost functions for operations on its GIS-specific data types.

2) Benchmarking and Tuning. These two activities dominated our attention for the past three months. As a benchmark, we have been using the Project Sequoia regional benchmark. This benchmark involves 11 different queries on a variety of different data types including large raster images from an AVHRR satellite plus large polygons of land-use data from California and Nevada. This benchmark has been used to extensively tune the Paradise software as well as to compare the performance of Paradise to both Postgres and Montage. Paradise has the best performance on 9 of the 11 queries in the benchmark.

3) Bug fixing. We are aiming for a release of the client-server version of Paradise at the end of August 1994.

4) Improvements to the user-interface. As we gain more experience with Paradise, we continue to evolve the user-interface. During the past quarter we added support for displaying objects with more than one spatial attribute as well as support for storing and displaying MPEG encoded video.

5) Ported Paradise to the HP Precision Architecture. The Paradise software now runs on both Sun and HP workstations. While no additional ports are currently anticipated, sometime in the future we will probably do a port to the NT operating system.

2.0 Plans for Upcoming Quarter

A number of activities are planned for the next quarter:

1) Continued tuning and bug fixing

2) Complete implementation of update and aggregate operations.

3) Extend set of join methods to include Grace hash-join.

4) Begin design of the parallel version of Paradise. Begin design of support for tertiary storage for both client-server and parallel versions.

5) Documentation - internal documentation and a user manual.

6) Release version 1.0 of client-server Paradise for beta testing. We hope to identify several beta sites at the July NASA AISRP meeting in Boulder.

3.0 Publications

"Paradise, A Parallel Geographic Information System," D. DeWitt, J. Luo, J. Patel, and J. Yu, *Proceedings of the 1994 ACM Workshop on Advances in Geographic Information Systems*, November 5, 1993.

"Client-Server Paradise", D. DeWitt, Navin Kabra, J. Luo, J. Patel, and J. Yu, accepted to the 1994 VLDB conference.

4.0 Presentations

David DeWitt gave an invited presentation titled "The Trend to Parallel, Object-Oriented Databases" at the 3rd Annual NASA GSFC Conference on Mass Storage Systems and Technologies, October 20, 1993.

David DeWitt presented a paper on the Paradise project at the 1994 ACM Workshop on Advances in Geographic Information Systems, November 5, 1993.

David DeWitt gave an overview of the Paradise project to the January meeting of the CESDIS Science Council.

(Task 16)

GEORGE WASHINGTON UNIVERSITY

Parallel Input/Output Evaluation

Tarek A. El-Ghazawi, Gideon Frieder
Department of Electrical Engineering and Computer Science

Task Objective

Input/output in high-performance computing systems presents a serious bottleneck for two main reasons. First, advances in technology have been increasing processor power, memory capacity, and disk capacity at a much higher rate than that of improving disk seek rate. Secondly as parallel computing started to emerge, the natural priorities were to find interconnection networks and parallel algorithms that can exploit the power of the massive number of processors in parallel machines. Thus, I/O has been receiving less attention and many of the I/O problems are yet to be discovered and solved. Consequently, real world problems frequently encounter I/O bandwidth limits that constrain the overall system performance.

The goals of this research are: (1) To identify the I/O resources required by ESS Grand Challenge applications and their unique I/O access patterns; (2) Identify features of the ESS Grand Challenge I/O requirements and patterns on which system designers should capitalize when planning the next generation of scalable supercomputers, as well as features that can be used to tune I/O-intensive software to parallel architectures; (3) Develop advanced experimental techniques for evaluating and analyzing high performance I/O subsystems; (4) Apply the methods to testbed architectures [such as MasPar MP-1, Convex Exemplar, TMC CM-5, Cray T3D, and Intel Paragon.

1.0 Parallel I/O Evaluation

Input/output speed continues to present a performance bottleneck for high-performance computing systems since technology improves processor power, memory capacity, and disk capacity at a much higher rate.

The MasPar I/O architecture, however, includes many interesting features. This work presents an experimental study of the dynamic characteristics of the MasPar parallel I/O. Performance measurements were collected and compared for the MasPar MP-1 and MP-2 testbeds at NASA GSFC. The results have revealed many strengths as well as areas for potential improvements, and are helpful to system designers, software developers, and system managers. Results were discussed and shared with MasPar technical staff. The experimental work was aimed at investigating the dynamic cache sizes for reading and writing (between the processing elements and the MasPar disk array), I/O prefetching, and I/O bandwidth scalability.

1.1 The MP-1 Testbed Measurements

Some of the results are briefly discussed here. Dynamic cache size can be affected by the specifics of implementations. Figure 1 presents the results of an experiment which investigates the effective I/O cache read size for the MP-1. Performance degrades rapidly when file sizes exceed the 4 Mbyte boundary which indicates that the effective read cache size is 4 Mbyte. However, the static cache size is actually 8 Mbyte. Thus, only 50% of the I/O cache is usable in the case of a read. The lack of ability to take advantage of the full cache size was attributed to the rigid cache block allocation scheme which statically assigns specific

numbers of blocks for specific functions (read, write, ...). Similar behavior was observed also in the case of writing. The Knee of the curve, however, was observed at 3 MB file size.

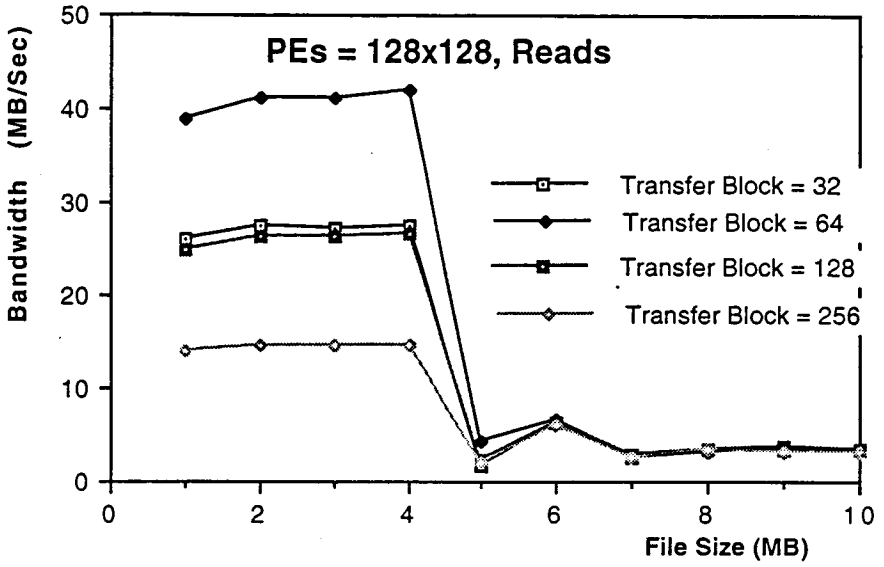


Figure 1. MP-1 Effective Read I/O Cache Size

The experiment of figure 2 was designed to study the scalability characteristics of the I/O subsystem in the MP-1 case study. Discontinuities, at which performance peaks, were observed when the used number of processors is a multiple of 32x32 (1 K processors). The reason for this is the fixed I/O-Router link of 64 bits in the used standard MasPar I/O (PVME) configuration. Note that each one of these wires can connect, via the global router, to a cluster of 16 processors. Thus, the 64 bit channel can connect to a 1K partition of the processor array. When the used file is much larger than the effective size of the cache, the disk array becomes the bottleneck. The sustained disk performance, shown in figure 2 as the 10 MB measurement, was found much lower than the published rates for peak performance as it was limited to only a few MB/Sec.

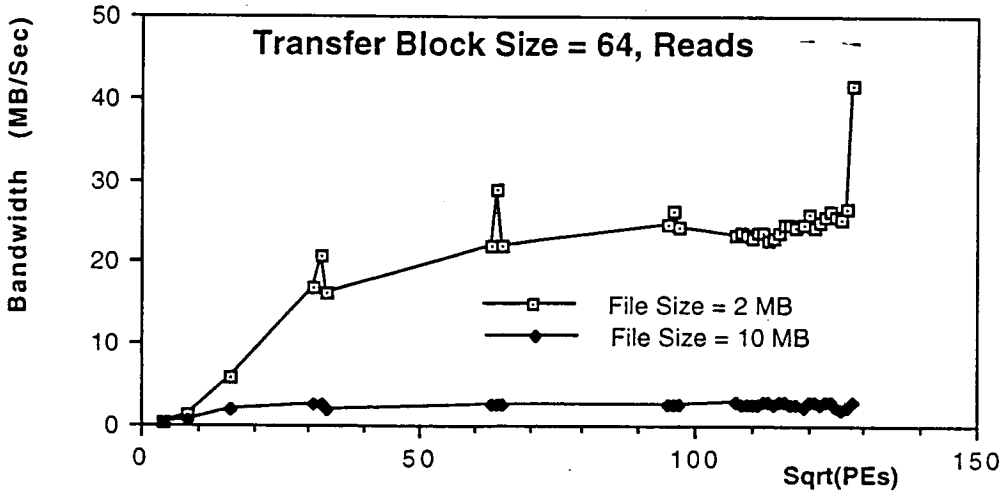


Figure 2. MP-1 Bandwidth Scalability with the Increase in PEs I/O

1.2 The MP-2 Testbed Measurements

A representative set of measurements was obtained once the MP-1 was upgraded to an MP-2 with the MasPar I/O channel, MPIOC, and I/O RAM of 32 MB. With the I/O RAM and the channel, the size of the global router to I/O channel was increased to become a 256 bit channel. The effective read cache size measurements are shown in figure 3. The performance of reads drops significantly when file sizes exceed 12 Mbytes. Thus, the effective cache size is 12 Mbytes for reads. The 12 Mbytes was also found to be the entire allocation for the MasPar file system (MPFS) out of the used 32 Mbyte I/O RAM. When file sizes exceed the 12 Mbyte limit, sustained disk array speeds of about 10 Mbyte/ Sec are observed, which is 33% less than the published rate. However, the published rate of 15 Mbyte/Sec is achievable with very large files (100MB or larger).

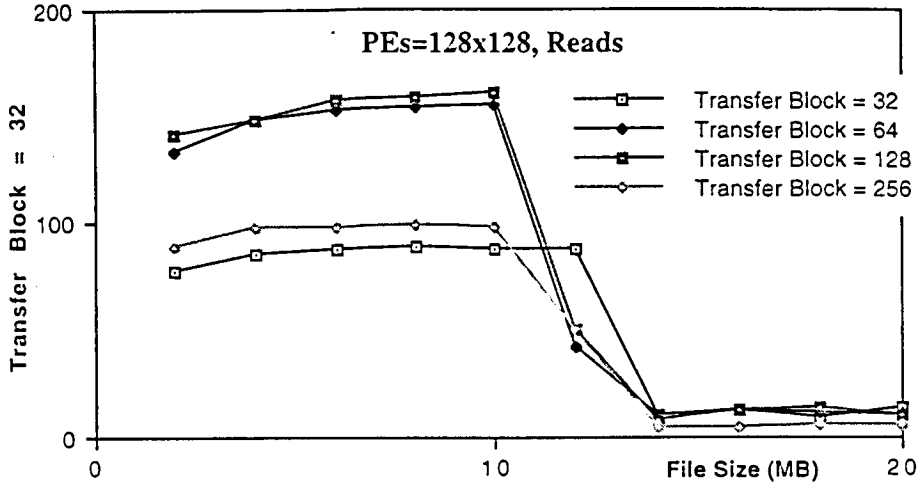


Figure 3. Effective Read Cache for the MP-2/MPIOC Case Study

Effective write cache size, as seen in figure 4, remains at 3 Mbyte even with the increased caching space due to the I/O RAM module. Furthermore, the write performance is about one order of magnitude worse than that of the read.

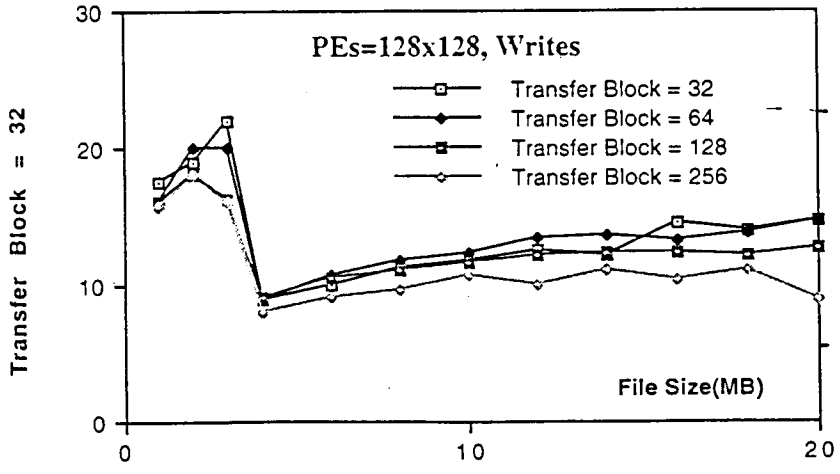


Figure 4. Effective Write Cache for the MP-2/MPIOC Case Study

Prefetching is not different from the first case study and is still following the same simple strategy of leaving a previously read file in the cache. Prefetching on this system was again studied by collecting individual (non-averaged) measurements with and without cache flushing in between.

Scalabiltiy measurements were collected for two files of sizes 10 Mbyte and 100 Mbyte under parallel read operations and plotted in figure 5. This figure resembles figure 2 in the general trend but now with much greater values. For the 100 Mbyte case, the system runs at the speed of the disk array which now demonstrates a sustained performance equal to the published 15 Mbyte/Sec. The 10 Mbyte file displays a great positive spike at 128x128 PEs. This is consistent with figure 2 and the fact that in this new configuration our I/O RAM module provides 256 wires that support 4K PEs through the global router.

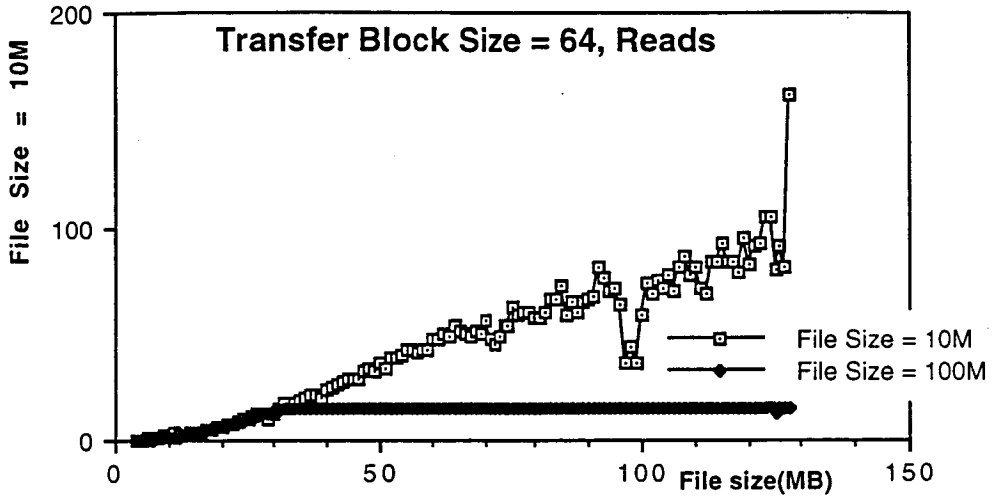


Figure 5. I/O Scalability for the MP-2/MPIOC Configuration

Thus, positive spikes are expected at dimensions that provide multiples of 4K, namely 64x64 and 128x128. No such spike was observed, however, at 64x64. The negative spikes are basically due to the system's heavy activities at the time of the measurements.

In conclusion, with the MasPar I/O channel and I/O RAM, the MasPar I/O subsystem will reach the published sustained performance of the disk array when using large sequential files. Disk technology is clearly the bottleneck and is likely to remain so for some time to come. Programmers of I/O-intensive scientific applications can tune their programs to attain good I/O performance by amortizing the I/O overhead using locality. To do so, they need to be at least aware of their I/O configuration, the specific I/O RAM size and how it is locally partitioned in an attempt to partition data into files that can fit into the I/O RAM. This might be seen still as a usability hurdle as one can not ask application programmers to be aware of such design issues. The work further establishes that system managers need to understand the I/O resource requirements of the applications running on their machines and tune the I/O RAM configuration for best performance. Specifically, efficient partitioning of the I/O RAM among the different I/O activities such as disk reads, disk writes, data processing unit (DPU) to front end communications, and interprocessor communications should be based on an understanding of the most common needs of the local application domain. Thus, the study establishes that a lot more work is needed in the areas of efficient caching of files and I/O data blocks, dynamic partitioning of I/O cache blocks, sources and remedies for discontinuous I/O scalability behaviors leading to operational behaviors that can stay at the best sustained performance points in the scalability curve. On the tools side, work is needed to develop I/O dynamic profiling and tuning tools that could assist users and system managers in making intelligent selections of the best I/O parameters for their applications and system configurations.

2.0 Publications

1. Tarek El-Ghazawi and Jacqueline Le Moigne. "Mutiresolution Wavelet Decomposition on the MasPar Massively Parallel System". *Journal of Computers and Their Applications*, in press.

2. Tarek El-Ghazawi. *Characteristics of the MasPar Parallel I/O System.* Submitted to Frontiers'95

3. A.K. Chan, Charles Chui, Jacqueline LeMoigne, H.J. Lee, Jyh Charn Liu, and Tarek El-Ghazawi. *The Performance Impact of Data Placement for Wavelet Decomposition of Two Dimensional Image Data on SIMD Machines.* Submitted to Frontiers'95.

4. Thomas Sterling, Steven Zalasak, and Tarek El-Ghazawi. "An Innovative Approach to Benchmarking Scalable Parallel Computers for the Earth and Space Sciences Problem Domain". *Scalable High-Performance Computing Conference'94*, IEEE Computer Society, Knoxville, TN, May 1994.

5. Tarek A. El-Ghazawi. *I/O Performance of the MasPar MP-1 Testbed.* TR-94-111. CESDIS, Code 930.5, NASA-GSFC, January 1994.

6. Jeff Pedelty and Tarek El-Ghazawi. "Mapping Morphological Filters onto High-Performance Computing Systems". *Supercomputing'93*, IEEE Computer Society, Portland, OR, November 1993.

7. Thomas Sterling, Tarek El-Ghazawi, Armagan Ozkaya, and Abdullah Meajil. "NASA Science Workload Characterization for Scalable Computer Architectures". *Supercomputing'93*, IEEE Computer Society, Portland, OR, November 1993.

(Task 17)

SYRACUSE UNIVERSITY

High Performance Input/Output Systems for High Performance Computing and Four-Dimensional Data Assimilation

Geoffrey Fox, Alok Choudhary, Kim Mills
Northeast Parallel Architectures Center

Richard Rood
NASA Goddard Space Flight Center

Task Objective

The purpose of this proposal is to apply computer science research in high performance input/output systems for parallel computers to the NASA Grand Challenge application of four-dimensional data assimilation. We will apply leading parallel computing research to a number of existing techniques for assimilation, and will extract parameters indicating where and how I/O limits computational performance. Using detailed knowledge of the application problem, our approach will be to:

- write a parallel I/O support system specifically for this application;
- extract the important I/O characteristics of the data assimilation problem; and
- build these characteristics as parameters into a run-time library for parallel I/O support.

The results of this work will be incorporated into FortranD/High Performance Fortran.

1.0 Progress

- In the application domain, we have studied and identified parts of the 4-D data assimilation application that would require extensive parallel I/O support. In the local model, this part is the Optimal Interpolation component. We have come up with basic design for hand coding the parallel I/O for this application. We are also investigating the Global optimal interpolation model which would require very large conjugate gradient solves. We are currently investigating parallel I/O requirements for this. The technique itself is in the design phase.

- In the system software domain, we have designed runtime primitives to perform parallel I/O (as discussed briefly in the next section). We have tested some of these routines. The eventual goal of the project is to both use the primitive as well as hand code the I/O in the application and compare the two strategies to determine how close can the "generalized" primitives come to hand optimized version in terms of performance. Currently, we have implemented some kernels, both structured as well as unstructured, using the parallel I/O runtime support. A detailed description is provided in the attached technical report.

- We have done some preliminary work on compiler and language issues for developing out-of-core applications in data parallel applications such as HPF. Compiler is able to generate calls to the runtime library to perform I/O in parallel and organize the data in memory as required by the distribution.

2.0 Technical Overview

We have concentrated our recent research on the correlation between data distribution and parallel I/O performance. In several studies, we have shown that the performance of parallel file systems can vary greatly as a function of the selected data distributions, and that some data distributions can not be supported. Further, in these studies we have described how the parallel language extensions, though simplifying the programming, do not address the performance problems found in parallel file systems.

As a result of our studies, we are developing a parallel I/O runtime package that uses an alternative scheme for conducting parallel I/O - the Two-Phase Access Strategy - which guarantees higher and more consistent performance over a wider spectrum of data distributions. We have designed and implemented runtime primitives that make use of the two-phase access strategy to conduct parallel I/O, and facilitate the programming of parallel I/O operations. Our performance results show that I/O access rates are improved by up to several orders of magnitude. Further, the variation in performance over various data distributions is restricted to within a factor of two of the best access rate.

A number of high level programming languages have recently introduced intrinsics that support parallel I/O through a runtime library. By using these primitives, I/O operation instructions within applications become portable across various parallel file systems. Also, the primitives are convenient to use; the instructions for carrying out parallel I/O operations don't involve much more than a declaration of the data decomposition mapping and the use of open, close, read, and write routines.

Yet, these language supported I/O primitives suffer from a serious drawback. Because they use a direct access mechanism to perform the I/O, the user data distribution mapping remains tightly linked to the file mapping to disks. Thus, they are susceptible to the same performance fluctuations and limitations (e.g., unsupported data distributions) that are observed of the parallel file systems.

Our system will provide the portability and convenience of language supported I/O primitives. In addition, because it makes use of the two-phase access strategy to carry out I/O, it effectively decouples user mappings from the file mappings of the parallel file system, and provides consistently high performance independent of the data decompositions used. Further, since these primitives are linked at the compile-time as a runtime library, they can be used with any MIMD (node+message passing) program, or from a data parallel program such as one written in HPF. The runtime primitives library provides a set of simple I/O routines. These include **popen**, **pclose**, **array_map**, **proc_map**, **pread** and **pwrite**.

We have developed models for out-of-core computations. The most prominent model is the Local Placement Model in which a virtual disk is associated with each processor of an MPP. For this model, we have developed preliminary runtime support to perform read/write accesses, prefetching to overlap I/O and computations and a communication system which handles out-of-core data as well.

2.1 Advantages of the Runtime I/O System

- The runtime system can be easily ported on various machines which provide parallel file systems. This makes the runtime primitives highly portable and easy to use.
- By using these primitives, the more complex data distributions (Block-Block or Block-Cyclic) are made available to the user. The only additional information required are the global, local array information and the processor grid information.
- Primitives allow the user to control the data mapping over the disks. This is a significant advantage since the user can vary the number of disks to optimize the data access time.
- Under certain conditions, the primitives allow the programmer to change the data distribution on the processors dynamically.

- The data access time is significantly improved and is made more consistent since the primitives use two-phase access strategy.

We are incorporating the use of these primitives into our HPF compiler development to allow for automatic introduction of I/O instructions into programs. This will facilitate programmer management of data access and of computation to communication overlap.

3.0 Plan

For the next two years, we plan to do the following.

- Develop a more extensive runtime system for the Local Placement Model. Incorporate optimizations in the runtime system including prefetching strategies and data reuse.
- Implement application templates using these runtime primitives.
- Incorporate automatic embedding of the runtime primitives in the Fortran 90D/HPC compiler.

References

R. Thakur, R. Bordawekar and A. Choudhary, *Compiler and Runtime Support for Out-of-Core HPF Programs*, to appear in the International Conference on Supercomputing, Manchester, England, July 94.

Choudhary Alok, Parallel I/O Systems, Guest Editor's Introduction, *Journal of Parallel and Distributed Computing*, January/February, 1993.

Juan Miguel del Rosario, Rajesh Bordawekar and Alok Choudhary, *Improved Parallel I/O via a Two-phase Run-time Access Strategy*, IPPS I/O Workshop, 1993.

Rajesh Bordawekar, Juan Miguel del Rosario, Alok Choudhary, *Design and Evaluation of Primitives for Parallel I/O*, Supercomputing '93, November, 1993.

Juan Miguel del Rosario and Alok Choudhary, High Performance I/O for Parallel Computers: Problems and Prospects, IEEE *Computer*, March 1994.

(Task 18)

UNIVERSITY OF VIRGINIA

High Performance Parallel I/O Support for Multidimensional Range Searches

James C. French
Andrew S. Grimshaw
Department of Computer Science

Task Objective

The high performance scientific data analysis typified by ESS applications is plagued by chronically inadequate I/O performance. The situation is aggravated by ever improving processor performance. For high performance multi-computers such as the Touchstone Delta that possess in excess of 500, 60-megaflops processors, I/O will be the bottleneck for many scientific applications. Viable solutions will never be achieved with conventional file systems.

This proposal describes ELFS (an *E*xtensib*L*e *F*ile *S*ystems). ELFS attacks the problems of:

- providing high bandwidth and low latency I/O to applications programs on high performance architectures;
- reducing the cognitive burden faced by applications programmers when they attempt to optimize their I/O operations to fit existing file system models; and
- seamlessly managing the proliferation of data formats and architectural differences.

The ELFS solution consists of language and run-time system support that permits the specification of a hierarchy of file classes.

1.0 Introduction

High performance computer systems are becoming increasingly unbalanced between the performance of the CPU and the I/O subsystems. Over the last decade, CPU speeds have increased dramatically, while at the same time I/O performance has improved only marginally. Thus, performance of many applications, in particular many scientific applications, has become bounded by the performance of the I/O subsystem and its related software. An example of this problem is examined in [DUQU85], where a deconvolution application uses only 20 minutes of CPU time on a CRAY X-MP computer, but due to inadequate I/O performance requires over 10 hours of wall clock time to complete. The advent of highly parallel architectures has made the problem worse because these architectures can consume data at an even faster rate than single CPU systems. In addition, it appears likely that the imbalance between the improvements in CPU and I/O system performance will continue into the foreseeable future, further aggravating the problem. Clearly, new approaches are needed to support I/O-intensive high performance applications.

We have developed a general approach for attacking the high performance I/O problem, namely the Extensible File Systems (ELFS) approach based on work in [GRIM91]. This report describes an implementation following the ELFS approach for a specific class of retrieval patterns, multidimensional range searches. Multidimensional range searches appear in a wide range of applications, including many scientific applications. Such applications view a data set as an n-dimensional data space, where each dimension represents

the values along a key field present in the data. The coordinates of each data record are its values for each of the n dimensions. Using this view, subvolumes of the data space can be defined by specifying a range of values for each dimension. For example, a data set containing a set of time indexed two dimensional images can be viewed as a three-dimensional data space (time,x,y). Possible range searches for such a data set include retrieving a specified region of each image (a rectangle in (x,y)) for all time values, retrieving full images for a certain range of times, etc.).

In the following sections we first present the current methods used for providing range search capabilities and then briefly describe the general ELFS approach and its benefits. The remainder of the report discusses an instance of this approach designed for high performance multidimensional range searches including details of the parallel structure of the implementation. Since this report describes work currently in progress, we cannot include a full description of the implementation nor a full performance analysis. However, we do describe the tests we are executing using interferometry data sets from the National Radio Astronomy Observatory (NRAO). More details on these tests can be found in [KARP94a,KARP94b].

2.0 Current Methods

There are several approaches typically employed to provide range searching capabilities on a set of data, each with varying degrees of implementation effort and performance. Many implementations store the data sets as simple sorted sequential files and scan the file, filtering out unwanted data outside of the desired subvolume. This approach is easy to implement, but performs poorly, especially when small amounts of data are desired relative to the file size. An improvement to this scheme uses an indexed, sorted file, to reduce the number of accesses needed to the file, improving performance at a modest complexity cost. This scheme works well for a one-dimensional space (for the sorted, indexed key), but does not generally perform well for multidimensional accesses. Some implementations designed for parallel applications, improve the performance of a single file by either replicating the data sets or partitioning the data set into separate disjoint sets. Each of these approaches is designed to alleviate the contention for the single file resource among multiple concurrent processes, but does not improve upon the basic access methods for each distributed file.

Another common approach is to use a commercial database management system (DBMS) which allows for the specification of range queries. Relational DBMS are particularly popular where range searches are easily defined using the Structured Query Language (SQL). This approach is easy from the implementation standpoint, but may not achieve acceptable performance. DBMS are built to support a wide range of possible access patterns and types of data and are not tuned to range searching in multiple dimensions. In addition, DBMS incur overhead for the guarantee of consistency within the database, which may not be an issue for many applications.

A less often used approach implements file structures specifically tailored for range searching such as PLOP files[KRIE88a,KRIE88b], grid files [NIEV84], k-d & k-d-b trees [BENT79,ROBI81], or quadtrees [SAME84]. These file structures offer performance advantages by attempting to preserve physical data locality in all of the dimensions of the data space and by providing efficient methods for finding particular regions of the data space. The drawback is that these file structures can be difficult to implement properly, especially in a distributed manner. Even when these structures are implemented, the implementations are often highly application-specific and not reusable, so the common practice is to build them virtually from scratch.

2.1 The ELFS Approach

The ELFS approach is to create file objects that satisfy four criteria:

(1) Match the file structure to the access patterns of the application and the type of the data. As the examples in the previous section point out, the organization and structure of the underlying file can greatly influence

performance by reducing the number of accesses required. For distributed file structures, effective data placement can potentially improve performance by reducing latency. Therefore it is important to match file structures with their use.

(2) Use parallel and other advanced I/O techniques. In a file type-specific manner exploit parallelism to overlap application computation with I/O requests to reduce the effective latency of a request (i.e. the wait actually experienced after issuing a request). For distributed file structures, true parallel access can be used to better utilize the file system's bandwidth. Other I/O-related functions, such as data conversion and sorting, may be performed in parallel to speed the overall performance of using stored data. Prefetching and caching are two other well-known performance-oriented techniques that can be employed when applicable.

(3) Improve the I/O interface to application programs. There are two main reasons for improving the file interface. Most importantly from a performance standpoint, is to allow the user to convey useful information that can be exploited by the file object implementation. For example, knowing the stride of accesses in a matrix file can be exploited to effectively prefetch data, or knowing that the file will be used in a read only fashion can allow the implementation to avoid potentially costly consistency protocols. The second reason for improving the interface is to make file objects easier to use by application programmers, reducing their programming burden.

(4) Encapsulate the implementation details in file objects. This goal is aimed at increasing the maintainability and reusability of the file objects. By using the object-oriented paradigm for the file objects, application programmers can derive new file objects from existing base objects and can then extend them and tailor them without reimplementing much of the file object functionality.

A suite of extensible file objects can be developed using this methodology, each performing best for a particular class of data types and access patterns. Application designers can then choose the best file object for their purposes and extend or tailor the file definition as needed, hopefully requiring only a modest amount of effort. Our early design work in this area has been reported in [KARP94c].

2.2 Parallel File Objects for Multidimensional Range Searches

Using the ELFS approach we have created a parallel file object designed to provide high performance for range queries in multiple dimensions. Our implementation uses the PLOP file as the basic underlying file structure. Though other file structures could be used for multidimensional range searches, it is our opinion that none of these candidates is clearly superior to PLOP files, while PLOP files have a relatively straightforward implementation. For a more in depth analysis of the choice of file structure see [KARP94a]. A PLOP file views a data set as a multidimensional data space. The data space is partitioned by splitting each dimension into a series of ranges called slices. The intersection of a slice from each dimension defines one logical data bucket. Data points are stored in the bucket that has corresponding values in each dimension. Therefore, within a bucket, the data points exhibit spatial locality in all dimensions. A tree structure for each dimension tracks the physical location of each bucket within the file, so that each bucket can be accessed very efficiently. This structure allows retrievals to eliminate parts of the file that do not correspond to values within the range search based on all dimensions, while quickly accessing those parts that may contain valid data.

We first implemented a sequential version of the PLOP file based file object. Though unable to take advantage of parallel techniques, this version exploits the structure of PLOP files to achieve efficient accesses. In addition to the obvious benefits of using the tailor-made structure of the PLOP file, a subtle performance improving enhancement was implemented for sorting by a key that is one of the dimensions. Because the data points contained in each slice along a dimension are disjoint, the data in each slice along the sort key can be sorted separately, and with each slice returned in order. By sorting in smaller batches, the complexity of the sort is reduced from O(nlogn) to O(n/plogn/p) in the ideal case (p = number of slices spanned by the request).

The parallel version is being implemented using Mentat [GRIM93], an object-oriented parallel processing system. The design of the parallel implementation includes several significant changes from the sequential version. First, PLOP files have been modified to accommodate distributed pieces. These pieces can be created and distributed in three patterns: segmented (or partitioned) along a dimension, striped along a dimension, or blocked by some set of dimensions (e.g. using two dimensions each piece would be a rectangular region). The distributed PLOP file allows not only parallel access to the data, but also allows an application program to map processes to nodes near the data they will require.

Second, parallel I/O workers have been added to access each distributed file piece and a manager has been added to coordinate their activities. The workers asynchronously handle all requests for data at their piece, including I/O device access, data conversion and, if possible, data sorting. Our initial design has only a single manager process for all worker processes and clearly does not scale well for increased numbers of application processes requiring data. We already plan to replace this design with a scheme that will scale for increases in the number of application processes by enabling the manager to replicate itself and assign different managers to different application processes.

Third, the interface has been improved. A major improvement to the interface to decouple the definition of a query request from the retrieval of the data. The idea is to allow the user to specify a query ahead of the time the data will be used whenever possible, and to submit the query to be performed. The file object can asynchronously begin buffering the request while the application continues to do useful work. When the application actually wants the retrieved values, a call is made to ask for the data.

2.3 Performance Tests

To test our implementation, we have converted two interferometry data sets from NRAO's Very Large Array (VLA) radio antenna installation. The first file, a line spectrum file, is ~50 megabytes and 126,092 records, while the second file, a continuum spectrum file, is ~270 megabytes and contains 8,440,092 records. Initial results for the sequential version have been very encouraging. The converted PLOP files utilize space fairly efficiently, 79% and 66% for the line and continuum files respectively (efficiency is calculated by comparing actual storage used for records versus the total storage allocated, including overhead and fragmentation).

To test the performance of range retrievals, a set of twelve representative queries for NRAO's applications has been developed. Both files have been tested using these queries for the sequential version, with impressive results. The parallel version will be tested with the same suite of queries for various file distribution patterns and numbers of file pieces. These results and a comparison of results across the various parameters can be found in [KARP94a,KARP94b].

3.0 References

[BENT79] J. L. Bentley and J. H. Friedman, "Data Structures for Range Searching", *ACM Computing Surveys* 11, 4 (Dec. 1979), 397-409.

[DUQU85] B. Duquet and T. Cornwell, "Deconvolution on the Digital Production Cray X-MP", *NRAO Newsletter*, July 1, 1985, 10.

[GRIM91] A. S. Grimshaw and E. C. Loyot, Jr., *ELFS: Object-Oriented Extensible File Systems*, Tech. Rep. CS-91-14, Dept. of Computer Science, University of Virginia, July 1991.

[GRIM93] A. S. Grimshaw, "Easy to Use Object-Oriented Parallel Programming with Mentat", IEEE *Computer*, May 1993, 39-51.

[KARP94a] J. F. Karpovich, A. S. Grimshaw and J. C. French, *Breaking the I/O Bottleneck at the National Radio Astronomy Observatory*, Tech. Rep. (in progress), Dept. of Computer Science, University of Virginia, Charlottesville, VA, 1994.

[KARP94b] J. F. Karpovich, J. C. French and A. S. Grimshaw, High Performance Access to Radio Astronomy Data: A Case Study, *Proc. 7th Inter. Conference on Scientific and Statistical Database Management*, 1994. To appear.

[KARP94c] J. F. Karpovich, A. S. Grimshaw and J. C. French, "Extensible File Systems (ELFS): An Object-Oriented Approach to High Performance File I/O", *Proc. OOPSLA '94: Object-Oriented Programming Systems and Languages*, 1994. To appear.

[KRIE88a] H. Kriegel and B. Seeger, "PLOP-Hashing: A Grid File without a Directory", *Proc. of the Fourth Inter. Conf. on Data Engineering*, Feb. 1988, 369-376.

[KRIE88b] H. Kriegel and B. Seeger, "Techniques for Design and Implementation of Efficient Spatial Access Methods", *Proc. of the 14th VLDB*, 1988, 360-370.

[NIEV84] J. Nievergelt and H. Hinterberger, "The Grid File: An Adaptable, Symmetric Multikey File Structure", *ACM Transactions on Database Systems* 9, 1 (Mar. 1984), 38-71.

[ROBI81] J. T. Robinson, "The K-D-B-Tree: A Search Structure for Large Multidimensional Dynamic Indexes", *Proc. of the Annual Meeting of the ACM Special Interest Group on Management of Data*, 1981, 10-18.

[SAME84] H. Samet, "The Quadtree and Related Hierarchical Data Structures", *ACM Computing Surveys* 16, 2 (June 1984), 187-260.

(Task 19)

UNIVERSITY OF FLORIDA

Distributed Search Structures

Theodore Johnson
Department of Computer and Information Science

Task Objective

The management of extremely large data sets requires the use of distributed storage for two primary reasons: scalability and throughput. The density of a storage device is limited by the current technology, so larger devices can't be counted on to provide additional storage. Further, extremely high volume devices do not solve the problem of throughput in a highly parallel system. Data storage must therefore be accomplished by using a large number of devices. Centralized management of these devices will result in an I/O bottleneck, so that the storage devices must be distributed throughout the system.

While distributed storage can solve the large volume storage problem, it opens another difficult problem, namely that of providing an index into the data. If the stored data is to be useful, it must be quickly and easily accessible to the users of the system. Thus a distributed index is needed in order to access the stored data. Maintaining a distributed index is difficult for several reasons. First, the index must reflect the current data that is stored in many different sites. Second, the index must be scalable and highly parallel, so that it doesn't become the bottleneck in the system. Finally, the index must be able to refer to a very large number of entries–perhaps billions of entries–so the index itself might require distributed storage.

We propose to develop, implement, and make available a scalable, highly parallel, and efficient distributed index. The distributed indices will automatically allocate new data files to the storage devices in a manner that balances the data load on each storage device. In addition, the distributed index will allow a researcher to rapidly access data files without being required to know the location of the storage device.

Our proposed research includes both theoretical and implementational components. Theoretically, we plan to study the application of lazy updates to distributed search structures. In addition, we will build a working prototype for the Goddard National Space Science Data Center's National Data Archive and Distribution Service (NDADS), and test its performance in a production environment.

1.0 Theoretical Development

We developed the theoretical underpinnings needed to implement distributed and replicated search structures. A discussion of the correctness theory is contained in the University of Florida Dept. of CIS technical report TR93-034. This report can be obtained by anonymous ftp at *ftp.cis.ufl.edu:cis/tech-reports/tr93/tr93-034.ps.Z.*

We have extended the theory contained in TR93-034 to handle some problems encountered in deleting and merging nodes. While we have not yet written up these algorithms formally, the ideas are in use in our experimental implementation.

1.1 Implementation

The implementation now handles search, insert, and delete operations. The extension to handle deletes properly requires a refined understanding of the semantics of increasing a node's range of responsibility. We have implemented the necessary algorithms and will write a formal description of the algorithms in the 1994-1995 period.

The current implementation incorporates mechanisms for monitoring performance and for load balancing. In the technical report TR94-009, we discuss the implementational issues we encountered, how we solved the problems, and present performance measurements of the implementation. The performance measurements are encouraging, as they show that the *width of replication*, or average number of copies of an interior node, is moderate, while the number of messages required to find an entry is also moderate. The technical report TR94-009 is available via anonymous ftp at *ftp.cis.ufl.edu:cis/tech-reports/tr93/tr94-009.ps.Z.*

We have implemented the distributed B-tree on a 96-node KSR1. Though the KSR1 is a shared memory machine, we do not make use of the shared memory. Each processor in the KSR1 runs an OSF/1 kernel, and supports Berkeley sockets. Thus, we use the KSR1 as a convenient testbed for our message-passing algorithm (as it is easy to allocate processors, and there is little network interference). Unfortunately, kernel bugs in the recent releases of the kernel cause problems in our socket allocation. We are porting the implementation to run on the network of SUN workstations maintained by the University of Florida CIS department.

1.2 Performance

The measurement studies on the implementation are encouraging, but experiments are difficult to set up and execute. To more quickly test a wide variety of replication and load balancing strategies, we wrote a simulation of a distributed B-tree. The simulator can report statistics such as width of replication, messages per operation, variation in storage requirements, etc. A timing study requires the implementation. Preliminary simulation results show that the width of replication is between 2 and 3, and the number of messages required for an operation is about 2, relatively independent of the node fanout or the number of processor that maintain the index (we ran simulations with up to 50 processors).

We have observed that the data balancing algorithm has a significant impact on performance measures such as width of replication and messages per operation. For example, we have observed that a random probing strategy is significantly better than a sequential probing strategy. We are investigating data balancing strategies that use locally available information about the structure of the index. We expect to write a report (to be posted at the ftp site) on these experiments in late July 1994.

2.0 Miscellaneous

With Dennis Shasha of New York University, we developed a high performance cache management algorithm. A paper describing this algorithm, *2Q: A Low Overhead High Performance Buffer Management Replacement Algorithm*, will be presented at the 1994 Very Large Data Base conference. The paper acknowledges the support of USRA.

(Task 20)

CLEMSON UNIVERSITY

High Performance Input/Output for Parallel Computing Systems

Walter B. Ligon, III
Department of Electrical and Computer Engineering

Task Objective

The goals for this projects are to:

- develop an understanding of the performance potential of I/O architectures with respect to the requirements of EOS applications,
- provide a mechanism to integrate this understanding into the Distributed Active Archive Center (DAAC) processing environment at NASA,
- conduct an evaluation of parallel I/O architectures relative to the requirements of parallel applications, and
- integrate the results of this study into the IIFS testbed being developed by NASA Code 935 (Information Science and Technology Office).

1.0 Project Goals

Our efforts during the first year have focused on two issues:

1) preparing an experimental environment for the proposed study, and

2) identifying appropriate processing algorithms.

Our experimental environment is to consist of two major components. The first and primary one is a simulation environment for studying the effects of various architectural parameters on I/O performance. For this component we are building on existing infrastructure in the Reconfigurable Architecture Workbench, which was developed under a previous research program. The second is a small-scale parallel computing system utilizing clusters of high-performance workstations built through an equipment grant from the NSF. This second facility is much less flexible in its capabilities, but does provide a means for validating some of the results obtained under simulation. This second approach also provides a view of the potential to be had in low-cost systems for Earth Science computations.

In our study we consider parallel systems as a collection of processors each with their own local memory and a network that provides basic message passing capability. This network can be a simple as an Ethernet, or as complex as a crossbar, torus, or hypercube network. I/O in these systems are provided by some number of I/O processors, where each I/O processor is a processor with local memory and directly attached storage devices such as disks and tape drives. I/O processors can be either dedicated or can also be used for computation. The number of I/O processors can be the same or different than the number of compute proces-

sors. I/O devices are attached to the I/O processors via a device bus (such as a SCSI bus) and a device controller. One or more devices can be connected to the device bus and one or more device controllers can be attached to a given I/O processor.

Our performance models include device behavior (focusing on Winchester disk and digital audio tape (DAT) devices), I/O bus performance, and interconnection network performance. The load placed on these facilities is determined by the behavior of software both at the system and application levels. System software consists primarily of device driver codes, message passing libraries, and file system software. Of these, the parallel file system code is unique to this system model. In addition, key design choices in the message passing software may have an impact on I/O performance. Key issues include the amount of data copying performed and details of the networking protocol.

Currently, the main focus of our search for applications is on processing level zero telemetry data to levels 1 and 2, data product generation, and automatic metadata extraction. Level 1 and 2 processing algorithms include sensor calibration, georegistration, correction and enhancement. Data product generation is highly application specific, but would include such things as vegetation index, snow and ice cover, sea surface temperature, atmospheric ozone content, etc. Metadata extraction is the process of preparing data sets for inclusion in an earth science database by generating browse products and summary data. These activities would comprise a large share of the constant processing requirements for a typical DAAC scenario in both preprocessing and reprocessing modes.

2.0 Progress Made

During the last year we have made significant progress both in preparing our experimental environment and in collecting target applications. We have developed two key pieces of software for use in the study: the Parallel Virtual File System (PVFS), and an enhanced version of the RAW simulator dubbed Tiger Parallel Architecture Workbench (TPAW). We have begun working closely with groups at NASA GSFC to gather appropriate codes for study and are currently working with Code 935 in the development of an Earth science data object database and calibration, georegistration, and data product generation codes.

2.1 PVFS Parallel File System

As discussed in section 1.0, a key piece of system software in the parallel system model we are studying is the parallel file system. This software is responsible for determining the distribution of data among the available devices and providing a unified user interface to facilitate access to the data by the tasks of a parallel program. A few such systems have been presented in the literature. Among these we selected the Vesta file system developed by IBM Research as a model for PVFS due to its emphasis on flexibility. It is hoped that this flexibility will provide us a good vehicle for experimentation with critical file system parameters. Our initial implementation utilizes the existing Unix file system on each processing node for local allocation of blocks within each physical partition of a file. TCP/IP is used for communication between processors, and a library is provided for user access. A version of the file system has been implemented on our Unix parallel system integrated with the Parallel Virtual Machine (PVM) message passing library available via the Internet. Another version has been implemented under TPAW. Future versions will be implemented using various communications protocols (in order to determine their effect) and the TPAW version will be studied using more sophisticated network such as those found on commercial parallel processors. The details of PVFS have been documented in a Masters Thesis written by Mr. A. Blumer and a paper co-authored by Mr. Blumer and the Principle Investigator is to be submitted to an international conference in July, 1994.

2.2 TPAW Simulation Environment

A key drawback to RAW as a tool for studying I/O systems is that RAW's processor simulator uses instruction-level simulation, which is to say target programs are compiled to an abstract assembly language and interpreted by the simulator. This was designed as such in order to facilitate the study of reconfigurable processing elements in a previous research program. In order to study I/O system, it is important that considerably longer programs than are practical to study with such a system (due to the long simulation time) are used. Since processor instruction set is not a key issue in this study, the processor simulator has been replaced with a new module that executes the target application compiled to binary code suitable for the host on which the simulation is to be run. This results in a simulation that is several orders of magnitude faster than under the old system. Available systems that utilize such a technique are limited in that they only work for one host processor type and are limited to MIMD, and in some cases only shared memory architectures. In order to maintain RAW's flexibility in these areas, a new simulation tool was developed. This tool uses a source-code augmentation technique that maintains a high degree of portability. All simulator code is written in C, and few vendor-dependent system calls are used. Where possible development utilized POSIX compliant calls. This system simulates both SIMD and MIMD architectures and focuses on message passing systems (though shared memory is supported as well). Control and Network modules from the RAW simulator transfer readily to the new platform and the PVFS file system and I/O device models have been developed with the new simulator in mind. The details of TPAW have been documented in a Masters Thesis by Mr. R. Agnew and a paper co-authored by Mr. Agnew and the Principle Investigator is to be submitted to an international journal in September, 1994.

2.3 Object Databases

One of the least understood applications areas in EOS DAAC processing is the area of intelligent data management. Of primary interest is the creation and access to an object database that records metadata needed to find specific earth science data in a terabyte sized archive. In addition, the object database must record knowledge on processing techniques in order to transform raw data into the desired end data products. Such a system would typically require significant amounts of storage in its own right, and must be able to support hundreds of simultaneous users. Considerable work in this area is being done by NASA Code 935 at GSFC. A Clemson ECE Ph.D. student is focusing on this area in order to study I/O subsystem behavior of such an application. This student (A NASA Graduate Research Fellowship holder) is currently spending his summer on-site at GSFC working with Code 935 on developing an object-oriented database for just such purpose. This is expected to be an important source of experiments for our study.

2.4 Remote Sensing Algorithms

The processing algorithms we expect to comprise the bulk of a typical DAAC scenario would consist of those algorithms that process raw instrument telemetry data into specific data products needed by the scientific community. Some of these algorithms would be run as a standard processing suite during data ingest, others would be the result of a specific request for data. In each case, large amounts of data would need to be output to and input from archival storage, in addition to transfers between secondary staging devices and main memory. During the last year we have established contacts with NASA Code 935 and have been working with them on codes for calibration and registration of radiospectroscope data, specifically AVHRR. We expect to continue with algorithms for standard data products such as vegetation index and sea surface temperature. This group is also looking towards the MODIS sensor as a natural follow-on to AVHRR. These algorithms are representative of those used on other similar radiospectroscopes, and could be adjusted to account for different spatial and spectral resolutions. We believe this is a good start at looking at critical and well-proven algorithms. In the future we hope to focus on more exotic and experimental algorithms.

3.0 Interaction with Other Projects

While this project is self contained, it does not operate in a vacuum but rather is influenced by interactions with other on-going projects at Clemson. In particular, three projects have had an impact on this project (and vice versa): The Macintosh Telemetry and Control (MacTAC) project, a recent NSF research equipment grant, and the Beowulf project at NASA GSFC.

3.1 MacTAC Project

The goal of the MacTAC project is to develop low-cost level zero processing capability for future direct broadcast capability. This is an on-going project between NASA Code 521 and the Clemson ECE department headed by the principle investigator. Our work with Code 935 has revealed a desire to move level 1 processing into the semi-programmable hardware modules in the MacTAC system. This, in turn would provide a primary source of input for a low-cost ground processing station. Thus a second area of consideration for the I/O project is alternate system configurations for low-cost earth science data processing. This scenario has lower data rate and storage requirements, but has a similar dependency on I/O capability balanced with moderately powerful processing capability.

3.2 NSF Equipment Grant

The Clemson ECE department was recently awarded a grant from the NSF for computer equipment to be used in part in the study of I/O subsystems for high performance parallel computers. This equipment provides both a platform for conducting simulation experiments, but also a system for limited experimentation with system software prototypes. We intend to study the behavior of the PVFS file system using this equipment as a mechanism for calibrating our simulator. Finally, this system is representative of the computing power that could be applied as a low-cost direct-broadcast ground processing station, as discussed in section 3.1.

3.3 Beowulf Project

This is a project conducted at NASA GSFC by USRA/CESDIS and NASA personnel (see Staff Scientist section, Task 38). We have maintained contacts within the project to track its progress with the hopes of providing our input. Currently the project is building a low-cost parallel processing system based on commodity parts. Such a system would be a prime candidate as the processing unit for a low-cost ground processing station. We are hoping that the PVFS file system will be adopted as the file system for the project.

4.0 Second Year Projections

The second year of our project will see initial experiments using the TPAW simulator and an extensive study of the PVFS file system. Significant effort will focus on the effect communications protocols and network design has on I/O throughput. Considerable effort will focus on algorithm collection and development through contacts with NASA Code 930. In addition, the process of integrating our results into ongoing NASA missions will begin both through interactions with Code 930 and Code 521.

5.0 Conclusions

Our first year seems to be on track. We have completed most of the key development tasks and are beginning the process of collecting suitable algorithms. Our contacts within NASA have become well established and we are beginning to make an impact.

(Task 21)

UNIVERSITY OF WASHINGTON

A Visual Database System for Image Analysis on Parallel Computers and its Application to the EOSRAM Project

Linda G. Shapiro, Steven L. Tanimoto, James P. Ahrens
Department of Computer Science and Engineering

Task Objective

Standard database systems were designed for business applications and are usually not equipped to handle scientific data. Thus, there has been a lot of interest in developing database systems that are more appropriate for the large volumes and different types of data that are used and produced by scientific applications. These applications create new problems in data storage, organization, and retrieval and can require monumental amounts of processing time.

The problem we are working on is the creation of a design and prototype implementation of a scientific database system that is particularly suited for handling the image, vision and scientific data associated with the EOSRAM project. We will focus on a database structure and facilities for querying that are designed to execute efficiently on parallel computers.

1.0 Introduction

During this year, we gathered information from the scientists working on NASA's Earth Observing System—Regional Amazon Model (EOSRAM) project here at the University of Washington about their database needs. Using this information, we identified effective representations for their database queries. These queries usually ask for data which is not currently available in the database and must be computed on demand. Using these representation, we formulated a collection of example queries. We have identified methods of executing these queries efficiently on parallel computers and are working to automate the application of these query optimization methods.

2.0 An Overview of the EOSRAM project

Our work is being done in conjunction with the NASA Earth Observing System (EOS) IDS project, *The Regional Amazon Model: Synoptic Scale Hydrological and Biogeochemical Cycles from EOS*, Jeffrey Richey, University of Washington, PI. The mission of the EOSRAM project is to contribute to understanding the dynamics of the Amazon system in a natural state, and how it would evolve under possible change scenarios (from instantaneous deforestation to more subtle longer term climatic/chemical changes). The overall goal of the project is to determine how extensive land-use changes in the Amazon would modify the routing of water and its chemical load from precipitation, through the drainage system, and back to the atmosphere and ocean.

The work is being undertaken by two groups here at the University of Washington: the Hydrology group which is part of the Oceanography department and is headed by Jeffrey Richey and the Remote Sensing

group which is part of the Geology Department and is headed by John Adams. The Hydrology group is currently working on creating a computerized model of the hydrology of the Amazon River Basin.

An integral part of this project is obtaining, manipulating, and understanding satellite images of these regions. The Remote Sensing group focuses their efforts on this aspect of the project. One of their main interests is the classification of the regions of a satellite image, so that the primitive components of a region (such as particular soil and vegetation types) are identified.

Adams and his team have developed a model of a satellite image as a mixture of a set of spectra called endmembers. Spectra are represented as vectors of signal measurements of different light wavelengths. The Remote Sensing group analyzes images from the AVIRIS satellite whose pixels are vectors representing spectra. Spectra can be also obtained for real-world objects, such as soil and vegetation types. A collection of these spectra are stored in a computerized library. The Remote Sensing group has formulated an algorithm which models the pixels of a spectral satellite image as a mixture of endmember spectra. The endmembers are chosen from the library by a scientist and input to the algorithm along with a spectral satellite image. Solution fractions are output by the algorithm, one for each endmember. A solution fraction contains the fraction of each endmember that is needed to accurately model the image. The Remote Sensing group uses this spectral mixture analysis algorithm to calibrate satellite images for atmospheric effects and to classify regions of satellite images.

They have implemented a version of this algorithm on an IBM PC with a 486 processor and an i860 co-processor. Their PC is equipped with Direct Memory Access device which allows I/O operations to execute in parallel with computations. The inner loop of the algorithm has been hand-coded in assembly language for speed. Executions takes approximately 300 seconds to complete on an satellite image of size 692x614 pixels each with a vector of 124 signal measurements called bands.

We have worked closely with members of the EOSRAM project to understand the types of computer experiments they run and their database needs. From these discussions, we have identified what these scientists would like in a database query mechanism. As part of a query, scientists would like to apply operators to the scientific data that is stored in the database to generate new data on demand. They would also like to interactively create and modify queries. Queries cannot be statically defined because scientists use their domain knowledge to interpret the results of queries and then use this information to guide subsequent query formulation.

3.0 Query Definition

We have identified an effective representation for the queries required by the EOSRAM scientists. The representation is a flow graph which is composed of scientific data and operators. A query is formed by connecting the outputs of operators through intermediary data elements to the inputs of other operators. With this representation a scientist can build up a complex processing request. For example, the spectral mixture analysis algorithm can be expressed as a query, as we will show later in this report. These processing requests can be modified dynamically by adding or deleting data or operators from the query graph. We envision the scientists utilizing an interactive visual interface to view and manipulate query graphs.

Formally, a query graph is a data-flow graph G with nodes N and edges E. Nodes are either data or operator nodes. Edges connect data and operator nodes, denoting the flow of data through the graph. There are two kinds of operator nodes: storage/retrieval and computation. Storage operator nodes store data in the database and retrieval operator nodes retrieve data from the database.

We have identified and will now describe three different types of query graphs: schema, instance and experimental.

3.1 Query Schemas

A query schema describes a general query structure. All the nodes of a query schema are constrained by a data or operator type. A data type is a data structure. An operator type defines an operator's input and output types. Nodes in a query schema may also be constrained by an instance value which is compatible with the node's type. A data instance contains data values. An operator instance is a specific function. In a query schema, at least one input data or operator node in the query is constrained only by type.

Working with Remote Sensing researchers, John Adams and Milton Smith, we have formulated a query schema which computes a spectral mixture analysis of a satellite image. The operators in the query are the sub-components of the spectral mixture analysis algorithm. This query schema is shown in Figure 1. In our visual representation of a query graph, ovals represent data and rectangles represent operators. The arrows represent edges and indicate the flow of data between the operators.

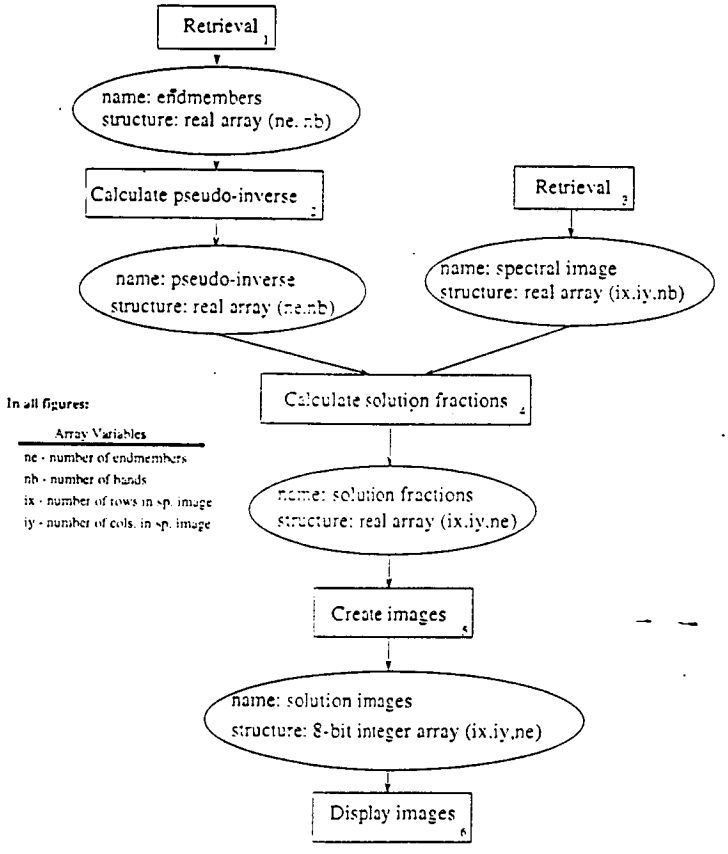


Figure 1. A query schema which computes a spectral mixture analysis of an image.

The central sub-component of the algorithm is a matrix multiply operator labeled "calculate solution fractions". This operator inputs a spectral satellite image and a pseudo-inverse matrix (generated by the "calculate pseudo-inverse" operator) and multiples them together to generate the solution fractions. The "create images" operator converts the solution fractions to grayscale images and the "display images" operator displays them for the scientist.

3.2 Query Instances

A query instance is a refinement of a query schema whose input data and operator nodes are all constrained by instance values. Query instances are executable. Execution of the query creates data values for all non-input data nodes in the query instance.

An example of a query instance is shown in Figure 2. It is a refinement of the spectral mixture analysis query schema shown in Figure 1. A spectral satellite image of a wine growing region in France is used as an input value for the spectral image and five endmembers spectra: alluvial soil, limestone, marls, green vegetation and shade, are formed into an array and used as an input value for the endmember array. All operators in the query instance have a sequential Unix implementation associated with them.

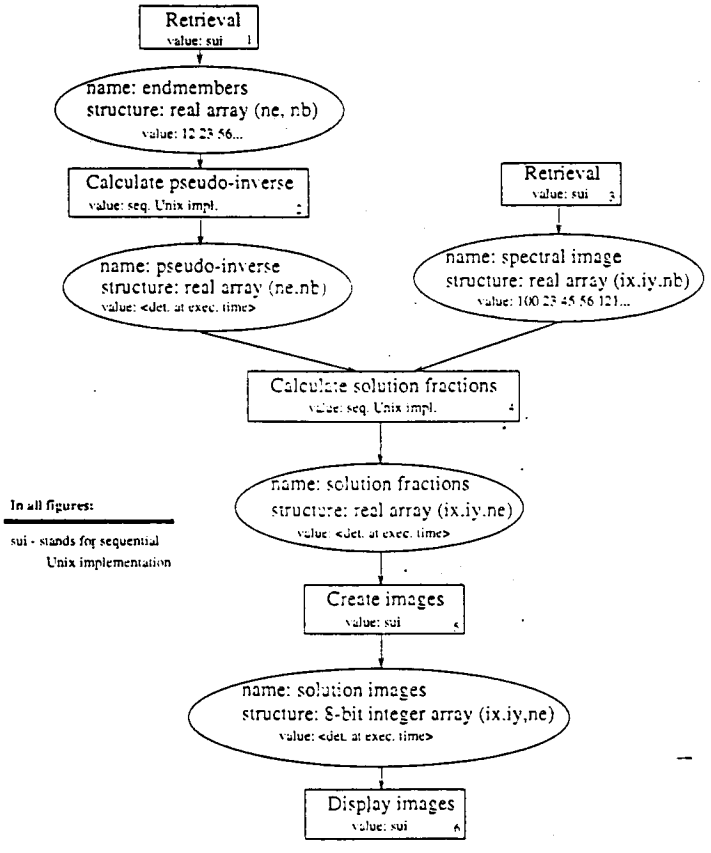


Figure 2. A query instance.

3.3 Experimental Queries

An experimental query is useful for scientists who want to study the effects of using many different data and operator values in a query instance. An experimental query is a refinement of a query instance in which the scientist replaces the values of at least one node in the query with a set of data or operator values to try. This set is called a substitution set. The values in the set must be compatible with the node's type. All possible combinations of substitutions of data and operator values into the query graph are made, creating multiple query instances. These query instances are then executed and their results recorded. Automatically creating and executing these query instances relieves the scientist from having to do this work by hand.

An example of an experimental query is shown in Figure 3. It is a refinement of the query instance shown in Figure 2. A substitution set of two endmembers has been created as well as a substitution set of "create images" operators. The two operators, "sui-ver1" and "sui-ver2" use different grayscale conversion schemes to create the output images.

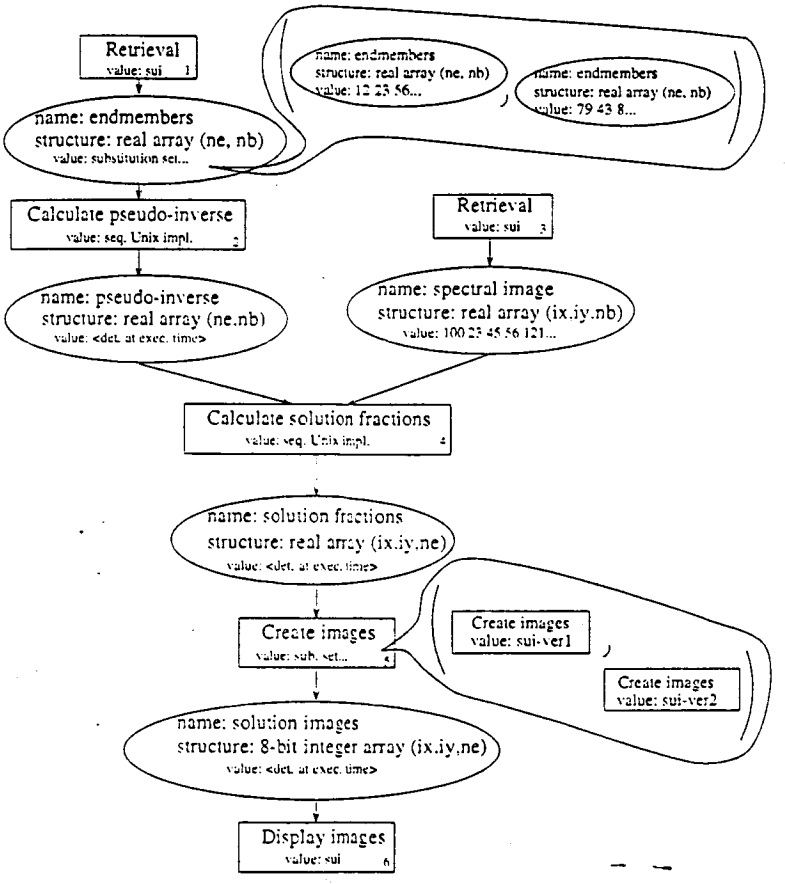


Figure 3. An experimental query.

3.4 Temporal Evolution of Queries

Queries can evolve over time. This evolution can be modeled using a temporal sequence of query graphs. The sequence consists of query graphs interleaved with simple change operators which describe the relationship between the graphs. For example, a simple change can consist of an addition or deletion of a node in the graph. Scientists utilize this historical sequence to understand how a query was derived.

An example of the temporal evolution of a query schema is shown in Figure 4. The sequence shows a simple version of the spectral mixture analysis query schema augmented with a "create images" operator and "solution images" data.

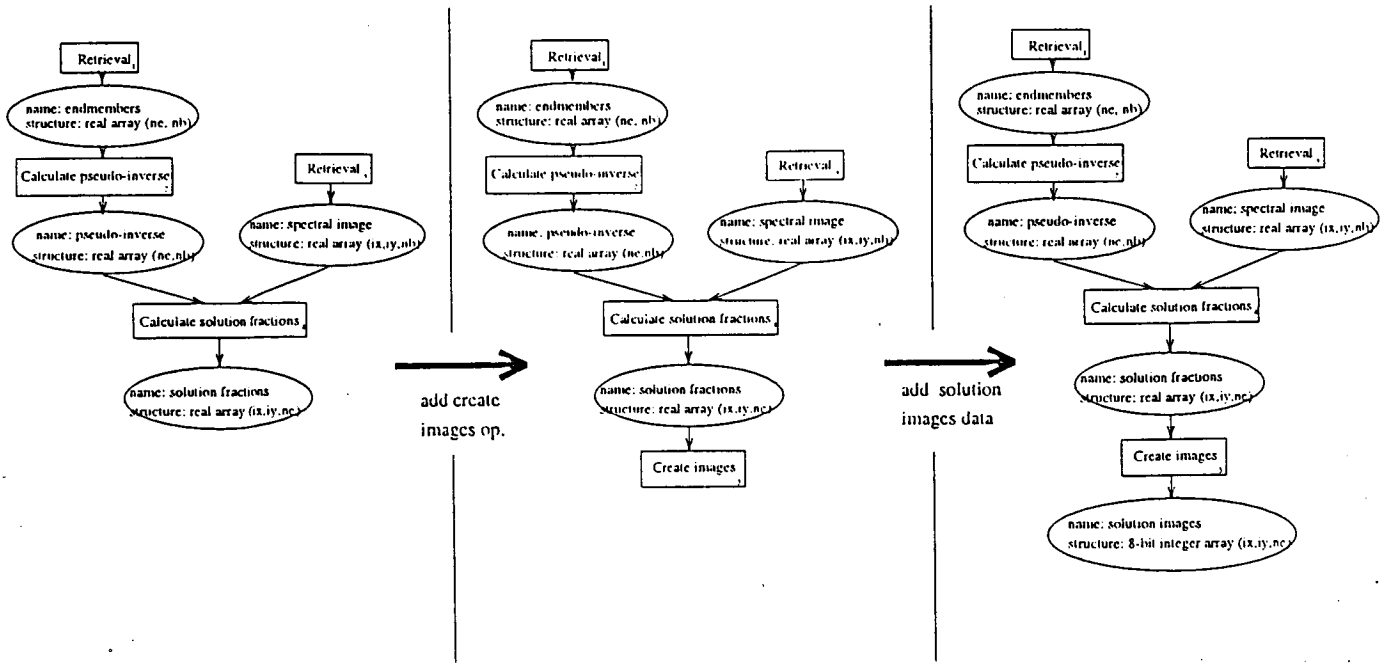


Figure 4. The temporal evolution of a query schema.

4.0 Parallel Execution of Queries

Our current research efforts are focused on understanding how to execute these different types of interactive queries efficiently on parallel computers. We have identified three methods of parallelizing a query: operator, pipeline and data-flow parallelism.

4.1 Operator parallelism

Operator parallelism is defined as the use of parallel instead of sequential implementations of operators in a query. We expect a parallel processing expert to create these parallel implementations for or in collaboration with the scientists. Once a set of operators are parallelized they can be re-used many times in different queries.

For example, some of the operators of the query instance in Figure 5 have parallel implementations (i.e. the "calculate pseudo-inverse" and "calculate solution fractions" operators).

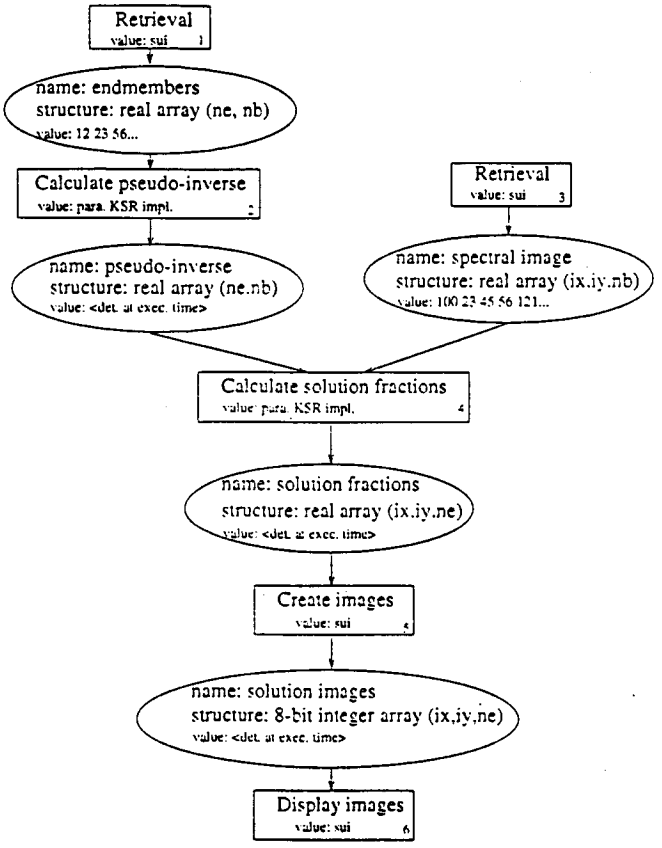


Figure 5. An example of operator parallelism.

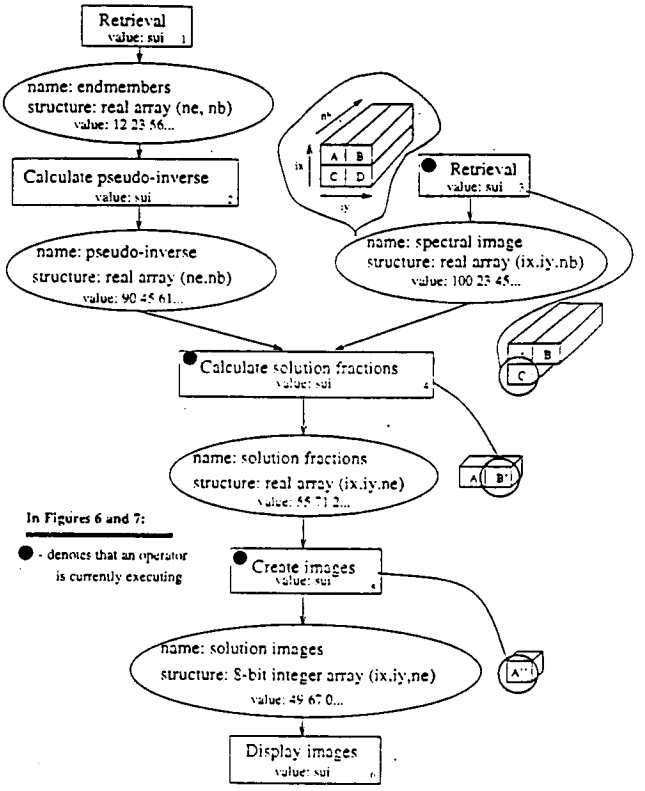


Figure 6. An example of pipeline parallelism.

4.2 Pipeline parallelism

Pipeline parallelism is defined as the parallel execution of multiple operators on different sub-blocks of the input data. The data must be divisible into smaller units. The data-flow semantics of the query graph are modified to allow pipelining of data through operators.

For example in Figure 6 the spectral mixture analysis query is shown with operators 3,4 and 5 pipelining blocks of the spectral image. Operator 3 is processing block C in parallel with operator 4 which is processing block B and operator 5 which is processing block A. Executing the retrieval operator (i.e. operator 3) in parallel with the compute operators (i.e. operators 4, 5 and 6) alleviates the major I/O bottleneck in the spectral mixture analysis query. We are currently studying how to identify when data can be partitioned into blocks for pipelining.

4.3 Data-flow parallelism

Data-flow parallelism is defined as parallel execution of all operators in the query graph with their input data available. A data-flow graph identifies all the data an operator needs for execution; if this data is available, the operator can be executed. Thus, all these operators can be executed in parallel.

For example, in Figure 7 the spectral mixture analysis query schema is shown with two retrieval operators (1 and 3) executing in parallel.

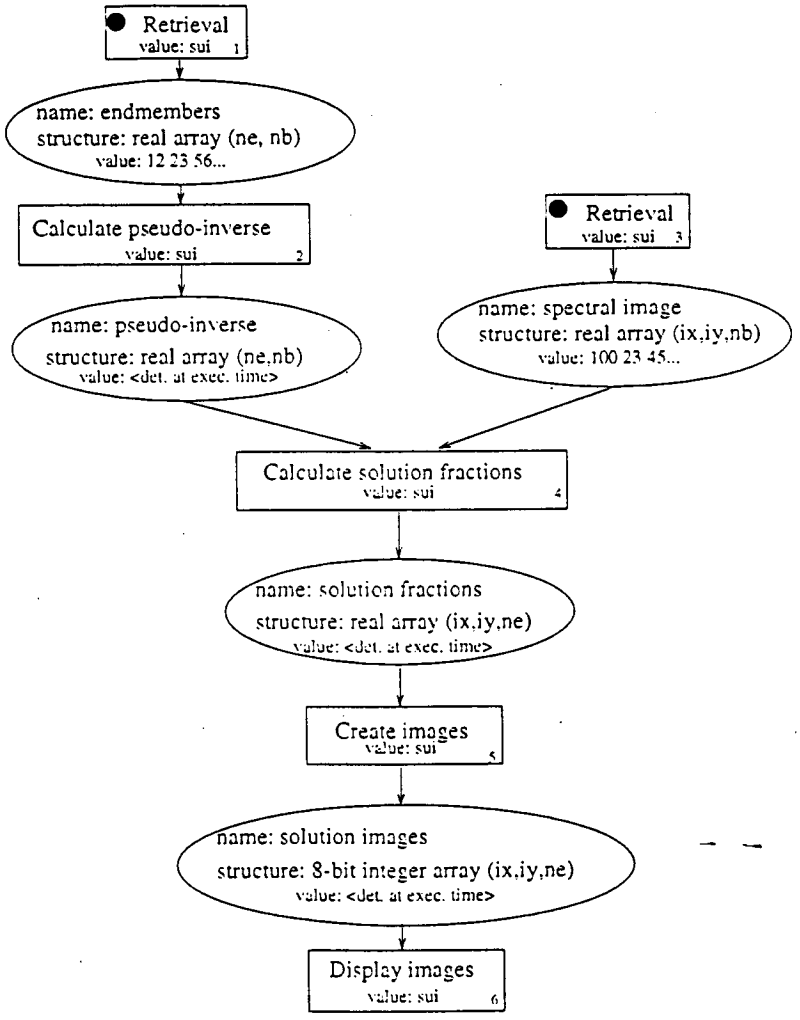


Figure 7. An example of data-flow parallelism.

4.4 Layouts

One conflict that arises when composing parallel operators in a query graph is how to distribute shared data on the processors of a parallel computer. The parallel version of each operator will have a preferred distribution of the data and this distribution may be different for two connected operators. The query optimization mechanism we are designing should utilize the distribution which provides the fastest execution of the query.

A simple example of this type of conflict is shown in Figure 8. The "calculate pseudo-inverse" operator can compute the pseudo-inverse using one of two different distributions. Option 1 distributes the data for each endmember to all the processors while Option 2 distributes the data for an endmember to a single processor. The fastest distribution for calculating the pseudo-inverse is option 1. For calculating the solution fractions, however, it is option 2. A third option, option 3, is also possible, using option 1 to calculate the pseudo-inverse, re-distributing the data to the option 2 distribution, and then calculating the solution fractions using the option 2 distribution. The distribution which results in the fastest execution is option 2 because most of the query execution time is spent calculating the solution fractions. Query execution time is dependent on a variety of factors, such as the size of the input data and the time complexity of the operators. We are working on a cost model which will quantify this information so that decisions about which option to use can be made automatically.

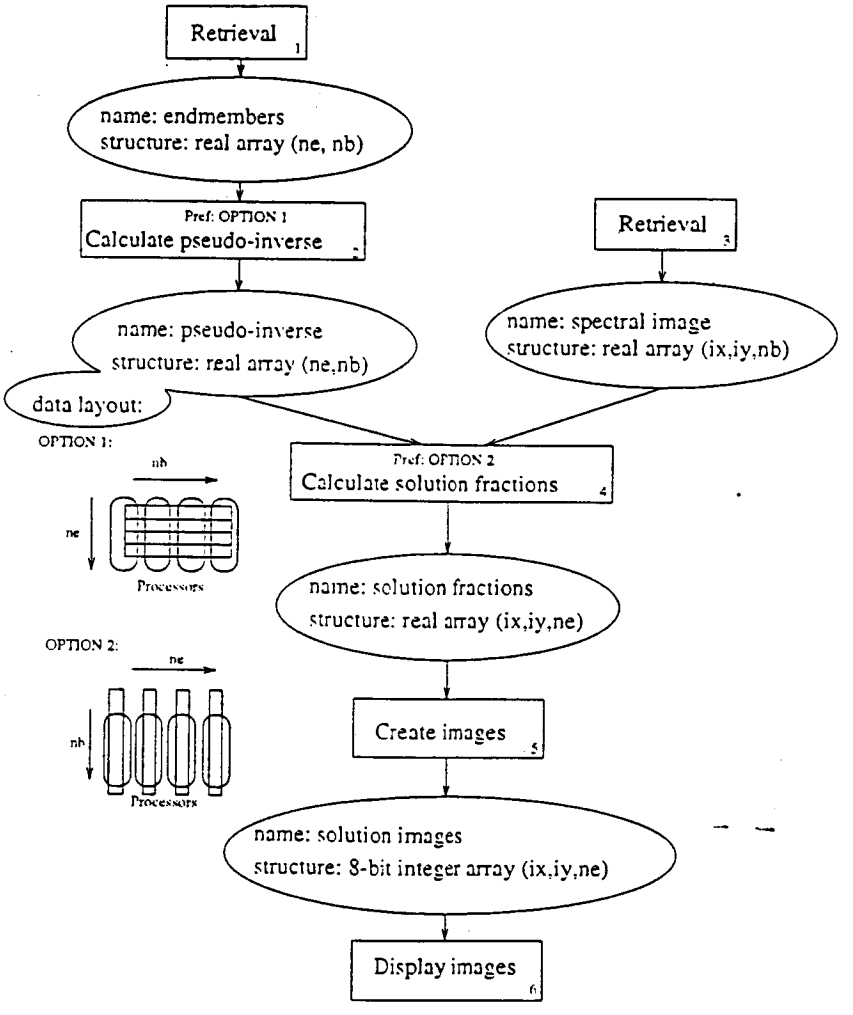


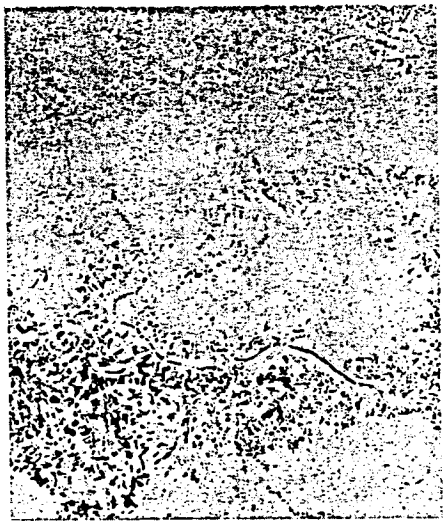
Figure 8. An example of a data layout conflict.

5.0 Current KSR Parallel Implementation

We have implemented a parallel version of some of the operators in the spectral mixture analysis query on our KSR-2 parallel computer. We parallelized the "calculate pseudo-inverse" and "calculate solution fractions" operators. We used KAP, an automatic FORTRAN parallelization tool for the KSR. We then did some optimization of the parallel operators by-hand, modifying and augmenting the parallelization directives output by KAP. The created query instance is shown in Figure 5.

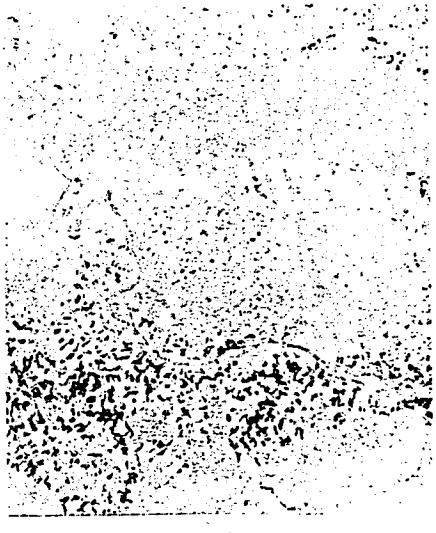
The output solution images, shown in Figure 9, contain a grayscale representation of the fraction of each endmember used to model the image. The value of output fraction image are mapped from the range 0 to 1 to grayscale range from black to white. Thus, a dark gray pixel means a very small fraction of the endmember is used to the model the image. A light grey pixel means a large fraction of the endmember is used to model the image. In order to better understand the output images, a description of the contents of the satellite image is helpful. The satellite image is divided diagonally, northwest to southeast, by a river. South of the river are vineyards and north of the river is a limestone ridge.

Alluvial soil

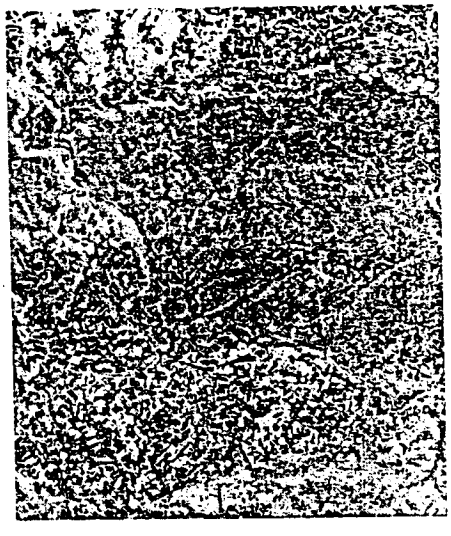


Limestone

Marls



Green vegetation



Shade

Figure 9. Solution fraction images output from our parallel operator implementation of the spectral mixture analysis query.

A graph of the execution times obtained for the spectral mixture analysis query graph using operator parallelism is shown in Figure 10. All execution times are for the same satellite image which is of size 692x614 pixels by 124 bands. Notice that using operator parallelism on the KSR provides a faster execution time, 170 seconds on 20 processors, than the version implemented by the Remote Sensing group which takes approximately 300 seconds. This performance improvement is beneficial for the Remote Sensing group because they can make more tests on more images when they get their results more quickly. A corresponding speedup graph is shown in Figure 11. The speedup is calculated relative to the query's performance on one processor of the KSR.

Figure 10.
Execution time.

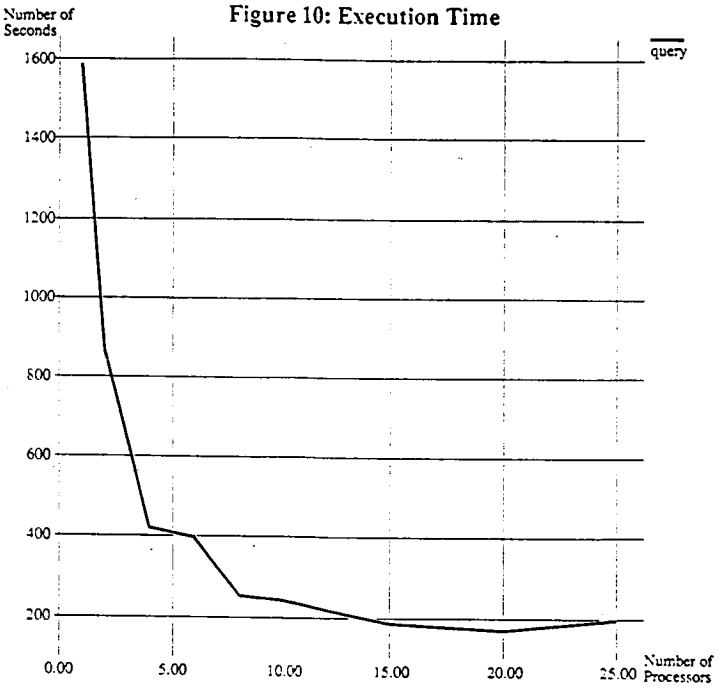
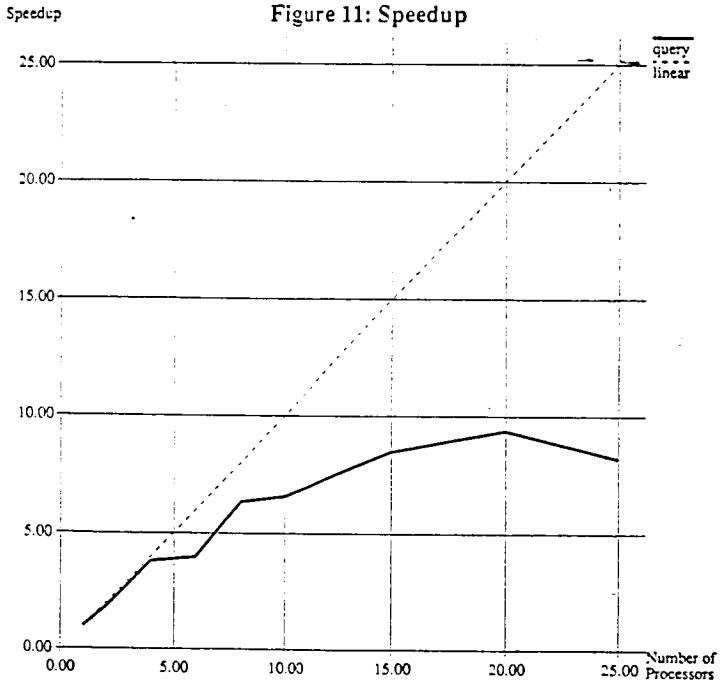


Figure 11.
Speedup.



When comparing the two implementations, it is important to understand how they have been optimized. The Remote Sensing group has done a great deal of low-level optimization including assembly coding. Our operator implementations are written in KSR FORTRAN - a version of FORTRAN that has been augmented with parallelization directives. Further performance improvements are possible with more efficient parallel implementations of the query's operators. Other improvements will result from the implementation of pipeline and data-flow parallelism.

6.0 Work Plan

Our long range goal is to create an automated query optimizer which utilizes all these types of parallelism. We want to insulate scientists from having to understand and make decisions about how to parallelize queries. To accomplish this goal we will work on solving the following sub-goals:

6.1 Identify Optimization Information

We will identify what information is needed to characterize optimization tradeoffs. One example of this type of information is operator execution times for different data sizes. We will also study how to identify the effect of different input data distributions on operator performance.

6.2 Identify Parallelism

We will continue working to identify other types of parallelism in query graphs. We will also characterize the ways in which input data can be partitioned for pipeline parallelism.

6.3 Formulate a Cost Model

We will develop a cost model to represent all this information and create effective algorithms which optimize the execution of the queries on parallel computers.

(Task 22)

UNIVERSITY OF ILLINOIS

High Performance Input/Output Systems for Parallel Computers

Daniel A. Reed
Department of Computer Science

Task Objective

In our original research proposal, we outlined a three phase approach to the input/output problem: a data-driven performance analysis of current application input/output access patterns and requirements, a performance study of existing parallel file systems, and the development of user interfaces for intelligent file caching and migration. In all three phases, the reprocessing of NASA satellite data provides the "focus problem" for the investigation; we have used the NASA SeaWiFS l0tol1a code as a basis for our work.

1.0 Introduction

In the following pages, we report on our annual progress in each of the three areas listed in the abstract. First, in sections 2-4 we describe, respectively, our progress to date in developing application characterization tools, in acquiring operating system software for system instrumentation, and in developing a flexible infrastructure for studying data caching and migration policies. This is followed in section 5 by a description of recent developments on related input/output projects that buttress and support the work supported by this contract.

2.0 Application Input/Output Characterization

The paucity of data on the access patterns of scientific codes on both sequential and parallel systems has motivated our development of tools and techniques to capture, analyze, and understand the input/output patterns of scientific codes. This data is the basis for informed application and input/output software design and is best obtained by software instrumentation of key application codes to capture file system access data.

To provide the necessary infrastructure for data capture and analysis, we have extended (and continue to extend) the Pablo performance instrumentation and data analysis environment [3,4,5] to instrument, capture, and analyze input/output activity. We have used this instrumentation software to analyze the input/output access patterns from the SeaWiFS l0tol1a translation code; see section 2.3 for details.

Below, we summarize our approach to instrumentation, describe the most recent developments to the Pablo instrumentation software, our evolving analysis software, and our plans for continued work. Though it continues to evolve, our input/output instrumentation infrastructure is now ready for installation and use by Goddard scientists–it has been in use at other sites for several months.

2.1 Characterization Levels

In our tri-annual reports, we noted that application input/output characterization is a description of the application's temporal and spatial access pattern, with sufficient ancillary information to compute desired statistics. The temporal component includes the distribution and duration of input/output requests during the code's execution. Similarly, the spatial component is the pattern of data accesses to different parts of one or more files.

The philosophy behind the design of our input/output instrumentation is simple. It should be as easy to use and as efficient as possible, it should minimally perturb application execution behavior, and it should provide the requisite information to improve input/output performance.

This philosophy is reflected in the four levels of input/output instrumentation provided by our software [2]: event traces, temporal summaries, spatial summaries, and temporal histograms.

The levels range from a complete trace of every input/output operation to simple summary statistics (e.g., a histogram of file access delays). Clearly, a detailed trace of every input/output access provides the information necessary to characterize an application's behavior at a less detailed level. However, capturing such traces is not practical for the entire execution of large, long running codes that involve hundreds of thousands of input/output operations — the volume of data and the perturbation induced by the instrumentation are simply too great.

Temporal and spatial summaries, described in our earlier tri-annual reports, allow user-specified grouping of input/output events based on windows of time or file access regions, respectively. Finally, a temporal histogram, provides more detailed information on the distribution of events, while still minimizing the volume of captured data.

2.2 Instrumentation Data Analysis

Instrumentation must balance intrusion and captured data volume against data utility. Although the dynamic data reductions provided by temporal/spatial summaries and temporal histograms can be computed at low cost, and they dramatically reduce the volume of captured data, there are times when more general statistics are needed. More often, the desired performance metrics are not known until after an initial examination of the detailed trace data. In such instances, a profile that includes information about temporal and spatial input/output patterns provides an application "thumbprint" that can be easily compared to the profiles of other applications. We developed, and continue to develop, an analysis program that computes input/output profiles from a combination of detailed input/output traces and spatial/temporal summaries.

2.3 SeaWiFS Analysis

Using our instrumentation and data analysis software, we have studied the input/output behavior of a version of the NASA Goddard SeaWiFS l0tol1a code. This code sequentially reads a large file of raw satellite data, partitions the data into logical segments, and constructs archive products in NCSA's Hierarchical Data Format (HDF) [1]. Below, we briefly describe the results of our analysis.

We captured and analyzed the behavior of the l0tol1a code, using NASA test problem input, on a Sun SPARCsystem 600 with a local SCSI disk. Because this code reopens a small number of files hundreds of times, we minimized the volume of instrumentation data by summarizing the input/output trace events via spatial summaries of file lifetimes; a single input/output trace record is generated for all the input/output activity to a file between its open and close.

Figures 1 and 2 show a portion of our analysis, based on processing one frame per input/output operation–the default behavior of the code. The two figures, and Table 1 confirm that input/output is dominated by small file reads and writes. Because the HDF library is stateless, to process each HDF input/output request, the

underlying file must first be opened, then the HDF meta-format descriptors are be read or written, and finally the data is read or written. Although this bottleneck is ameliorated by processing the data in multiple frame increments, the qualitative behavior is intrinsic to the design of the HDF library.

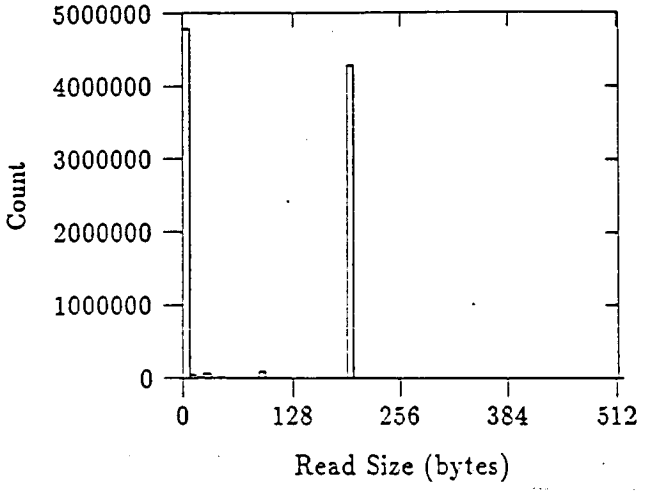


Figure 1. Read 10to11a Write Size Histogram

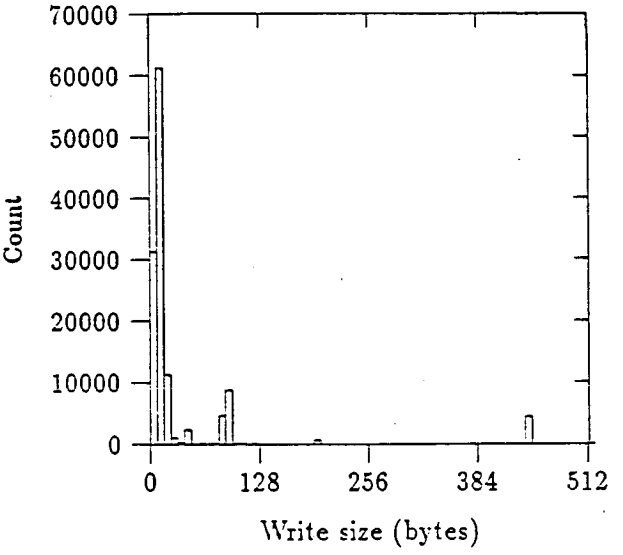


Figure 2. SeaWiFS 10to11a Write Size Histogram

		Operation Size		
Operation	Count	Mininum	Maximum	Mean
Read	9,299,936	0	21,504	112.6
Write	141,166	0	20,608	759.9
Seek	9,745	0	23,322,465	848,583.6

Table 1. SeaWiFS 10to11a Operation Size Statistics

Because HDF is the standard representation format for SeaWiFS data, we have begun instrumentation discussions with the developers of HDF at NCSA. They have expressed great interest in our preliminary data because it provides a basis for software optimization design decisions. We are working with the HDF developers with the goal of developing performance instrumentation that can be distributed with the HDF library source code. Such instrumentation would allow SeaWiFS developers to obtain input/output performance data merely by linking their code with an instrumented version of the HDF library.

3.0 Operating System Input/Output Studies

In April 1994, Intel shipped us the OSF/1 source code for the Paragon XP/S. We have begun studying the code to identify the best places to instrument system-level instrumentation. Our goal remains the capture and recording of the the pattern of physical input/output operations (i.e., the requests to storage devices). Because the file system mediates application input/output requests and generates physical input/output requests in response to application demands, the combination of physical and application input/output requests provides the data needed to assess file system performance.

We believe system instrumentation software developed for the Intel Paragon XP/S at Illinois can be used with little or no change on NASA JPL Paragon and the larger Caltech Paragon. We have already installed our application instrumentation software at Caltech and have begun collaborative instrumentation of their input/output intensive codes as part of an NSF Grand Challenge group. Moreover, an extension of this operating system instrumentation effort is an integral part of the planned scalable input/output initiative; see section 5.2.

4.0 Flexible File Caching Policies

Our limited experience with parallel systems and multiple disk arrays suggests that a single, system-imposed file system policy is not likely to yield good performance for all applications. Rather than a single technique for data distribution and management, we are developing a flexible file system architecture that will exploit application-specific knowledge of file access patterns to specify efficient file caching algorithms and distributions of file blocks across storage devices.

To test these policies, we continue to implement a portable parallel file system (PPFS) built atop standard Unix file systems. A parallel file consists of a collection of standard Unix files, each of which contains some portion of the file. These Unix files are distributed across the secondary storage devices of the parallel system; our initial target architectures are a workstation network and the Intel Paragon XP/S.

As described in our earlier reports, our parallel file system design is based on the client/server model. The servers are input/output nodes that manage storage devices and respond to client (application) requests. The file system itself consists of file management policies that are realized by the servers and the clients. Caching and prefetching may occur at the input/output nodes, the clients, or at a designated caching node, called the caching agent. Most file system policies can be changed dynamically during application execution (e.g., changing a prefetching algorithm or increasing client cache sizes). Using the server and client infrastructure, the parallel file system supports a parallel file abstraction. Like standard Unix files, parallel files can be created, removed, opened, closed, read, and written. Interfaces for file open and access support specification of the distribution of file blocks across input/output servers, the desired prefetch policy (both on the client and server sides), the expected access pattern (e.g., disjoint sequential), and the file consistency algorithm (if needed). Our premise is that given an associated set of file data management alternatives, the application developer can tune input/output management to the application, rather than tuning the application to the input/output system as is common now.

During the past year, we have implemented an initial version of the framework described above. We have based our distributed memory implementation on the Intel NX communication primitives. NXLIB [6], a socket-based version of these primitives, allowed us to develop and test the file caching software on a workstation network. This greatly reduced development time by allowing us to exploit high-performance workstations for compilation and testing.

At this writing, the file system software is being tested on the Intel Paragon XP/S using an input/output intensive parallel genome sequencing code. We plan to begin detailed experiments with specific caching and prefetching policies shortly.

5.0 Related Work

Several other projects naturally complement the work supported by this contract. In particular, the Pablo performance analysis environment provides a natural tool base for capturing and analyzing the input/output access patterns of sequential and parallel application codes. In addition, our work on data immersion and parallel input/output provide support staff and software development infrastructure. Below, we briefly describe these research projects and how they provide leverage for our current work.

5.1 Application Characterization

Our application characterization efforts are supported by a complementary, ARPA-sponsored project that supports a post-doctoral associate and a staff software developer; its primary focus is extending the Pablo performance tool suite to support the capture of application input/output access patterns and conducting a broad-based application input/output characterization study. Working with the research assistant supported by this contract, the staff have developed the instrumentation infrastructure needed to capture and analyze input/output patterns of codes written in both C and Fortran, executing on both single processor UNIX systems and the Paragon XP/S.

In addition, we are part of an NSF grand challenge team with Caltech to analyze parallel applications on the Caltech Paragon and to instrument a set of domain-specific input/output libraries and object-oriented databases, developed by our collaborators, to understand the dynamics of their responses. We have used our input/output instrumentation at Caltech to capture data from one production application, an electron scattering code for low temperature plasmas. This interaction has allowed us to test and validate our instrumentation software, and to identify scalability problems on large (hundreds of processors) hardware configurations. We are ready to begin similar collaborations with Goddard scientists.

5.2 Scalable Input/Output Initiative

During the past eighteen months, an active group of researchers has coalesced to initiate a broad-based, concerted attack on the input/output problem. This group, originally organized by Paul Messina from Caltech, has proposed a wide-ranging research program that includes application and system input/output characterization, networking, file systems, persistent object-stores, compiler and language support, and basic operating system services. At this writing, ARPA, DOE, NSF, and DOE have orally committed to funding the project.

As part of this project, Caltech would upgrade the Intel Paragon XP/S to include 256 disks and multiple HiPPI interfaces. In addition, the 128 node IBM SP-1 at Argonne National Laboratory would serve as a second input/output testbed. The focus of our proposed work, which will be funded by NASA, is broader scale application and system input/output characterization (i.e., a unification of application and physical input/output data). NASA funding will support additional staff, and will allow us to expand and accelerate the scope of the work in our current contract.

5.3 Data Immersion

An NSF-funded project is extending the Pablo performance analysis environment to include data immersive performance presentation techniques. Instrumenting massively parallel systems can result in megabytes or gigabytes of dynamic performance data, far more data than can be assimilated by direct examination.

The primary goal of the data immersion project is to increase user interaction modality by immersing the performance analyst in dynamic performance data via a head-mounted display, tracking system, and three-space audio cues. Because performance data often is of high dimension, involving many metrics with widely varying range and characteristics, this "virtual reality" differs substantially from more common applications of data immersion technology.

At present, we have an operational virtual reality system that immerses users in dynamic performance data and permits real-time adaptive control of application parameters. We have tested this data immersion system using a parallel genome sequencing code with controls for input/output access granularity, and we will be demonstrating the software at the upcoming ACM SIGGRAPH VROOM (virtual room) display. In the coming months, we will be extending this infrastructure to include additional data presentation metaphors and a wider variety of application codes.

References

[1] NCSA HDF, Version 3.2, University of Illinois at Urbana-Champaign, National Center for Supercomputing Applications, Feb. 1984.

[2] Noe, R. J. *Pablo Instrumentation Environment User's Guide.* Tech. rep., University of Illinois at Urbana-Champaign, Department of Computer Science, Aug. 1993.

[3] Reed, D. A. Performance instrumentation techniques for parallel systems. In *Models and Techniques for Performance Evaluation of Computer and Communications Systems*, L. Donatiello and R. Nelson, Eds. Springer-Verlag Lecture Notes in Computer Science, 1993.

[4] Reed, D. A., Aydt, R. A., Noe, R. J., Roth, P. C., Shields, K. A., Schwartz, B. W., and Tavera, L. F. Scalable Performance Analysis: The Pablo Performance Analysis Environment. In *Proceedings of the Scalable Parallel Libraries Conference*, A. Skjellum, Ed. IEEE Computer Society, 1993.

[5] Reed, D. A., Olson, R. D., Aydt, R. A., Madhyastha, T. M., Birkett, T., Jensen, D. W., Nazief, B. A. A., and Totty, B. K. Scalable Performance Environments for Parallel Systems. In *Proceedings of the Sixth Distributed Memory Computing Conference* (1991), IEEE Computer Society Press.

[6] Stellner, G., Lamberts, S., Bode, A., and Ludwig, T. *Nxlib — Paragon Parallel Programming Environment on a Network of Workstations.* Tech. rep., Institut fur Informatik, 1994.

(Task 23)

UNIVERSITY OF MINNESOTA

Fast I/O for Massively Parallel Applications

Matthew O'Keefe
Department of Electrical Engineering

Thomas Ruwart, Paul R. Woodward
Army High Performance Computing Research Center

Task Objective

I/O bandwidth and storage capacity have not kept pace with the processing speeds of today's high performance computing machines. This issue has become even more pressing with the arrival of massively parallel processing, which has the potential to achieve sustained computational rates of tens to hundreds of Gigaflops in the near future on machines with very large memories. These machines will allow very detailed studies of large, complex physical phenomena; understanding the results of these large calculations requires a sophisticated graphics and high performance I/O environment, such as the one found in the Graphics and Visualization Lab at the University of Minnesota. The work undertaken in this project involves innovative I/O research that builds on our existing strong graphics and I/O foundation, relates very closely to several major areas of NASA interest , and capitalizes on the unparalleled high performance computing infrastructure available to us at Minnesota.

1.0 Disk Array Research

The signal achievement for the contract this year was the achievement of 510 Megabytes per second sustained data transfer rate from our workstation-based (SGI) disk array. This experiment — known as the M.A.X. (Maximum Achievable Transfer) project — was a collaboration between University's Graphics and Visualization Lab, Ciprico, and Silicon Graphics to determine the maximum possible transfer rate through a high-end SGI workstation server. Ciprico provided 24 disk controller boards and test setup that allowed us to build a "virtual" disk array with 31 RAID 3 devices, 24 of which had no drives at all! The drive behavior was simulated within the disk controller firmware; as part of the project Ciprico modified their controller firmware to achieve this. A paper summarizing these results will be made available as part of this final report.

We have implemented and tested the performance of a single-level disk array based on the latest Seagate 7200 rpm 2-Gigabyte Barracuda drives. This single level array is similar to configurations at Goddard; we are currently finishing a paper that describes these results.

Jeff Stromberg has written a file system utility that allows a system manager to measure disk fragmentation for a file system. File fragmentation plays a large role in reducing performance for large transfers. We are using this program to measure fragmentation on GVL file systems across a variety of systems to gain some understanding of this effect.

Steve Soltis and the I/O group are concentrating on testing and instrumentation of SGI's new xFS file system which provides new capabilities that we intend to exploit and study in the coming year.

2.0 NASA Interactions

In June Tom Ruwart and Matt O'Keefe traveled to NASA Goddard, where Tom reported on the MAX results and Matt described recent work in oceanographic modeling and its impact on our I/O work. We had much interaction with the Goddard storage team and in particular we obtained a copy of their requirement specs. Several visits earlier in the year were described in previous reports.

Tom Ruwart is currently working with Lisa Hamet of NASA in getting her and several other NASA researchers time on the AHPCRC CM-5.

3.0 The Future

We intend to continue our I/O experiments in the primarily SGI hardware domain of the GVL. For year two of the project we intend to instrument the new xFS and IRIX 6.0 operating system to obtain even more detailed performance information to aid in system design and improvement. We hope to visit NASA Goddard again in the near future to continue the early fruitful interactions we have had.

Furthermore, based on our meeting with Milt Halem at NASA Tom Ruwart is pursuing Hierarchical storage system performance evaluation on a high-performance Ampex tape library subsystem. The trip to Goddard helped us to better focus our efforts on problems in this area that are mutually beneficial to NASA and the AHPCRC.

Both Steve Soltis and Jeff Stromberg are now full-time graduate students on the project.

4.0 References

[RuO93] Thomas M. Ruwart, *QM.A.X.: The Maximum Achievable Transfer Experiment, University of Minnesota, May 1994.

(Task 25)

DUKE UNIVERSITY

Image Compression: Algorithms and Architectures

John Reif, Steve Tate, Tassos Markas
Department of Computer Science

1.0 Objective

This project has been concerned with research efforts in the development of data compression algorithms and architectures. Our goal was to provide NASA with some innovative solutions for handling large volumes of information in terms of software programs that can be executed within a reasonable amount of time given today's technology, and hardware devices that can perform real-time data compression. The key difficulties addressed by our research that characterize NASA's data compression problem are (1) the diversity of types of data including multispectral image data, video data, audio data, and large computer programs, each with distinct compression (lossy or lossless) and quality requirements which depend critically on the subsequent scientific processing to be done, (2) the need for adaptive compression algorithms which modify and improve their compression in response to changes in the data stream, and (3) the need for controlled compression algorithms that allow the tradeoff between the amount of compression and the quality of the compressed data to be adjusted to the NASA mission's specific scientific computing requirements.

In response to these problems our approach has been to develop data compression techniques with the objectives that they (1) are efficient, in that we provide faster new algorithms and improvements to known data compression methods, and exploit the speedup of parallel processing, (2) achieve as high a compression ratio as possible for the given degree of fidelity required by the scientific processing application, (3) allow adjustable compression and quality ratio, (4) are dynamically adaptable to changing input data, (5) address the unique compression aspects of multispectral data, and (6) support hierarchical and progressive decompression techniques needed for browsing and preview.

2.0 Accomplishments

During this research effort we devised several novel and improved data compression algorithms and investigated their theoretical aspects as well as evaluated their performance against existing methods, designed parallel architectures for lossy and lossless compression, and developed software programs to implement several of the new algorithms and to simulate various hardware architectures (in addition to software for basic existing compression methods and basic image processing functions that were necessary for our study). A number of papers describing work on this project have been accepted and presented at international conferences, and notably we aided in instigating the first national conference on data compression (DCC) as well as the organization and operation of a recent workshop on data compression for scientific data. In overview, our significant results include the development of:

Algorithms for lossless and lossy compression:

- Dynamic compression schemes for vector quantization
- Distortion-controlled compression algorithms
- Multispectral image compression algorithms, both

- lossy using techniques such as discrete wavelet transform, and
- lossless using band ordering
- Lossless compression of edge map files

Parallel architecture designs for lossless and lossy compression:

- Architectures for real-time lossless compression and decompression
- Systolic architectures for parallel tree-structured vector quantization
- Optical techniques for image compression

Software to:

- Implement basic image processing tasks such as image display, image partitioning, image conversion and image comparison, and existing compression algorithms such as various vector quantizers including the k-means training algorithm, the laplacian pyramid coder, and various transform methods.
- Implement and evaluate performance of the newly developed algorithms
- Simulate hardware architectures.

The remainder of this report briefly summarizes the results and progress in each of these areas, with references to where full descriptions may be found. We begin, however, by first describing two significant events —the completion of a dissertation on compression supported under this contract, and the organization of the first national conference on data compression (DCC) as well as a recent workshop on data compression for scientific data.

3.0 Significant Events

3.1 Conference and Workshop on Data Compression

The First Data Compression Conference (DCC 91) was held at Snowbird, Utah, in 1991, instigated and co-chaired by John Reif and James Storer. This was the first national conference on data compression, and is now held annually. It was organized with the assistance of CESDIS, in particular Nancy Campbell.

Also in 1993 we organized and held the "Workshop on Data and Image Compression Needs and Uses in the Scientific Community" at the Goddard Space Flight Center. The workshop was organized in participation with Jim Tilton of NASA, and publicized and assisted by the CESDIS office at Goddard. This workshop was designed to enhance communication between researchers in data and image compression, and the potential users of data compression in the scientific community. The goals of the workshop were twofold: education and collaboration.

The workshop hosted 10 presentations, with talks on data compression by Tassos Markas, Jeff Vitter, Irving Linares, Edward Seiler, and Manohar Mareboyana, talks on scientific data use and collection by Gene Feldman, Mary James, and Jim Pfaendtner, and related talks on managing large amounts of scientific data by Kan Salem and Robert Cromp. There were 58 registered attendees of the workshop, with affiliations from industry, government, and academia.

The meeting was a productive one with contacts made between physical scientists and compression researchers. Contacts were made with Immanuel Freedman of the Cosmology Data Analysis Center who was interested in the vector quantization work developed by this project. Contacts were also made with the various users of scientific data who will supply test data for our compression projects. The discussion period was useful in understanding the needs of the scientific users, and was a motivation for the work in lossless compression of multispectral images described earlier in this report.

A full report of the workshop is available as a CESDIS Technical Report [Tat93b] .

3.2 Dissertation

In the spring of 1993, the graduate student research assistant on the project, Tassos Markas, finished his dissertation involving most of the work done on this project. His dissertation, entitled *Data Compression: Algorithms and Architectures*, has been sent to CESDIS and should appear in the CESDIS dissertation series [Mar93].

4.0 Algorithms for Lossless and Lossy Compression

4.1 Dynamic compression schemes for vector quantization

A significant effort in the area of lossy schemes was the development of dynamic vector quantizers. We modified the basic lossy compression algorithm for vector quantization to incorporate dynamic training and encoding, thus capturing the temporal locality of the incoming images. These schemes build their vocabularies in a dynamic way from the input data. In addition pruning features have been implemented to dynamically adjust the vocabularies so that we can capture the temporal characteristics of large image databases. The pruning algorithm uses the average cumulative MSE of each subtree for deleting sections of the binary tree that do not reduce significantly the distortion of the reconstructed image with respect to the number of bits required to encode this information.

At the same time we evaluated several preprocessing methods that were used to transform images in a different domain of representation to increase the compression efficiency of the quantizer. The evaluated methods include the difference transform, the pyramid coding, and the 2-dimensional Discrete Fourier Transform.

Another research effort was to speed-up the computation of the classic tree-structured vector quantizer (TSVQ) scheme. A number of techniques were implemented and evaluated. We compared the full-search scheme with the tree-structured vector quantizers (TSVQ), the variable-length TSVQ with the fixed-length TSVQ, the multidimensional k-d trees with the TSVQ. From these schemes, the variable-length TSVQ gave the best SNR (Signal-to-Noise ratio) over execution time factor. We also developed some other algorithms that trade off distortion for speed. The results of this effort are summarized in a survey report titled Fast Computations of Vector Quantization Algorithms that was submitted to NASA [MR9 1 b] .

In addition, we developed methods for compression of constant coefficient PDEs, in collaboration with Victor Pan.

4.2 Distortion-controlled compression algorithms

We have developed a new class of distortion-controlled compression algorithms that are capable of compressing images at various rates so that the reconstructed images meet certain distortion criteria. These Distortion Controlled Vector Quantization methods (DCVQ), more specifically, are techniques capable of controlling the amount of information that is lost by trading off between compression rates and distortion. Three such algorithms are summarized in [MR9 1 a, MR92] .

The basic idea of the distortion-controlled algorithms is to iteratively approximate an image block using the current approximation and a best-match block obtained from a local vocabulary (vector quantization). This operation is executed recursively until the overall distortion of an image block is below a given value. The first DCVQ method that was developed is the multiresolution algorithm (MRVQ) that utilizes different size image blocks to encode efficiently areas with different information content. The second algorithm encodes recursively the error block, defined as the difference between the current approximation and original block, to approximate the original block. The third algorithm, the Error-Coding Multiresolution algorithm, combines the features of the first two algorithms in that it uses both quad-trees and vector quantization of the difference between original vector and current approximation.

The Error-Coding Multiresolution algorithm achieves better performance compared to the other two algorithms and it outperforms the classic vector quantization algorithm in terms of better image quality at high bit rates and wider range of distortion/compression performance. The distortion-controlled algorithms achieve a wider distortion/compression range than the combined range of the 2x2, 4x4, and 8x8 vector quantizers. The compression ratios of these algorithms range from low rates, suitable for quick browsing of of large amounts of image data, to high rates for reconstructing accurately images with high information content.

We also managed to improve the performance of multiresolution algorithms by using two different Huffman coders to losslessly compress the quad-tree representations and the vector quantization indices. We have been also investigating how to incorporate transform coding methods as a preprocessor in an attempt to exceed the compression/distortion performance of the JPEG standard .

4.3 Lossy multispectral image compression algorithms

Over the last three years great progress was made in the lossy compression of multispectral images. In particular, we have developed algorithms that exploit both spectral and spatial redundancy within multispectral images, and combined these methods with a novel hierarchical encoding of the data, which gives multispectral compression with excellent results. When tested on the Landsat Thematic Mapper data of the Washington, D.C. area, compression ratios of 20-30:1 were obtained with perceptually lossless quality, and over 100:1 compression was achieved for browse quality.

The spectral redundancy is removed by a two-step process: first, the bands are put through a histogram equalization process to minimize the variance along the spectral domain, and then a one-dimensional transform is applied along the spectral dimension of the data. In general, histogram modification is an irreversible process, but by posing several constraints to equalizing process we can define a reversible transformation that can be used to reconstruct the original image without any loss of information. After equalization of the bands, a one-dimensional transform such as the KLT (Karhunen-Loeve Transform) is applied to the values of all bands at each pixel location.

After removing spectral redundancy as described above, two-dimensional wavelet transforms are applied to remove spatial redundancy. The resulting coefficients are fed into a uniform scalar quantizer, where different levels of the wavelet coefficients are quantized independently. To encode the locations of the significant coefficients after quantization, a novel hierarchical approach is used. Simulation results show that the hierarchical encoding process requires 5-10% less space than the more traditional run-length encoding process. As a final stage of the compression process, the coefficients and bitmap codes were encoding using a lossless entropy encoder, for which we tried both a Huffman coder and an arithmetic coder.

A paper describing some of the results of this work was presented at the 1993 Data Compression Conference [MR93].

4.4 Lossless multispectral image compression algorithms

During the last year we also began work on a new project on lossless compression of multispectral images.

We used the new CM-5 parallel computer at Duke to analyze the large amount of data in multispectral imagery (up to 210 bands of AVIRIS data) . The compression process is divided into three parts: band ordering, modeling, and coding. The modeling and coding phases are based on previous work in lossless image compression, using linear prediction and contextual arithmetic coding, respectively. Differences with previous work exist due to the fact that for multispectral compression, it is vital to determine the linear prediction coefficients for the particular data being compressed, rather than relying on fixed coefficients as is typically done in lossless image compression (or relying on a small set of possible coefficients as is done in the lossless JPEG compression scheme). It has also been determined that assumptions made in typical image coding, such as the assumption of a Laplacian distribution for prediction errors, do not hold for some types of satellite data. Removing this assumption adds some complexity to the coding phase, but can have dramatic results: the compression ratio for 5 bands of CZCS data went from 2.2:1 to 2.9:1.

The truly unique part of this work is the work on band ordering. This phase of the compression method can be examined independently of the modeling and coding phases, and the work of this project will be applicable to any lossless compression scheme that uses interband relations in the coding of multispectral images. In particular, we give an efficient algorithm for computing the optimal coding ordering of the bands of a multispectral image, given any modeler and coder. Previous methods for such problems have performed exhaustive search on the band orderings to compute such an order, requiring n ! time to find the optimal ordering. Such an exhaustive search algorithm run on a Sparc 2 would take approximately 10"(386) years (longer than the age of the universe!) to compute the optimal ordering of the 210 bands of AVIRIS data; the newly designed algorithm finds the optimal ordering in less than 20 seconds on a Sparcstation 2. We also examine the ordering problem with some restrictions placed on the ordering that allow for random access to bands within the compressed file. We have proved that finding an optimal ordering under such restrictions is NP-hard, so is computationally infeasible.

The algorithms have been implemented, and experiments are now being run on NASA data. A paper describing this work and the results obtained is now in preparation, and will be submitted for publication once it is completed [Tat93a] ..

4.5 Lossless compression of edge map files

We have also worked in a collaborating effort with other data compression researchers at NASA to develop lossless compression techniques for bitmaps defining regions of images. The outcome of this effort was a context-based statistical modeling technique that was fed into an arithmetic coder. This technique showed a significant improvement over previous methods.

The compression method we have designed for the edge map files is a simple adaptive context modeler fed into an arithmetic coder. For a pixel location's context, we use the pixel locations immediately above and to the left of the current location. It is an important property of edge maps that we are allowed to have such meaningful contexts with a relatively small number of different possible contexts (so that the gathered statistics are reliable). The modeler keeps track of how many times each pixel value has occurred in each context, and these counts are used as the prediction probabilities for the arithmetic coder.

Five sample edge map files were used, all of which were produced from the same original image (a Landsat Thematic Mapper image of the Ridgely area), but with various quality thresholds. The highest quality image file is labeled eOO and the quality decreases in the progression of files that ends at eO4. According to Tilton, the quality represented by eO4 is typical of the reconstructed image quality desired in his study. For comparison, we compressed all the files with the various compression methods available in the "crush" compression package, which includes LZC (the standard UNIX compress utility), WNC (the Witten, Neal, and Cleary arithmetic coder), ADAP (an arithmetic coder with an adaptive model), and LZRW3A (a fast variant of Lempel-Ziv compression due to Ross Williams). In addition, we used the lossless JPEG compression method in an attempt to draw out some of the two dimensional dependencies in the data. For all test files, our new compression program attained higher compression ratios than any of the previous methods. It should be also noted that the savings on the file eO4 is particularly impressive, and that this is the file described as typical by Tilton.

5.0 Parallel Architectures for Lossless and Lossy Compression

5.1 Architectures for real-time lossless compression and decompression

Our effort in the area of real-time data compression systems has been focused in defining a new architecture for hardware textual compression that improves on the earlier Lossless Data Compression and Decompression (LDCD) system [SRM90] by increasing the compression speed and by minimizing the size of implementation (the number of ASIC's required to implement a full compression system). The outcome of this effort was the development of a new systolic-type parallel architecture that utilizes Content Addressable Memories (CAM's) to perform parallel dictionary matching. The data compression algorithm that is implemented by this architecture is a generalization or the universal LZW (Lemplel, Ziv, Welsh) compression algorithm and the original LDCD algorithm that is described in [SRM90].

Several simulation experiments on large text files showed that the compression performance of the implemented algorithm achieves equivalent to LDCD performance when it is implemented with 32-64 processing elements, each having a local dictionary of 128-64 entries, respectively. The circuitry required to implement this algorithm can be possibly included in a single ASIC (compared to 30 required at the LDCD) thus significantly reducing the cost of the system. In addition, higher compression speeds can be achieved using this architecture because of the fast matching speed of the CAM devices.

A hardware simulator of the CAM-LDCD system, for both the encoding and the decoding functions, has been also completed. All this work is summarized in [MRS93].

5.2 Systolic architectures for parallel tree-structured vector quantization

We have also designed two systolic-type parallel architectures for data compression that implement a parallel version of the tree-structured vector quantization algorithm. The proposed architectures are based on a parallel, memory-shared system organization that offers significant advantages compared with the memory organization of other vector quantization systems. In particular, previous designs have required memory sizes that grow exponentially with their position in the systolic pipe. The new design uses a pool of shared memory so that the stages of the systolic pipe are more uniform, and access is granted to the PEs in the pipe using a time-slotted protocol. The scheduling of these accesses is timed such that all processors remain busy.

This effort was pursued in two phases: A simplified version of this design has been implemented in a single board using standard parts to demonstrate the feasibility of such system. This design has been completed and has been demonstrated in a PC using a standard interface that was build for this purpose. During the second phase, we developed a high performance architecture that is capable of compressing data at extremely high rates. This system has been designed around an Application Specific Integrated Circuit (ASIC), and it offers a high degree of flexibility compared with other existing designs. The architecture is capable of implementing the fast binary-searched vector quantization algorithm using variable size image blocks and variable size vocabularies. Both systems have been designed using the Mean Squared Error (MSE) distortion measure.

A paper that describes the parallel vector quantization architecture along with critical design issues was presented at the 1993 Image Coding Symposium [MREE93].

5.3 Optical techniques for image compression

Optical computing has recently become a very active research field. The advantages of optical devices is their capability of providing highly parallel operations in a three dimensional space. In this research effort we have investigated the optical implementation of a variety of data compression techniques such as transform coding, vector quantization, and interframe coding. The main outcome of this research is that many transform coding methods, such as the cosine transform, can be implemented by a simple optical system, and the operation can be carried out in constant time.

This work has been also focused in the implementation of vector quantization using holographic associative matching. Holographic associated matching provided by multiple exposure holograms can offer advantageous techniques for vector quantization systems. This is achieved using photorefractive crystals, which provide high density recording in real time, as our holographic media. The reconstruction alphabet can be dynamically constructed through training, or stored in the photorefractive crystal, prior to the encoding process. The encoding of a new vector is carried out in constant time by holographic associative matching. An extension to interframe coding is also being investigated. More information on this research effort can be found in [RY92].

6.0 Software Developed Under This Contract

We have developed software for performing basic image processing functions that were necessary for our study. This software includes image displaying, image partitioning, conversion of images in different formats, and image comparison.

We have also developed prototype software (written in C, running on a SUN-4) for our newly developed algorithms as well as a number of existing data compression algorithms. A list of some of the most important programs is given below. This software can be available to anyone upon request.

Transform Data Compression Methods:

- Two-dimensional discrete cosine transform
- Bilinear Transformation
- Pyramid coding
- Histogram equalization

Data compression algorithms:

- Full-searched vector quantization (FSVQ)
- Tree-structured vector quantization (TSVQ)
- Pruned tree-structured vector quantization (PTSVQ)
- Variable-length tree-structured vector quantization (VLVQ)
- Multiresolution vector quantization (MRVQ)
- Error-coding Vector Quantization (ECVQ)
- Error-coding Vector Quantization with multiresolution capabilities(ECMR)
- Clear-bit method for lossless compression
- Fast computations of the TSVQ algorithm:
  - Difference TSVQ algorithm
  - Vector reduction
  - Mean error measure

Decompression programs:

Distortion evaluation programs for various measures such as:

- the mean squared error,
- the mean error, the weighted squared error,
- and the reduced size mean squared error.

(These programs also include distortion display capabilities such as the error images, and error histograms.)

In particular, the research on wavelet compression has resulted in a releasable version of the wavelet compression package, and our quad-tree vector quantization program is being integrated into Dr.Immanuel Freedman's compression package for COBE data.

Publications

[MR91a] Markas, T., and J. Reif, "Image Compression Methods with Distortion Controlled Capabilities", *Proc. of IEEE Data Compression Conference* (DCC 91), Snowbird, Utah, April 1991, pp.93- 102.

[MR91b] Markas, T., and J. Reif, *Fast Computations of Vector Quantization Algorithms*, NASA Technical Report TR-91-58, 1991.

[MR92] Markas, T., and J. Reif, "Quad-Tree Structures for Image Compression Applications", *Information Processing & Management*, Vol.28, No.S, 1992.

[MR93] Markas, T. and J. Reif. "Multispectral Image Compression Algorithms". *Proc. of the Data Compression Conference* (DCC '93), Snowbird, UT, March, 1993, pp. 391-400.

[MREE93] Markas, T., J. Reif, W. Elliot, and E. Elliot. "Memory-shared Parallel Architectures for Vector Quantization Algorithms". *Proc. Picture Coding Symposium*, Lusanne, Switzerland, March, 1993 .

[MRS93] Markas, T., J. Reif, and J. Storer, "On Parallel Implementations and Experimentations of Lossless Data Compression Algorithms", *Proc. Picture Coding Symposium*, Lusanne, Switzerland, March, 1993.

[Mar93] Markas, T.. *Data Compression: Algorithms and Architectures.* Ph.D. Dissertation, Duke University, Dept. of Electrical Engineering, 1993.

[RY92] Reif, J., and A. Yoshida, "Optical Techniques for Image Compression", *Proc. 2nd Annual IEEE Data Compression Conference* (DCC 92), Snowbird, UT, March 1992, pp. 32-41. Also in *Image and Text Compression*, edited by James A. Storer, Kluwer Academic Publishers, 1992.

[SRM90] Storer, J., J. Reif, and T. Markas, "A Massively Parallel VLSI Design for Data Compression using a Compact Dynamic Dictionary", *Proc. of IEEE Workshop on VLSI & Signal Processing*, 1990, San Diego, CA, pp.329-338. Also published as "A Massively Parallel VLSI Compression System Using a Compact Dictionary", *VLSI Signal Processing*, No. 4, 1990 (edited by H.S. Moscovitz and K. Yao and R. Jain), IEEE Press, 1990, New York, NY, pp. 329-338.

[Tat92] Tate, S. *Lossless Compression of Region Edge Maps.* Duke University Computer Science Technical Report CS-1992-09.

[Tat93a] Tate, S. *Lossless Compression of Multispectral Images.* In preparation.

[Tat93b] Tate, S. *Report on the Workshop on Data and Image Compression Needs and Uses in the Scientific Community*, CESDIS Technical Report TR-93-99.

RESEARCH ACTIVITIES

Peer Reviewed Projects

Applied Information Systems Research (AISRP)

In the spring of 1990, CESDIS supported the peer review process of NASA NRA-89-OSSA-21, *Applied Information Systems Research.* Selected projects were funded and administered directly by NASA through Goddard's procurement and contracting offices. Early in 1992, CESDIS was asked to administer the projects where Principal Investigators were not Federal government employees. This support has involved the collection of periodic progress reports, obtaining approval for equipment purchases, obtaining permission for foreign travel, invoice processing, and information dissemination.

CESDIS personnel have attended the annual AISRP Workshop organized by Glenn Mucklow (NASA HQ, ST) and hosted by Randal Davis of the University of Colorado's Laboratory for Atmospheric and Space Physics (LASP) in Boulder. CESDIS will help make arrangements for and provide on-site support for the 1994 workshop in July 1994.

The projects listed in this section were funded for three years and are approaching the end of the third year. Final reports are being collected and will be compiled into one summary document which may be requested by contacting the CESDIS administrative office at cas@cesdis1.gsfc.nasa.gov. Software developed will be deposited with the Software Support Laboratory (SSL) at LASP which is acting as a repository for software developed through this NASA program and others. For more information about the Software Support Laboratory and/or access to the software described in this section, send an e-mail request to ssl@sslab.colorado.edu. Addresses for project personnel are included in Appendix E of this report.

Boston University

Development of a Tool-set for Simultaneous, Multi-site Observations of Astronomical Objects

Supriya Chakrabarti
Center for Space Physics

Task Objective

A network of ground and space-based telescopes can provide continuous observation of astronomical objects. In a target of opportunity scenario triggered by the system, any telescope on the network can request supporting observations. The investigator intends to develop a set of data collection and display tools to support these observations. He plans to demonstrate the usefulness of this toolset for simultaneous multi-site observations of astronomical targets.

Brandeis University

High Performance Compression of Science Data

James Storer
Computer Science Department

Task Objective

The investigators plan to develop algorithms that can be a basis for software and hardware systems that compress a wide variety of scientific data with different criteria for fidelity/bandwidth tradeoffs. The algorithmic approaches will be targeted for parallel computation where data rates of over 1 billion bits per second are achievable with current technology.

California Institute of Technology

Multivariate Statistical Analysis Software Technologies for Astrophysical Research Involving Large Data Bases

S. G. Djorgovski

Task Objective

We have developed a system, called SKICAT, for producing, managing, and analyzing catalogs from the digitized POSS-II survey. The system classifies and matches catalogs from multiple, overlapping plate scans as well as CCD calibration sequences. In this proposal, we describe how we would also like to integrate and extend the analysis tools provided by SKICAT, to facilitate more sophisticated scientific investigations of these expanding survey data sets. The tools we intend to provide would include our already developed STATPROG multivariate statistical analysis package, and a wide variety of new Bayesian inference tools, objective classifiers and other advanced data a management and analysis packages and algorithms. We are also separately applying to the NASA ADP program to extend SKICAT to accommodate catalogs from other sources, such as from the space-based IRAS and ROSAT missions. The finished system should thus be of a considerable utility to a much wider NASA community, going beyond the immediate task of processing 3 Terabytes of digitized POSS-II information.

Hughes Applied Information Systems

Advanced Data Visualization and Sensor Fusion: Conversion of Techniques from Medical Imaging to Earth Science

Vance Mc Collough
Colorado Engineering Laboratories

Task Objective

The investigators plan to transfer existing medical imaging registration algorithms to the area of multi-sensor data fusion. The University of Chicago's algorithms have been successfully demonstrated to provide pixel-by-pixel comparison capability for medical sensors with different characteristics. The research will attempt to fuse GOES, AVHRR, and SSM/I sensor data which will benefit a wide range of researchers. The algorithms will utilize data visualization and algorithm development tools created by Hughes in its EOSDIS prototyping. This will maximize the work on the fusion algorithms since support software (e.g., input/output routines) will already exist.

Institute of Global Environment and Society

The Grid Analysis and Display Systems (GRADS): A Practical Tool for Earth Science Visualization

James Kinter

Task Objective

The investigator proposes developing and enhancing a workstation-based grid analysis and display software system for Earth science dataset browsing, sampling and manipulation. The system will be coupled to a supercomputer in a distributed computing environment for near-real time interaction between scientists and computational results.

Massachusetts Institute of Technology

Topography from Shading and Stereo

Berthold Horn
Artificial Intelligence Laboratory

Task Objective

Methods exploiting photometric information in images that have been developed in machine vision can be applied to planetary imagery. Present techniques, however, focus on one visual cue, such as shading or binocular stereo, and produce results that are either not very accurate in an absolute sense or provide information only at a few points on the surface. The investigators plan to integrate shape from shading, binocular stereo and photometric stereo to yield a robust system for recovering detailed surface shape and surface reflectance information.

National Center for Atmospheric Research (NCAR)

Interactive Interface for NCAR Graphics

Bill Buzbee
Scientific Computing Division

Robert Lackman
Scientific Visualization Group

Task Objective

NCAR Graphics is a FORTRAN 77 library of over 30 high-level graphics modules which are heavily used by science and engineering researchers at over 1500 sites worldwide. These Earth science-oriented modules now have a Fortran callable subroutine interface which excludes their use by non-programming researchers. The investigators plan the development of a fully interactive "point and click" menu-based interface using the prevailing toolkit standard for the X Window system. Options for direct output to the display window and/or output to a Computer Graphics Metafile (CGM) will be provided. X, PEX, and PHIGS will be implemented as the underlying windowing and graphics standards. Associated meteorological and geometric data sets would exploit the network extended NASA Common Data Format, netCDF.

Science Applications International Corporation (SAIC)

SAVS: A Space Analysis and Visualization System

Edward Szuszczewicz
Laboratory for Atmospheric and Space Science

Task Objective

The investigators propose to develop a powerful and versatile data acquisition, manipulation, analysis, and visualization system which will enhance scientific capabilities in the display and interpretation of diverse and distributed data within an integrated user-friendly environment. The approach will exploit existing technologies and combine three major elements into an easy-to-use interactive package: (1) innovative visualization software, (2) advanced database techniques, and (3) a rich set of mathematical and image processing tools. Visualization capabilities will include 1-, 2-, and 3- dimensional displays, along with animation, compression, warping, and slicing. Analysis tools will include generic mathematical and statistical techniques along with the ability to use large-scale models for interactive interpretation of large-volume data sets. The system will be implemented on Sun and DEC UNIX workstations and on the Stardent Graphic Supercomputer.

Space Telescope Science Institute

Data Reduction Expert Assistant

Glenn Miller, Mark Johnson, Robert Hanisch

Task Objective

The investigators propose to develop an expert system tool for the management and reduction of complex data sets. The reduction of such data presents severe challenges to a scientist: not only must a particular data analysis system be mastered, large amounts of data can require days of tedious work and supervision for even the most straightforward reductions. The proposed Data Reduction Expert Assistant will help the scientist overcome these obstacles by developing a reduction plan based on the data and producing a script for the reduction of the data in the language of the analysis system. The script will then be executed to perform the reduction. A powerful user interface and a customizable knowledge base will enhance the usefulness of this tool.

Texas A&M University

An Interactive Environment for the Analysis of Large Earth Observation and Model Data Sets

Kenneth Bowman
Climate System Research Program

Robert Wilhelmson
University of Illinois, Urbana
Department of Atmospheric Sciences
National Center for Supercomputing Applications

Task Objective

The investigators propose to develop an interactive environment for the analysis of large Earth science observation and model data sets. The investigators will use a standard scientific data storage format and a large capacity (>20 GB) optical disk system for data management; develop libraries for coordinate transformation and regridding of data sets; modify the NCSA X Image and X DataSlice software for typical Earth observation data sets by including map transformations and missing data handling; develop analysis tools for common mathematical and statistical operations; integrate the components described above into a system for the analysis and comparison of observations and model results; and distribute software and documentation to the scientific community.

University of Colorado, Boulder

A Land-Surface Testbed for EOSDIS

William Emery
Colorado Center for Astrodynamics Research

Task Objective

The investigators propose to develop an on-line data distribution and interactive display system for the collection, archival, distribution and analysis of operational weather satellite data for applications in land surface studies. A 1000 km square scene of the western U. S. (centered on the Colorado Rockies) will be extracted from Advanced Very High Resolution Radiometer imagery (AVHRR). These AVHRR data will be navigated and map registered at CU/CCAR and then be transferred to NCAR for storage in an on-line data system. A display workstation software will be developed and fully distributed on-line that will display and further process the AVHRR data for studies of vegetation monitoring and snowpack assessment. This experiment with an active on-line and indicative analysis system will provide experience with a small scale EOSDIS.

University of Colorado, Boulder

Experimenter's Laboratory for Visualized Interactive Science

Elaine Hansen
Colorado Space Grant Consortium

Task Objective

The investigator proposes adapting and upgrading several existing tools and systems to create an experimenter's laboratory for visualized interactive science. Intuitive human-computer interactive techniques have already been developed and demonstrated at the University of Colorado. A Transportable Applications Executive (TAE+), developed at Goddard Space Flight Center, is a powerful user interface tool for general purpose applications. A 3D visualization package developed at NCAR provides both color-shaded surface displays and volumetric rendering in either index or true color. The Network Common Data Form (NetCDF) data access library, developed by Unidata supports creation, access and sharing of scientific data in a form that is self-describing and network-transparent.

University of Illinois

A Distributed Analysis and Visualization System for Model and Observational Data

Robert Wilhelmson
Department of Atmospheric Sciences
National Center for Supercomputing Applications

Task Objective

The objective of this proposal is to develop an integrated and distributed analysis and display software system which can be applied to all areas of the Earth system science to study numerical model and Earth observational data from storm to global scale. This system will be designed to be easy to use, portable, flexible and easily extensible and to adhere to current and emerging standards whenever possible. It will provide an environment for visualizing the massive amounts of data generated from satellites and other observational field measurements and from model simulations during or after their execution. Two and three dimensional animation will also be provided. This system will be based on a widely used software package from NASA called GEMPACK and prototype software for three dimensional interactive display built at NCSA. The underlying foundation of the system will be a set of software libraries which can be distributed across UNIX-based supercomputers and workstations.

University of Maryland, College Park

VIEWCACHE: An Incremental Pointer-based Access Method for Autonomous Interoperable Databases

Nicholas Roussopoulos
Department of Computer Science

Task Objective

VIEWCACHE is intended to provide an interface for accessing distributed datasets and directories. It allows database browsing and search performing inter-database cross-referencing with no actual data movement between database sites. This organization and processing is especially suitable for managing astrophysics databases which are physically distributed all over the world. Once the search is complete, the set of collected pointers pointing to the desired data are cached. VIEWCACHE includes spatial access methods for accessing image datasets, which provide much easier query formulation by referring directly to the image and very efficient search for objects contained with a two-dimensional window. A VIEWCACHE External Gateway Access to database management systems will be developed and optimized to facilitate distributed database searches.

University of Wisconsin

Planetary Data Analysis and Display System: a Version of PC-McIDAS

Sanjay Limaye, L. A. Sromovsky

Task Objective

The investigators propose developing a system for access and analysis of planetary data from past and future space missions based on an existing system, the PC-McIDAS workstation. This system is now in use in the atmospheric science community for access to meteorological satellite and conventional weather data. The proposed system would be usable by not only planetary atmospheric researchers, but also by the planetary geologic community.

RESEARCH ACTIVITIES

Additional Tasks

The projects included in this category are short term tasks generally of only a few months duration. They are initiated by individuals within NASA Headquarters or Goddard's Space Data and Computing Division who provide funding from budgets allocated to them. The investigators contribute reports to the CESDIS technical report series, take part in CESDIS-sponsored workshops and seminars, and are given access to NASA computers when on-site.

(Task 24)

UNIVERSITY OF MARYLAND, BALTIMORE COUNTY

Research on Digital Libraries and Computer System Performance

Timothy Finin, Yelena Yesha
Department of Computer Science

Task Objective

Objective 1: Digital Libraries

Perform research on the problems and issues related to requirements for distributed database systems to support digital libraries. The goal is to develop a prototype information system suitable for use on a large network. The four major objectives are to design algorithms and architecture for the information system, to study theoretical issues relating to the effectiveness and scalability of the developed algorithms, to make a prototype implementation of the information system using a small number of workstations, and to adapt the prototype to allow growth. The result of this research will be an information system of general applicability which is capable of handling our current and future information management needs.

Objective 2: NASA's Computer System Performance

Provide a report on the analytical studies of NASA's Center for Computational Sciences plus a report on the contractor's experience with NASA's computing facilities as a new user.

Final reports are not yet available, but interim reports follow.

Implementation Status of Alibi

David Flater and Yelena Yesha

1.0 Introduction

The work we have been doing for CESDIS is called Alibi, for Adaptive Location of Internetworked Bases of Information. Alibi is a network of information servers whose purpose is to locate and retrieve information from databases and archives scattered across the Internet. It eliminates the need for users to be intimately familiar with the Internet sites providing the information, and it eliminates duplication of effort among users by caching frequently used data at nearby sites. Furthermore, it avoids the complexity and confusion of a navigation interface by accepting simple keyword queries.

Alibi currently consists of two programs. The smaller program, which has been named Alibi for the convenience of users, is a user interface which accepts keyword queries, communicates with an information server, and assists users in dealing with the responses. The larger program, named Unetd, is the server which does all the actual work for answering queries and communicating with other servers. The remainder of this report

RESEARCH ACTIVITIES

Additional Tasks

The projects included in this category are short term tasks generally of only a few months duration. They are initiated by individuals within NASA Headquarters or Goddard's Space Data and Computing Division who provide funding from budgets allocated to them. The investigators contribute reports to the CESDIS technical report series, take part in CESDIS-sponsored workshops and seminars, and are given access to NASA computers when on-site.

(Task 24)

UNIVERSITY OF MARYLAND, BALTIMORE COUNTY

Research on Digital Libraries and Computer System Performance

Timothy Finin, Yelena Yesha
Department of Computer Science

Task Objective

Objective 1: Digital Libraries

Perform research on the problems and issues related to requirements for distributed database systems to support digital libraries. The goal is to develop a prototype information system suitable for use on a large network. The four major objectives are to design algorithms and architecture for the information system, to study theoretical issues relating to the effectiveness and scalability of the developed algorithms, to make a prototype implementation of the information system using a small number of workstations, and to adapt the prototype to allow growth. The result of this research will be an information system of general applicability which is capable of handling our current and future information management needs.

Objective 2: NASA's Computer System Performance

Provide a report on the analytical studies of NASA's Center for Computational Sciences plus a report on the contractor's experience with NASA's computing facilities as a new user.

Final reports are not yet available, but interim reports follow.

Implementation Status of Alibi

David Flater and Yelena Yesha

1.0 Introduction

The work we have been doing for CESDIS is called Alibi, for Adaptive Location of Internetworked Bases of Information. Alibi is a network of information servers whose purpose is to locate and retrieve information from databases and archives scattered across the Internet. It eliminates the need for users to be intimately familiar with the Internet sites providing the information, and it eliminates duplication of effort among users by caching frequently used data at nearby sites. Furthermore, it avoids the complexity and confusion of a navigation interface by accepting simple keyword queries.

Alibi currently consists of two programs. The smaller program, which has been named Alibi for the convenience of users, is a user interface which accepts keyword queries, communicates with an information server, and assists users in dealing with the responses. The larger program, named Unetd, is the server which does all the actual work for answering queries and communicating with other servers. The remainder of this report

will discuss the algorithms and techniques which have been employed in the implementation of Unetd to make Alibi a success.

2.0 Unetd Architecture

The information network of Alibi (sometimes called the Übernet) consists of a large number of information servers (Unetd's) that maintain links with their closest neighbors over the Internet. At the center of each Unetd is an asynchronous network driver that services all the active connections in-a round-robin fashion. Whenever new data arrive or a connection becomes ready to receive queued data, the server makes a sweep through all the active connections, insuring that each one is serviced in a timely fashion. The rest of the time the server remains in a dormant state, only waking up occasionally to update its network performance statistics and reconnect any links which have been broken. Even the establishment of connections is done asynchronously; other connections continue to be serviced while confirmation of a network connection is awaited.

The topology of the information network is roughly determined by the list of servers with which each individual Unetd is told to link up. This list is selected by the local Unetd administrator; each installation is free to link up with whatever servers provide the best localized performance, and there are no mandatory guidelines for how many or what kind of servers to contact. The actual topology of the information network varies over time anyway since poor Internet performance and machine downtime cause links to be broken. The servers routinely re-establish links that have been broken and route messages via whatever links are up and running at the time.

Point-to-point routing is supported through the use of an improved version of the old NETCHANGE protocol [1]. We chose to start with NETCHANGE because of its simplicity; however, the stock protocol did not meet our requirements because it figured distance only in terms of the number of hops and did not really support the removal of hosts from the distance and routing tables should they become inaccessible. We succeeded in modifying the protocol to measure distance as the number of microseconds needed to send a message, to cause distance and routing tables to be updated as the observed delays change, and to remove inaccessible hosts from distance and routing tables. The tables themselves are hashed by destination to reduce overhead.

3.0 Handling of Queries

When a user wishes to enter a query, he or she runs the Alibi program to contact an information server. A query is then entered as a series of key words. When the server receives the query, it removes all punctuation, deletes junk words, and uses a thesaurus to replace words with more popular equivalents whenever possible. For example, the query "picture of Elvis" would end up as "image Elvis." The parsing of queries is still being improved; boolean keywords (and, or, not) are not currently recognized as special.

The query is then passed to a classifier which looks for buzzwords and attempts to classify the query. In the above example, "image" is a buzzword specifying the type of data being requested. The classifier scans the list of known classification and specializations, looking for the most general classification which matches the maximum number of buzzwords in the query. "Image" is a specialization of the general data class "blob" (acronym for binary large object, author unknown), so "image Elvis" would be classified as "blob image." In the event that a site somewhere specialized in pictures of Elvis, there might exist a data class "blob image elvis" which would be preferred since it matches more buzzwords in the query. However, a specialization such as "blob image nasa" would not be preferred to "blob image" since it matches the same number of buzzwords but is less general.

Before the query is forwarded, the information server first lets its own resources have a look at the query to see if it can be answered locally. Most resources will simply look at the classification and return the query unanswered. However, a resource which knows something about the data class being sought will search its information base for relevant data and answer the query if relevant data are found. The cache resource, which is general-purpose, always searches for relevant data before passing on the query.

The forwarding of queries is actually implemented as another resource. The "forwarding resource" is last in line to receive the query, and it always succeeds in handling the query so long as there are network connections on which to forward it. In the ugly event that there are not, an apologetic message is returned to the user who entered the query explaining that there is nowhere to send it.

Queries are forwarded according to the advice of a "point-to-data" router (as opposed to a point-to-point router, which gives the next host on the way to a specific host). The routing table for point-to-data routing is indexed by the class of data being sought, rather than the destination host. Furthermore, it does not merely provide a single next host in line to which to forward the query; statistics are kept on each of the neighboring hosts to help determine the best route for a query that may already have been seen by some of the neighbors. If responses of a certain class arrive most frequently from neighboring site A, queries on that class of data will usually be forwarded to site A. If site A has already seen the query, it will be sent to the next most likely neighbor which has not already seen the query.

The routing information for queries is updated based on the traffic of response messages through the local site. When a response message is processed, all possible generalizations of the classification applied to the response by its creator are updated in the query routing table. Thus, a site does not need to recognize the specialization of a data class to be able to route a query. A query classified as "blob image elvis" can be routed using "blob image elvis," "blob image," "blob," or the universal class.

A detailed explanation of the query routing algorithm and the keeping of statistics can be found in [2].

4.0 Resources

We have just discussed the workings of the query forwarding resource, which is the last resource to be given a chance to handle the query. In fact, there is nothing preventing other resources from themselves forwarding a query, rather than answering it. The main program simply needs the query to be "handled" one way or the other. A specialized resource, which recognizes the specialization of a query but cannot itself provide an answer to the query, might nevertheless know in which direction the answer lies and send the query off in that direction.

This capability is exploited by the netnews resource to perform searches over Usenet news in parallel with other operations. Whenever the netnews resource recognizes the class of a query, it tells the main program that the query has been handled and then queues up a series of commands for the NNTP server. When the responses arrive from the NNTP server, they are passed to the netnews resource. At this point either the query is answered, or the netnews resource hands the query to the query forwarder so that it can continue on its way towards an answer. This technique even allows the NNTP commands for several different queries to be interleaved and multiplexed over a single NNTP connection without them interfering with one another.

The netnews resource turns portions of the newsgroup hierarchy into specialized textual information bases. It was the first mediator written for Alibi; its purpose was mainly to help with research and development, but it does provide the service of finding useful information in netnews. Each netnews resource handles a different portion of the news hierarchy as specified in a configuration file. It is possible to handle non-hierarchical slices of the newsgroup space by specifying a substring that must appear in the newsgroup name; the substring ".politics." would include newsgroups from both the talk.politics and alt.politics subtrees. An example netnews resource would be a site that handles all newsgroups in the comp.infosystems subtree (currently comp.infosystems, comp.infosystems.gis, comp.infossystems.wais.comp.infosystems.gopher, and comp.infosystems.www). The resource would attempt to answer queries looking for information about infosystems in general or any of the specific infosystems listed above by retrieving articles from the NNTP server whose subject lines contain keywords from the query. Responses are assigned classifications starting with "netnews" and then qualified in accordance with the netnews hierarchy. "netnews comp infosystems gopher" is an example. We expect the netnews data class to go away or be made a specialization of a more general class such as "text" at some time in the future.

The data cache is also a resource. Each information server maintains a data cache whose size is determined by the local administrator based on the desired performance and the available computer resources. When a query is presented, the cache is searched for data of an applicable class whose descriptions contain keywords from the query. If a good cache hit is found, the query is answered. The data cache is filled with the responses to earlier queries in accordance with a decision function that combines a number of statistics to insure an efficient use of the cache space that is fragmented across many sites. Neighboring sites effectively cooperate to make the best possible use of their collective cache space. A more detailed discussion of the caching strategy will be provided in the next section.

The only other resource that is already implemented is the blob resource. This resource searches a directory tree for index files that provide descriptions for other files residing in the same directory tree. These files are then returned as binary data in response to matching queries. The pathname relative to the root blob directory, like the dot-delimited keywords in the netnews hierarchy, is used to qualify the general class of data that the blob resource handles when assigning classes to responses. If a blob resource handling "blob image nasa" has subdirectories called ozone and xray, files in those directories would be classified as "blob image nasa ozone" and "blob image nasa xray." This technique insures that the metadata inherent in the directory structure itself is preserved and propagated.

5.0 Caching Strategy

Our novel approach to distributed cache management takes direct control of the level of cache turnover at individual sites, thus insuring that no site is burdened with runaway cache turnover [3]. This technique works well in combination with our query router because routing queries towards cached replicas is only helpful if those replicas are still likely to exist by the time the query arrives. There are two parts to the cache mechanism, the decision function and the replacement algorithm. The replacement algorithm is an enhanced version of LRU (Least Recently Used). The enhancement takes the form of hint values [4] which are added to the last access times of cached replicas to tweak the LRU mechanism into giving preference to more valuable data.

The decision function trivially determines which data are valuable in addition to determining which data will be cached at the local site. The collective behavior of a group of sites using this decision function is such that duplication of effort among that group of sites is nearly eliminated once caches become full (if there is room to spare, there is no reason not to have full replication).

The decision function has two stages. The first stage works for "valuable" data, which are those that are not likely to be replicated nearby. The second stage merely fills up unused cache space with any data that will fit, assigning a hint value that designates them as "not valuable." If a datum marked as "not valuable" is later used to answer another query, its hint value is upgraded.

The following terms are needed to define the decision function:

- rnd. This term is replaced by uniformly distributed random numbers in the interval [0,1). The random factor helps to create a uniform distribution of replicas.

- Hops so far. This is a counter of the number of times the response message has been forwarded since the last replica of the datum in the response message was created or collided with on the return path. This counter is bundled into the response message and is updated by each site on the return path.

- Running link cost. This counter is managed exactly like the hops counter except that the sum of the costs of the links (i.e., total transmission delay) traversed by the response message is kept instead of the number of hops.

- Running cache sum. This counter is similar to the other two but contains the sum of the amount of cache space, both used and unused, at each site visited since the last replica. Used cache space is

included since sites with larger caches can handle a larger number of cache replacements while maintaining the same level of turnover. In other words, we want sites to cache data in accordance with their abilities to do so.

- Sum of neighboring link costs. This is the sum of the costs of the links connecting the current site with each of its neighbors.

- TCF. This is the Turnover Control Factor. Each site maintains its own TCF: it is not transmitted from site to site. The TCF is increased when a site wants to lower its turnover and decreased (subject to > 0) when a site wants to increase its turnover. Turnover is quantified as a smoothed function of the difference between the current time and the hint-adjusted timestamps of the items being replaced in cache.

The stage one decision function is:

$$rnd < \frac{Hops\,sofar}{Hops\,sofar + TCF} \times \frac{Running\,link\,cost}{TCF \sum neighboring\,link\,costs} \times \frac{Running\,cache\,sum}{TCF}$$

This function is derived simply as the product of three factors which determine a desirable placement of replicas. The three factors on the right hand side of the function are, from left to right:

1. HOPS. This factor discourages the caching of many copies extremely close to one another. As distance increases, this factor vanishes towards 1.

2. LINK COSTS. This factor controls the cost of repeated queries on the same datum by caching more copies as this cost mounts. The use of the costs of the links to the current site's neighbors greatly reduces the undesirable effects resulting from large variations in link costs.

3. CACHE SPACE. This factor prevents under-utilization of cache space when the distribution of such space is uneven.

If a datum fails the stage one test, it can still be cached with a "cold" hint if there is sufficient free cache space to hold the datum. The purpose of this second stage is to fill up idle cache space at each site (and hence in the entire system). Idle cache space is produced by a new site coming on-line with an empty cache and by fragmentation. If it is implemented, expiration of data with a limited lifespan will also contribute. The idle space is quickly filled up with small data objects having cold hints.

Items which go unused for a long period of time are simply removed by LRU; it is not necessary to downgrade warm hint values or forcibly eject items from the cache. Although our architecture does not support reliably locating every replica of a datum, if a new version of a datum is released, the new version will propagate outward from its source, causing old versions to be removed whenever they are encountered. Newer versions automatically replace older versions whenever they collide. The expected number of obsolete replicas in the system approaches zero unless there exists some site which never receives a response message. A version of the proof from [5] follows.

If the expected number of obsolete replicas does not reach zero, then there must exist at least one "immortal" replica. We cannot expect some group of sites to perserve the obsolete datum indefinitely by periodically creating a new replica and destroying the old; the continuation of such behavior becomes increasingly unlikely as time passes. Let us then consider the conditions which are necessary for a replica to become immortal. Since rnd takes on random values which are uniformly distributed in the interval [0,1), the only way to prevent the decision function from eventually firing is to force the right hand side to zero every time a response is received. This can only occur if one of the following holds:

1. Hops so far is zero. This cannot happen since no messages are sent when a query is answered from locally cached data.

2. Running link cost is zero. There is no such thing as negative link delay, so this too is impossible.

3. Running cache sum is zero. This implies that we have no cache space, which contradicts our possessing an immortal replica.

Therefore cache replacement can only be prevented by not receiving any response messages. Any time that cache replacement occurs, it is possible for the obsolete replica to be victimized. One large datum can flush the entire cache to make room. Therefore an immortal replica can only exist if a site receives no response messages.

A site receiving no response messages must never generate queries which it cannot itself answer. For obsolete data to be used, these queries must refer to data for which the site owns immortal, obsolete replicas. If even one query is entered which cannot be answered from local data, the obsolete data are in jeopardy of being replaced. In practice, therefore, the conditions for indefinitely preserving obsolete data are not likely to be met.

6.0 Ongoing Work

We currently are looking at the possibility of using NFS (Network File System) to leverage the blob resource into operating on large public archives such as wuarchive.wustl.edu and oak.oakland.edu. We originally intended to write a separate resource which would do anonymous FTP, but this will probably no longer be necessary.

We are also in the process of making the existing code base faster, more reliable, and more readable. One of the ideas being considered to insure a consistent level of performance is to establish separate Internet connections for the transfer of individual data that are so big that they interfere with the flow of other data through the virtual network. This would handle the problem of flow control by pushing it down into the network layer, where it is supposed to have been solved already. "Normal" sized data (under one megabyte) do not usually cause flow problems: of course, when there are serious problems with the underlying network, any transmission at all can take arbitrarily long, and there is nothing that an application program can do to improve the situation.

Soon we will be enhancing the query parser to permit limited boolean operators. Of particular interest is the "not" operator, which is useful when the first attempt to answer a query did not return exactly the desired response. The unsatisfied user could re-enter the original query, but with an additional clause asking that the datum sent the first time *not* be sent again. Because of the caching system, a repeated query with no additional qualifiers is almost guaranteed to result in a repeated answer, even if other answers exist somewhere in the system. We intend to enhance our user interface to automatically qualify and resubmit the previous query in response to a simple command.

We will also be creating mediators for more real-world information bases as they are made available to us. Only by providing access to a large number of actual preexisting public databases can the full capabilities of Alibi be demonstrated. After bringing in FTP sites through the use of NFS, we intend to make a mediator for databases conforming to the standard for Transitional SQL[6]. Transitional SQL is rumored to be next year's target of choice for database vendors seeking conformance to FIPS SQL, and it provides a standard way of accessing metadata which will be essential if a portable mediator is to be built. A mediator for libraries of reusable source code is also planned. Although source code could be archived with the blob resource, it should be possible to have a specialized resource that makes use of software specifications to understand more about the source code being archived and thus a better chance of correctly responding to requests. It would also be nice to have a mediator for a full text retrieval system that makes use of advanced text retrieval

techniques. We are searching for software reuse and text retrieval systems that already exist for public use so that we may make them available through Alibi.

An option we will consider in the uncertain future is to implement an interface which would allow users who can only access the Internet through e-mail to send queries to Alibi and receive responses using only e-mail. Users who are trapped behind Internet firewalls, who access the Internet via UUCP, or who use dialup services that only support e-mail will then be able to use Alibi. Binary data will have to be unencoded or converted to text some other way and broken into chunks small enough to survive transmission across these external networks, and the users will need to decode the binaries themselves. However, a surprising number of people have only e-mail access to the Internet, and this kind of service is better than nothing for them.

Lastly, there will always be room for enhancements to the Alibi program, the user interface. Image files should be displayed, sound files should be played, and all different kinds of compression and encoding should be recognized without relying on the user to manually call up the auxiliary software needed to perform these actions. The only limit on how much could be built into the client program is how much effort we wish to put into it; it is worth significant effort to make the system friendly to all users.

References

[1] William D. Tajibnapis. A correctness proof of a topology information maintenance protocol for a distributed computer network. *Communications of the ACM,* 20(7): 477-485, July 1977.

[2] David W. Flater and Yelena Yesha. An efficient management of read-only data in distributed information system. *International Journal of Intelligent and Cooperative Information Systems, Special Issue on Information and Knowledge Management,* 1994. To appear.

[3] David W. Flater and Yelena Yesha. Managing read-only data on arbitrary networks with fully distributed caching. *International Journal of Intelligent and Cooperative Information Systems,* 1994. To appear.

[4] Henry M. Gladney. A model for distributed information networks. Technical Report RJ5220, IBM Almaden Res. Lab, 650 Harry Road, San Jose, California 95120-6099, July 1986.

[5] David W. Flater and Yelena Yesha. Properties of Networked Information Retrieval with Alibi. In *Proceedings of the Second International Conference on Information and Knowledge Management,* pages 31-38, Washington, DC, U.S.A., November 1993. ACM Press.

[6] FIPS PUB 127-2. *Federal Information Processing Standards Publication for Database Language SQL.* X3H2-93-106, January 1993.

Performance Analysis of the Unitree Central File Manager at NASA's Center for Computational Sciences Part I: System Overview and Data Analysis

Odysseas Pentakalos

1.0 Introduction

The purpose of this report is to give a description of the system at NASA's Center for Computational Sciences (NCCS) in order to serve as a basis for the performance analysis to follow. The first section will cover an overview of the hardware and software configuration of the system at the center. The second section describes the Mass Storage System Reference Model developed at the Lawrence Livermore National Lab.

The Unitree File system has been designed after the Mass Storage System Model so in order to understand the internal structure of the Unitree you must be familiar with the Model. The third section will describe the ftp log files which have been examined in the past by the support personnel at Convex Corporation in order to extract user access patterns. The last section will describe the other log files which are generated by the Unitree Central File Manager in order to detect any potential of extracting useful information from them. This report is just an informal description of the system and is to be used by the author and Dr. Yelena Yesha as a log of the research progress.

2.0 Hardware/Software Configuration

This section of the report will describe the hardware/software configuration of the section of the system at NCCS which is under consideration in this performance analysis effort [4]. The system is made up of a Convex C3240 which is a ConvexOS (Unix 4.3 BSD variant) based, four processor system. It has 512 MB of memory, 150 GB of disk space and it is connected to three StorageTek 4400 Silos with a total nearline robotic storage of 3.6 TB (terabytes). This massive combination of storage media is managed by the Unitree Central File Manager version 1.7. The Unitree is a hierarchical mass storage system management facility which provides transparent access to massive amounts of information through the FTP and NFS network utilities. The system also includes a Cray C98 with 6 processors, 2048 MB of memory and 4 GB of disk space running the Unicos operating system. The Convex, the Cray and about 750 workstations in NASA are interconnected with an Ethernet network running at 10 Mbits/sec and a high speed Ultranet network providing gigabit/sec access to the Cray C98. Figure 1 gives a graphical representation of the system being considered for performance analysis.

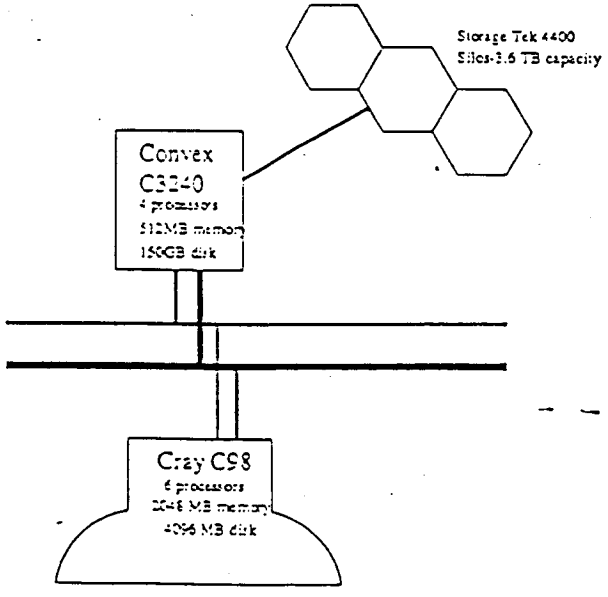


Figure 1. NCCS System Overview

3.0 Mass Storage System Reference Model

3.1 Introduction

This section will describe the Mass Storage System Reference Model (MSSRM) developed at the Lawrence-Livermore Laboratory [1]. The purpose of the MSSRM was to develop a standard for the standardization of Mass Storage Systems. It is envisioned that the modules of the model can be integrated in various

combinations to support a wide variety of storage needs and platforms expanding the flexibility of Mass Storage Systems but at the same time reducing their price. The standard describes the various modules needed in a Mass Storage System and how they interface with each other, but it does not describe their internal structure in order to leave the actual implementation up to the individual designer. The model follows an object oriented paradigm in describing the various modules and also adheres to the client/server model in order to make distributed system implementations possible.

3.2 Top Level Modules

The overall system is made up of seven primary modules as shown in Figure 2.

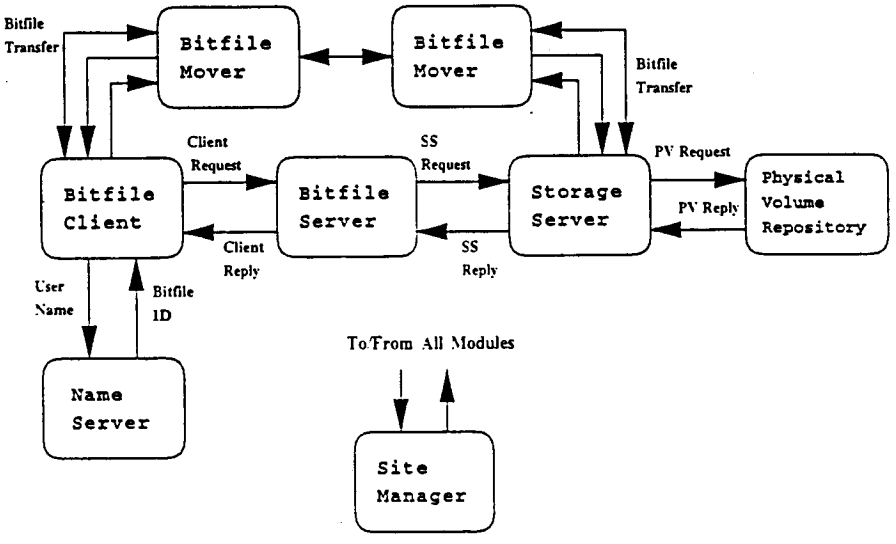


Figure 2. Mass Storage System Reference Model

A bitfile is a term coined by the IEEE CS Technical Committee on Mass Storage Systems and Technology and is used to refer to a string of bits unconstrained by size and structure. The description of the various modules is given below:

- Bitfile Server: This module is responsible for the management of the logical storage of bitfiles. It recognizes files only in terms of their logical identifier and it interfaces with the Storage Server to send requests to the Storage Server.

- Bitfile Client: This module is the one that interacts directly with the user via the application programs. It obtains requests for access to files. Then its responsibility is to convert this named request into a logical request for the Bitfile Server using the Name Server.

- Bitfile Mover: This module establishes the facilities and the protocols for implementing high speed transfer of bitfiles between the Physical Volume Repository and the Bitfile Client and vice-versa.

- Name Server: This module is responsible for the mapping between logical bitfile identifiers and named files. It receives requests from the Bitfile Client to perform conversions between the human readable file names and the logical bitfile identifiers used by the Bitfile Server.

- Storage Server: Is responsible for handling the physical aspects of bitfile storage. It maintains the data structures for the file systems and it can form a hierarchy of multiple levels of physical storage.

- Physical Volume Repository: Is the module which deals directly with the various physical media available for mass storage such as hard disks and tape media.

- Site Manager: The responsibility of the site manager is to monitor and maintain statistics of the operation of all the other modules.

In general the overall system operates as follows: The one or more instances of the Bitfile Client module interact with the user through an Application Interface. The user requests for access to human readable named files are converted by the Name Server to logical bitfile identifiers. The user request may be a request to create, delete, retrieve or store a file. As soon as a bitfile identifier has been obtained from the Name Server, the request for the bitfile is passed on from the Bitfile Client to the Bitfile Server. The Bitfile Server keeps track of the requests received from the client and the replies received from the Storage Server. The Bifile Server then authenticates the access rights of the requestor and then forwards the request to the Storage Server in the form of action commands. The Storage Server maintains tables about each device included in the mass storage system and given the bitfile identifier can identify the exact location of the file in one or more of the hierarchical levels of the filesystem. Once the request has been authorized and the data has been located in the Storage Server or a file identifier has been allocated for the new file, one of the Bitfile Movers is used for the transfer of the data. The Bitfile Mover represents the high-performance data transfer path between the Bitfile Client and the Storage Server. It is assumed that all the clients and the servers shown in Figure 2 are interconnected through a communication service, which must handle all of the interprocess communications involved in synchronous data transfer.

4.0 The Unitree Central File Manager

4.1 Introduction

The Unitree Central File Manager (UCFM) is a hierarchical distributed file system. UCFM is a mass storage manager which provides a transparent uniform Unix like file system to the user. The layers of the hierarchy consist of hard disks at the first layer and both on-line or off-line tape storage at the second layer.

The data stored on the UCFM can be accessed from any local machine using either the FTP protocol or the NFS protocol. For performance reasons only the FTP protocol method is being used at NCCS. When files are first transferred to the UCFM they are stored on the first layer of the hierarchy. Then, through a process called migration, a copy of each file is made to a lower layer of the hierarchy so that the lowest layer will have a copy of every single file. Based on certain configurable parameters files from the highest layer are removed if they have not been accessed for a certain period of time. When the user tries to recall the file, UCFM knows the highest layer location of the file and accesses it from there. Thus files which are accesed often will be retrieved quickly whereas files which are not accessed too often will have longer access time. In a sense the disk at the highest layer is being used as a cache for the slower lower layers of the hierarchy. Some of the advantages of UCFM are listed below:

- UCFM does not impose any limit on the number of files and the size of the files in the file system.

- UCFM presents the familiar structure of a UNIX file system to the user.

- UCFM uses a client/server based architecture which allows for expansion into a distributed storage system.

- UCFM accesses the storage media at the raw device level so it supports any storage device supported by the file server machine.

4.2 UCFM System Architecture

UCFM is composed of a number of servers which manage the storage hierarchy. Each server is responsible for one specific task and thus the overall storage management task is distributed. This distribution of the responsibility and the functional separation of the components allows for load distribution, enhances the scalability of the storage system and provides more fault tolerance. Figure 3 shows a diagram of the UCFM servers and their interelation.

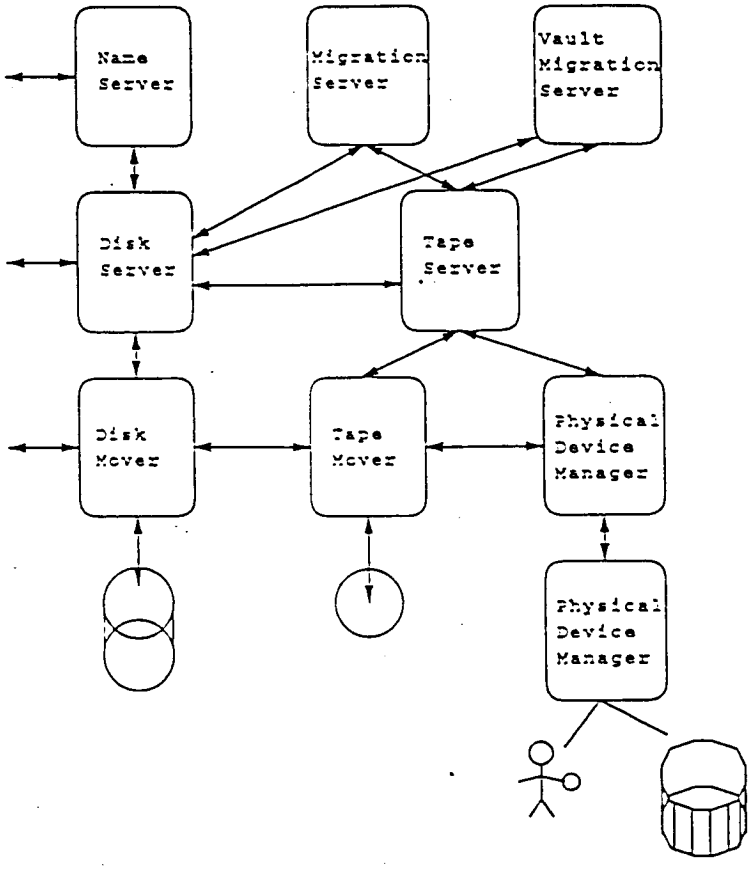


Figure 3. UCFM System Architecture

The above diagram basically shows the interconnection of the servers with each other. It serves as a good overview of the structure of the UCFM but it is not detailed enough to form the basis of a queueing network model since it is very general. Based on the request made by the user a job may have to visit each of the servers more than once and in a different order from what seems apparent. A description of each of the servers in the figure follows.

- Name Server: Its job is to maintain the Unitree filesystem structure and provide a transparent, Unix like interface to the Mass Storage System. It resolves human-oriented names to a globally unique machine-oriented resource identifier (bifile id).

- Disk Server: Provides the logical means for storing and retrieving data from the disk cache. It maintains the necessary header information for mapping a bitfile id into the actual file stored on the disk.

- Disk Mover: Its only purpose is to transfer file data to and from the disk cache. All requests to read and write data to and from the disk cache originate from the disk server. A response to each request is sent directly to the recipient of the file rather to the disk server.

- Tape Server: The tape server performs the equivalent service to tapes that the disk server performs to the disk cache. Its objective is to maximize the use of the storage media by archiving files. It maintains all the necessary information so that it can retrieve the information back from the tapes. It receives requests from the disk server and the migration server for access to files.

- Tape Mover: Its only purpose is to transfer file data to and from the tapes. It receives all its requests from the tape server.

- Physical Device Manager: Its job is to manage the tape mounts. It receives requests to mount tapes from the Tape Server and communicates its requests to the Physical Volume Repository to mount and dismount tapes. It schedules the tape mounts so as to optimize the efficiency of the mass storage system.

- Migration Server: As its name implies it moves data from the disk cache to lower levels of storage in the hierarchy in order to increase the size of the disk cache. It is triggered by two events: either the number of files which require migration has passed the set threshold or the period of no migration has reached the predefined threshold. Both thresholds are set by the system administrator.

4.3 UCMF and the IEEE MSSRM

It is interesting to examine how closely the UCFM model for their Mass Storage System adheres to the IEEE Mass Storage Systems Reference Model (MSSRM). This section describes this relationship and discusses which modules of the standard have not been implemented.

- Bitfile Server: The Bitfile Server has been implemented in two component modules in the UCFM, namely the Disk Server and the Tape Server. Both of these components perform the responsibilities of the Bitfile Server but each on a different level of the storage hierarchy. The mapping is not that clear because the Disk and Tape Server perform also some of the responsibilities of the Storage Server module, maintaining the filesystem structure.

- Bitfile Client: The Bitfile Client Server maps directly into the FTP and NFS daemons. Since the UCFM provides two methods for accessing the storage system it has one Bitfile Client for each method.

- Bitfile Mover: The Bitfile Mover has been implemented by two different modules in the UCFM, one for each of the levels of the storage hierarchy. The Disk Mover moves files between the disk and the user while the Tape Mover moves files between the tapes and the user.

- Name Server: This is a direct mapping from the MSSRM Name Server to the UCFM Name Server. The fact that the name of the component is consistent between the MSSRM and the specific implementation is an advantage and hopefully the rest of the components will eventually be renamed to correspond to the MSSRM standard.

- Storage Server: As it was already mentioned above the Storage Server's responsibilities are implemented by the Disk Server and the Tape Server modules of the UCFM. Separate modules are used for the different levels of the hierarchy.

- Physical Volume Repository: This module has been implemented by the Physical Device Manager.

- Site Manager: The Site Manager has no specific component that represents it in the UCFM. The monitoring and maintenance of statistics of the various modules is done by each individual module. A true Site Manager module would be a necessary addition to the existing UCFM and it would help tremendously in diagnosing system problems as well as doing performance analysis.

5.0 Data Analysis

In order to evaluate the performance of the overall system at NASA's Center for Computational Sciences we collected the log files from the ftp server of the Unitree. The format of the log files is described in the next section. The files were then processed with some awk scripts which were used to generate various statistics based on the ftp log file data. Another section describes the information which was generated by the awk scripts and finally the last section describes the results obtained by analyzing the generated graphs.

5.1 FTP Log Files

The ftp daemon of the Unitree system creates a log file every time the daemon is requested to perform an operation such as to open a connection, get a file, or put a file. Every time an operation is performed by the server an entry is made in the log file which contains a timestamp, the operation, the username of the person requesting the operation and the IP address of the host from where the request is made. The log files were processed by Edward Bender at Convex Systems. The reason for the processing is to make them more suitable for data analysis. The new processed files contain the following fields:

- Fields 1-8: Contain the date of the request (DDMONYY format)
- Fields 9-17: Contain the time of the request (hh:min:sec)
- Fields 18-25: Contain the command to be executed
- Fields 29-38: Contain the size of the file to be transferred in data transfer operations else its blank
- Fields 39-48: Contain the delay time of the tape drive in data transfer operations else the field is blank
- Fields 49-58: Contain the time required to transfer the file
- Fields 59-67: Contain the rate of the data transfer
- Fields 71-75: Contain the username of the user who made the request
- Fields 81-95: Contain the IP address of the machine from where the user made the request
- Fields 98-105: Contain the process id of the process which made the request
- Field 108: Contains a character which specifies the network type. An E stands for Ethernet and a U stands for Ultranet.

5.2 FTP Log File Processing

The log files for the month of August were processed and analyzed. The reason this specific month was chosen for analysis is because it is the most recent month of collected data and also because during the month of September the system at NCCS occasionally is off-line for hardware/software upgrading so the data wouldn't be representative of the usual system performance of the system.

Four different kinds of information were extracted from the log files using scripts written in the Awk language. In order to reduce the amount of redundancy only the log files from five randomly chosen dates of August were used. The first set of data was a histogram of the number of get and the number of put requests every hour of the day for the following days: August 1, August 5, August 12, August 18 and August 23.

The next set of data extracted for the same five days were the average file size transferred in get and put operations. The average file size was computed for each hour of the day.

The next set of data extracted were the average file transfer time for get and put operations.

In order to make the information in the plots that follow easier to read Tables 1 and 2 display averages over the range of a day. Table 1 shows the average file size in megabytes transferred for each of the get and put commands over the period of a day. Table 2 shows the average file transfer rate in kilobytes/sec for each of the get and put commands over the period of a day. This information will be very useful in the design of the queueing system and of the scheduler which will follow this study.

Day	8/01	8/05	8/12	8/18	8/23
Get	33.75	8.42	10.40	8.41	4.18
Put	10.07	9.42	10.25	11.99	10.57

Table 1. Average File Size Transferred in MBytes

Day	8/01	8/05	8/12	8/18	8/23
Get	44.81	138.95	161.41	85.44	47.33
Put	84.39	113.12	175.22	106.36	87.24

Table 2. Average File Transfer Rate in KBytes/sec

The last set of data extracted were the execution rate of the various ftp commands. We discovered that there were only four commands in the data files. Those are the LOGOUT, LOGIN, get and put commands. It is possible that mput and mget commands are inserted into the log files as multiple but individual put and get commands respectively. It is very likely that the other commands were filtered out of the data before the data was given to us. The information in Table 3 will be very important for developing a queueing model of the system for performance analysis.

Command	8/01	8/05	8/12	8/18	8/23
LOGIN	941	6761	1815	1300	1110
LOGOUT	936	6768	1791	1235	1080
get	182	1317	1254	583	281
put	102	1269	1517	1453	646

Table 3. Execution Rate of ftp Commands

6.0 Conclusions

From the point of view of the system performance analyst the Unitree system behaves as a black box since the source code is not available. The only possibility for making measurements on the system is by analyzing the data collected from the ftp logs as was done on the previous section. Based on the results of the analysis some insight was gained on the usage of the Unitree system which will be vital in designing a proxy scheduler which will increase the utilization factor of the system.

Looking at Figure 4 we can get a good idea of the utilization of the system over the period of a day. It is obvious that between the hours of 12:00 pm and 6:00 pm the system is being used heavily whereas between 12:00 am and 9:00 am the system is quite idle. This encourages the design of a scheduler which will provide more uniform load distribution on the system.

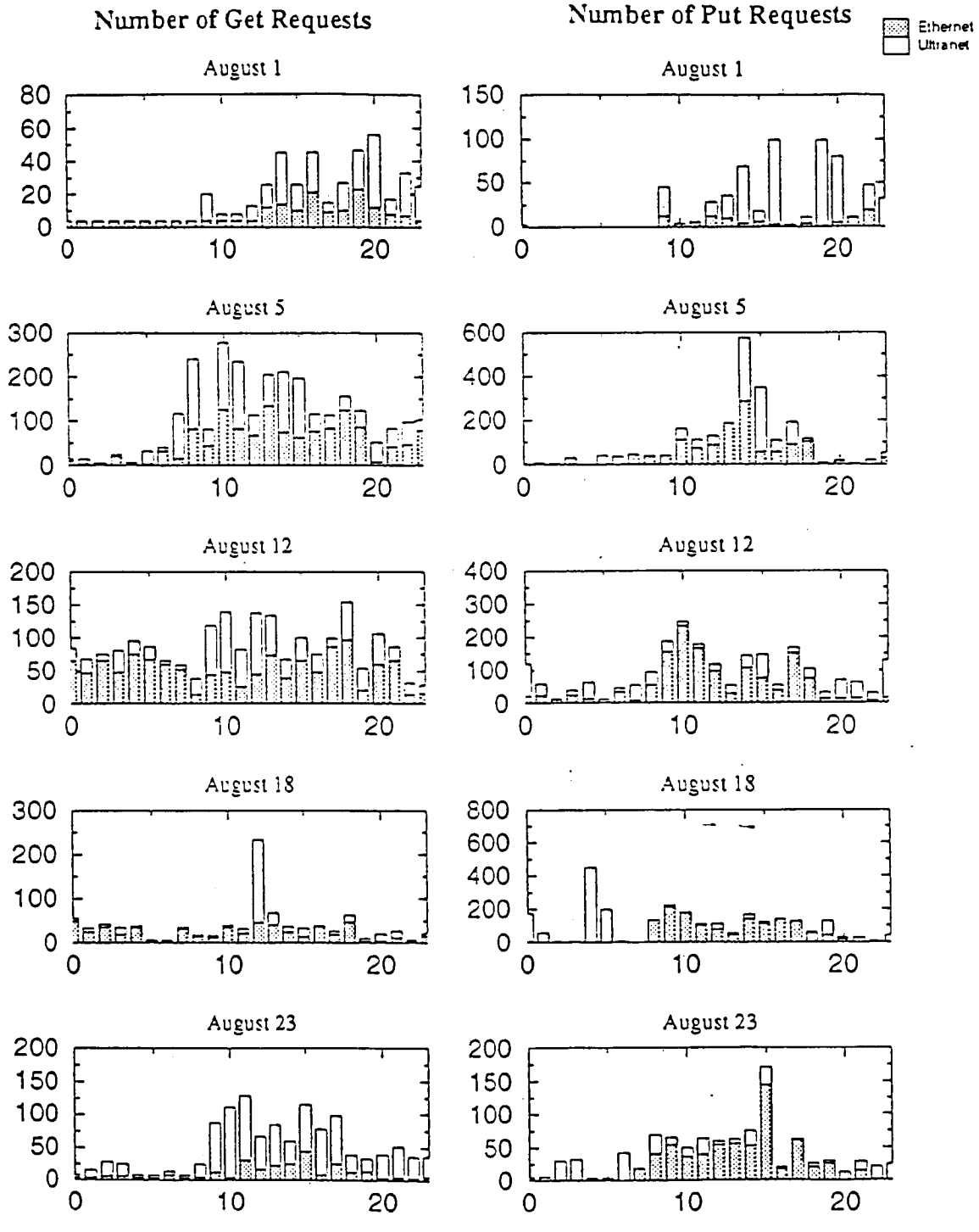


Figure 4. Histogram of Get/Put Requests Over the Hours of the Day.

Also some more information that we can gain from these diagrams is that the Ultranet is not getting as much use as the Ethernet even though it is the higher speed network and it should be the preferred choice of the user.

Finally the lack of a site manager and of a method for collecting statistics on the operation of the individual servers of the system make it very difficult for the performance analyst to model the system in an accurate way. The addition of a site manager for UCFM would be a worthwille addition to the system from the point of view of both the system administrator but also of the performance analyst.

References

[1] Sam Coleman and Steve Miller. Mass Storage Reference Model: Version 4. In *Goddard Conference on Mass Storage Systems and Technologies*, pages 1-76, 1992.

[2] P. J. Leach. Uids as internal names in a distributed file system. In *Proceedings of the Symposium on Principles of Distributed Computing*, pages 34-41, Ottawa, Canada, 1982.

[3] J. Mullender and A. S. Tanebaum. Protection and resource control in distributed operating systems. *Computer Networks*, 8: 421-432, 1984.

[4] Adina Tarshish and Ellen Salmon. The growth of the Unitree mass storage system at the NASA center for computational sciences. Technical report, NASA Goddard Space Flight Center, 1993.

[5] R. W. Watson. Identifiers (naming) in distributed systems. In *Distributed Systems - Architecture and Implementation*, 191-210. Springer-Verlag, 1981.

(Task 32)

UNIVERSITY OF COLORADO

Software Support Laboratory

Randal Davis
Laboratory for Atmospheric and Space Physics

In July 1993, CESDIS was tasked with coordinating communication between various research projects as well as collecting software, documentation, demonstration videos, and other materials resulting from funded research projects. This effort was subcontracted to the Laboratory for Atmospheric and Space Physics at the University of Colorado at Boulder which operates the Software Support Laboratory (SSL).

The SSL is a research project and an active software repository and distribution center which provides four services to users:

1. Online information: a listing of public domain and commercial software products that are likely to be of interest to NASA scientists. An online summary specifies the hardware and software operating environment for each program listed as well as examples of the output from the program and instructions for acquisition.

2. Online software archive: a number of NASA public domain software products available for online access through the Internet. Other software packages are maintained by the developer but can be accessed through the SSL.

3. Software testing and maintenance: The SSL is producing a set of CD-ROM disks for use in software testing. Representative datasets from most NASA science disciplines are provided, including images, maps, spectra, and tables in the most common formats for NASA applications.

4. Other services and products: Personnel at the SSL assist scientists in locating software for particular needs and help scientists use available software effectively.

Software described in the section of this report entitled *Research Activities: Peer Reviewed Projects, Applied Information Systems Research* will be deposited at the Software Support Laboratory. For information on available software or connections through Mosaic, send an e-mail request to ssl@sslab.colorado.edu. Summaries of project presentations at the July 1994 Applied Information Systems Research Program Workshop in Boulder will also be available through the SSL.

(Task 40)

UNIVERSITY OF NEBRASKA

Distributed Intelligent Data Management in Computer Vision Systems

Ashok Samal, Stephen Reichenbach, Phillip Romig
Department of Computer Science and Engineering

Task Objective

To bring together four on-going research projects:

- the DeViouS project at the University of Nebraska at Lincoln,
- the Intelligent Data Management project at the NASA Goddard Space Flight Center,
- the KRONOS scheduling program developed jointly by IDM and Honeywell, and
- the COLLAGE project at NASA Ames.

Each of these projects is looking at the problem of intelligent data retrieval and management systems problems from a slightly different angle. The main emphasis of this effort will be the integration of these projects, particularly with respect to computer vision systems.

CESDIS began providing support for Phillip Romig, the graduate student designated to work on the project, in May 1994. Phil is currently a Ph. D. student in the Department of Computer Science and Engineering. He has extensive programming experience in the design and implementation of distributed systems. He is working closely with the task originator at Goddard and has experience with all the systems to be dealt with as well as a general knowledge of remote sensing and ongoing NASA missions.

Phil will be working on the coordination of the projects listed in the task objective. This will include:

1. The integration of the NASA Ames COLLAGE planner with Goddard's Intelligent Information Fusion System. The IDM project at NASA Goddard has developed a prototype Intelligent Information Fusion System (IIFS) which provides a superior test environment for elements of a data management system.

2. Development of EOS domain applications using COLLAGE using KHOROS image processing. The COLLAGE planner was developed to be domain independent with flexible constraint modules that can be added or modified to run tasks in any domain. A target audience for the COLLAGE system is EOS. Remote sensing constraint modules for the COLLAGE systems based on KHOROS will be developed.

3. Exploration of the possibility of integrating COLLAGE with Honeywell's KRONOS scheduler.

(Task 44)

UNIVERSITY OF MARYLAND, COLLEGE PARK

Unsupervised Robust Estimation-based Clustering of Multispectral Images

Nathan Netanyahu
Center for Automation Research
Computer Vision Laboratory

Task Objective

To prepare for the challenge of handling the archiving and querying of terabyte-sized scientific spatial databases the Goddard Information Science and Technology Office has developed a number of characterization algorithms that rely on supervised clustering techniques. This task is aimed at continuing the evolution of some of these supervised techniques, specifically the neural network and decision tree-based classifiers, plus extending the approach to incorporating unsupervised clustering algorithms, such as those based on robust estimation techniques. The algorithms developed under this task should be suited for use by the Intelligent Information Fusion System metadata extraction modules, and as such these algorithms must be fast, robust, and anytime in nature. Finally, so that the planner/scheduler module of IIFS can oversee the use and execution of these algorithms, all information required by the planner/scheduler must be provided to the IIFS development team to ensure the timely integration of these algorithms into the overall system.

Dr. Netanyahu began work on this three-year task in May 1994 and is located on-site at Goddard. His efforts to date are reported below.

- Have further examined the TIROS Operational Vertical Sounder (TOVS) Pathfinder data. Conducted more discussions with various colleagues to appreciate prospective needs of NASA scientists, as far as automating TOVS data processing is concerned. Due to the very low resolution (level III data), we will focus for now on higher resolution data in an attempt to extract content out of images containing prominent weather phenomena (e.g., storms).

- Familiarized myself with the Probabilistic Neural Network (PNN)-based module developed at the ISTB for image classification. Checked that current implementations ran properly on a sequential machine and on the MasPar.

- Located a large source of Landsat TM data. Established a mechanism for reading and shipping these data onto the new mass storage device at the ISTB. Assistance will be provided by the Goddard Systems and Computer Operations Branch. Looked into a similar framework to populate our disk storage with a large amount of AVHRR data. Will obtain a small amount of sample data from the USGS.

- Interacted with LNK, Inc. personnel on matters related to classification of remotely sensed images. Will incorporate LNK's expertise and tools to gain "ground truth" data that is required for running our own clustering/classification modules.

- Have learned more about the Condor batch system that will assist in handling the large number of classification jobs to be submitted.

- Have read more literature on robust estimation and its application to image processing. Conceived the following idea:

  * Our robust estimation-based clustering scheme can also be used in the context of supervised clustering in this way:

    - Pick representative samples (i.e., training sets) of the various class types that are expected to be found in a scene.

    - For each class type, compute its estimated mean and covariance matrix by invoking Rousseeuw's minimum volume ellipsoid (MVE) robust estimator.

    - Classify the rest of the scene by employing classical Bayesian-like methods. An improved performance is expected since the mean and the covariance matrix use are more robust than those obtained due to a maximum likelihood estimator (MLE).

CONSULTANTS

The tasks included in this category have periods of performance ranging from a few days to a few months and require the expertise of an individual rather than several people in a university department. The contracted work may be performed on site at Goddard in close collaboration with Code 930 personnel or independently at an off-site location.

(Task 29)

Development of a Prototype for SE-Tree Learning for Automatic Data Cataloging and Characterization

Ron Rymon, President
Modeling Labs, Pittsburgh, PA

Task Objective

To develop a LISP prototype of SE-Learn, an SE-tree-based learning system.

Work on this task began in May 1994. Initial efforts involved streamlining and extending the current code and user interface towards porting the implementation to a NASA platform; the particular platform will be determined by IDM. Much of this work is aimed at parameterizing the user interface to make the program more easily accessible to scientists who are not intimately familiar with its internals. Many of the package's options are now accessible in a special file: parms.lisp.

SE-Learn has since been ported to a NASA platform (node dunloggin.gsfc.nasa.gov). This effort focused on adapting the code to the Allegro dialect; SE-Learn was originally developed in Lucid Lisp. The installed package is now being tested. Experimentation will hopefully begin in late July or August 1994.

(Task 30)

Multi-year K-12 Educational Outreach Plan

Margo J. Berg
MJB Consulting, Minneapolis, MN

Task objective

To assist in the formulation of a multi-year K-12 educational outreach plan for the NASA Office of High Performance Computing and Communications.

Several visits to NASA Headquarters and most of the NASA centers planning or executing K-12 educational programs under the HPCC program are expected. The purpose of the field center visits is to review center plans and activities in HPCC educational activities and provide guidance on improving the efficacy of the plans to maximize the educational benefit of the activities. Specific recommendations on the center visits will be provided to both the centers and Headquarters after each visit.

In May 1993 I reviewed proposals to develop High Performance Computing and Communications (HPCC) outreach programs to the K-12 education community from six NASA Field Centers involved in the HPCC program. I provided my assessment of the proposals to Paul Hunter in the HPCC HQ office.

I visited LeRC, ARC, JPL, GSFC, and LaRC in June and July to consult with program staff on the development of their HPCC outreach programs and to begin building collaborative ties between the center-based programs. A common need expressed by most of the program staff was the need to develop some overarching guidelines for the HPCC education program within which the centers could execute their programs. Also, center personnel expressed a desire to have regular and open communication among the various programs. Concensus among the personnel I talked with was that a workshop in the fall would be valuable to discuss common problems and solutions and to establish inter-center communication. Input and questions gathered during meetings with program staff at the site visits were used in the plan for a workshop to be held for center HPCC K-12 program personnel, which HQ staff expected would be held in early 1995 rather than the Fall of 1994.

HPCC HQ staff decided to issue a NASA Research Announcement (NRA) for FY94 funding, for which I drafted portions that would fund efforts to support curriculum development utilizing HPCC technologies. I reviewed several sources of information, including the U.S. EPA's Environmental Education Grants solicitation notice, the NSF's Applications of Advanced Technologies Program announcement, the *National Science Education Standards* in draft form being developed by the National Research Council, and the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for School Mathematics*. A preliminary draft of these portions of the NRA was submitted to Paul Hunter in late September, 1993.

I proposed developing a directory listing the teachers who are involved in the HPCC K-12 program along with their teaching discipline and grade level, email address, etc. My requests for information for this directory were answered by only a few participants. A follow-up with staff at the field centers on the agenda for the HPCC K-12 workshop and schedule also received very little response as of the end of September, 1993.

(Task 39)

Digital Libraries Consulting

Task Objective

Provide consulting services for digital libraries to develop plans and coordinate interagency, university, and industry collaborations.

This task provides the capability for entering into consulting agreements during the next three fiscal years. The first agreement was with Dr. Hans Mark of the University of Texas at Austin. Dr. Mark has been traveling extensively. His report will appear in the next annual report.

FELLOWSHIPS

CRAY RESEARCH SPACE SCIENCE FELLOWSHIP

Background

In May of 1990 Cray Research Inc. approached CESDIS with the idea of establishing a Cray Research Space Science Fellowship with CESDIS and Goddard Space Flight Center. The primary objective from Cray Research's perspective is to encourage the availability of high quality computer and information scientists for Earth science programs, such as those implemented by Goddard. The guidelines for this program were:

- Cray Research will fund a one-year graduate fellowship award for one student enrolled in a fulltime Ph.D. program at $20,000 per annum and will fund this award for the academic years 1991/92, 1992/93, and 1993/94.
- The fellowship program is to be established with CESDIS and is to be administered by Universities Space Research Association (USRA).
- The fellowship award is to be made to support research in applications relevant to NASA's global change program, specifically the Earth Observing System and Mission to Planet Earth.

Research topics of interest include, but are not limited to database management, parallel programming languages, data visualization, parallel algorithms, data compression, and image processing.

Recipients

The 1991/92 fellowship was awarded to Douglas Smith of Carnegie Mellon University who proposed to develop an intermediate language and virtual architecture for high performance image processing applications. The result of this effort is described in CESDIS Technical Report TR-93-96, *A Virtual Machine for High Performance Image Processing*, prepared by Mr. Smith.

The 1992/93 recipient was Kathleen Perez-Lopez for her proposal entitled *Use of an Index/Browse Set of Images for Database Management.* Ms. Perez-Lopez is an advanced doctoral student in the Department of Computer Science at George Mason University, and is advised by Dr. Arun Sood, professor of computer science and Director of the Center for Image Analysis at George Mason University. The final report submitted by Ms. Perez-Lopez may be requested as Technical Report TR-94-120.

The 1993/94 award went to Jonathan D. Bright, an advanced Ph.D. student in the Johns Hopkins University Computer Science Department. The goal of Mr. Bright's proposed work is to develop fault tolerant software for parallel processing and other high performance architectures while concentrating on problems derived from space science applications at the Hubble Space Telescope Science Institute. Areas of research are to include the design and analysis of parallel algorithms which use certification trails and the study of specific techniques for certifying these algorithms. Mr. Bright's principal advisors are Dr. Gregory Sullivan and Dr. Michael Goodrich. As with the other recipients, Mr. Bright's final report will become part of the CESDIS technical report series.

OTHER CESDIS ACTIVITIES

Task 27: High Performance Computing and Communications Program Coordination

Task 1: Conferences and Seminars

Task 1: Science Council Meetings

Task 41: Workshop on Multimedia Presentation Production

Task 32: Peer review support for NASA NRA-93-OSSA-09, *Applied Information Systems Research*

Task 33: MU-SPIN

Task 36: Computational meteorology and computer sciences support for the Earth Science Data Operations Facility

Tasks 26 & 39: Digitial Libraries

High Performance Computing and Communications
Program Coordinaton (Task 27)

Task Objective

To provide programmatic, logistical, and administrative assistance to the NASA High Performance Computing and Communications Program Manager. This includes:

- Assembling scientific, technical, and programmatic information,
- Reviewing and analyzing material,
- Exploring planning and development concepts and preparing papers, and
- Organizing meetings and workshops.

Activity Highlights
Thomas Hood, Technical Policy Analyst

Performed extensive work on the 1993 revision of the NASA HPCC Program Plan:

- Released versions 1.0, 2.0, and 2.1 of the level 1 document in August 1993.
- Began information gathering and editorializing of both level 2 volumes (HPCC and IITA).
- Released the final print version of 2.3 with all required approval signatures in September.

Attended teleconference between Lee Holcomb and Harvard University regarding policy issues for the National Information Infrastructure (NII) and the formation of a workshop to discuss those ideas.

Attended a teleconference between Lee Holcomb and Steve Wolfe of NSF in which workshop topics were discussed regarding the NREN Program in the development of the NII.

Attended the National Information Infrastructure Testbed (NIIT) Industry briefing to Congress and prepared a written report for Lee Holcomb.

Attended a Science, Space, and Technology subcommittee staff meeting with Lee Holcomb and Paul Hunter. The purpose of the meeting was to inform the House staff members of the accomplishments to date of the NASA HPCC Program (especially the ASTA component) as well as progress to date made by other Federal HPCC agencies.

Attended a teleconference among Lee Holcomb, Steve Wolfe, Mike St. Johns (ARPA), and John Cavallini (DoE) regarding the NREN Program in the development of the NII. Recorded minutes and drafted an issues letter to Brian Kahin of Harvard University's Kennedy School of Government.

Attended the CAS Quarterly Review held at NASA HQ.

Attended HPCC Working Group Executive Committee to discuss Systems Software R&D for the HPCC Program.

Compiled a list of levels 1 and 2 deliverables for both the HPCC and IITA Programs.

Released multiple versions of the Volume I (HPCC) Level 2 Program Plan in December to the HPCC Program Manager and the CAS and ESS Project Managers.

Attended two meetings with Codes 930 and 505 personnel and the CESDIS Acting Director to discuss the development of beneficial linkages among the ESS project, industry, and academia.

Distributed the final draft version of the HPCC Level 2 Program Plan to the HPCC Working Group Executive Committee members, the CAS and ESS project managers, the HPCC Program Manager, and the Director of the HPCC Office in January for their review and approval.

Assisted the ESS Project Manager in the compilation and distribution of the monthly ESS Project Reports which included technical progress, schedules, and financials.

Assisted the ESS Project Manager in the preparation of ESS review materials for the annual independent review and the comprehensive review of the HPCC program.

Began work on the ESS project's Mosaic presentation as part of the Space Data and Computing Division's Mosaic home page. Significant progress was made on the development of a Mosaic presentation that describes CESDIS and its programs.

Wrote a section on the national software exchange and the ESS Project for inclusion in the draft version of the 1994 Level 1 Program Plan for NASA HQ.

Larry D. Picha, Senior Program Coordinator

July 1993 — Prepared materials for the director of the NASA HPCC office for a confidential White House briefing on the NASA Information Infrastructure Technology and Applications (IITA) activity.

August — Traveled to participating NASA research facilities (Ames Research Center, Goddard Space Flight Center, Jet Propulsion Laboratory, Langley Research Center, and Lewis Research Center) to attend the HPCC program reviews. This was done to track the program as well as gather the necessary data and information to compile the 1993 HPCC technical highlights report.

Wrote responses to the White House Information Infrastructure Task Force (IITF) request for identification of Federal application projects that are related to the National Information Infrastructure efforts, across participating Federal agencies, and what they will accomplish. The task force's purpose is to coordinate the White House administration's efforts and promote applications of information technology in manufacturing, electronic commerce, education, health care, government services, libraries, and other areas.

Planned, organized, and attended the quarterly reviews for the Earth and Space Sciences (ESS) project and the Computational Aerosciences (CAS) project.

September — Prepared materials and presentations for the NASA Office of Aeronautics (Code R) including the HPCC Quarterly Review and the HPCC Baseline Data.

October — Wrote material, prepared presentations and attended the HPCC working group executive committee and the HPCC working group technical committee meetings.

November — Collaborated with new NASA headquarter's staff to set up a new monthly reporting system for HPCC projects and the Numerical Aerodynamic Simulation (NAS) program at NASA research centers (ARC, GSFC, JPL, LaRC, LeRC).

Planned, coordinated, prepared materials for, and attended the High Performance Computing and Communications working group technical committee meeting. Emphasis was placed on systems software, software exchange, and basic research issues.

December — Organized an open meeting with scalable computing systems software vendors sponsored by the Federal Coordinating Council for Science, Engineering and Technology (FCCSET) and the High Performance Computing, Communications and Information Technology (HPCCIT) subcommittee. The National Coordination Office for High Performance Computing and Communications assisted on this project.

Coordinated the terminal-based packet video teleconference to be conducted via Sun workstations. This was a first-time demonstration of this capability over the Internet at NASA Headquarters.

January 1994 — Planned, organized and attended the quarterly reviews for the Earth and Space Sciences project and the Computational Aerosciences project.

Coordinated and communicated requirements for the Annual Independent Review of the High Performance Computing and Communications Program to Research Center Management at NASA Headquarters.

February — Attended the open meeting with the scalable computing systems software vendors sponsored by the Federal Coordinating Council for Science, Engineering and Technology (FCCSET) and the High Performance Computing, Communications and Information Technology (HPCCIT) subcommittee, held at the National Coordination Office for HPCC.

Planned, coordinated, prepared materials for, and attended meetings of the High Performance Computing and Communications working group technical committee

March — Wrote and prepared materials for the Annual Independent Review of the High Performance Computing and Communications Program at NASA Headquarters.

April — Planned, organized, and attended the quarterly reviews for the Earth and Space Sciences project and the Computational Aerosciences project.

May — Wrote the FY 1995 HPCC level-1 program plan based on previous year's programmatic changes and funding shifts.

June — Finalized work on the NASA HPCC technical highlights report as well as the FY 1995 HPCC level-1 program plan.

Began writing the program plan addendum to be submitted to the Office of Management and Budget.

Michele O'Connell, Program Coordinator

January 1994 — Prepared graphics and text coordination between NASA Headquarters Printing and Graphics Office and the Government Printing Office for the Global Quest Brochure created by NASA National Research and Education Network (NREN) Project in cooperation with NASA's Education Division. The NREN Project is part of the High Performance Computing and Communications program, a Presidential Initiative to sustain and extend U.S. leadership in all areas of advanced computing and networking.

Coordinated tracking material and attended Quarterly Reviews for the Earth and Space Sciences Project and the Computational Aerosciences Project.

February — Assisted in the coordination of the first Enabling Technologies for PetaFLOPS Computing Workshop held in Pasadena, California. More than 60 experts in all aspects of high-perfor-

mance computing technology met to establish the basis for considering future research initiatives that will lead to the development, production, and application of PetaFLOPS scaled computing systems. Follow-on activities include presentations and briefings to senior NASA and White House administration personnel, as well as the final publication of the workshop report, "Enabling Technologies for PetaFLOPS Computing."

March — Coordinated the HPCC K-12 Educational Workshop at Goddard Space Flight Center. This was a three day Information Infrastructure Technology and Applications (IITA) activity to review NASA Center K-12 programs. Sessions were held to discuss school selection processes, teacher resource center technology strategies, curriculum development, network infrastructure, as well as evaluation and repositories necessary to support current and future projects.

April — Assisted with the HPCC Cooperative Agreement Notice preproposal conference (CAN-OA-94-1) *Public Use of Earth and Space Science Data Over the Internet* held at Goddard Space Flight Center. The notice solicited proposals for development of innovative applications of U.S. Earth and space science remote sensing databases via computer networks, the development of digital libraries technology and the establishment of a remote sensing public access center.

May — Attended the Cooperative Agreement Notice Proposal Selection Conference, *Public Use of Earth and Space Science Data Over the Internet*, held in Greenbelt, MD. Performed statistical analysis of proposal submissions and logistical coordination throughout the conference.

June — Finalized the NASA HPCC Classic Milestone Tracking Reporting System. Also finalized the HPCC Graduate Student Research Program Implementation Process. This consisted of developing a tracking system for grants awarded graduate students, their funding renewals, and the number of contracts. Also wrote the 1995 Graduate Student Research Program submission.

Conferences and Seminars (Task 1)

Data Compression Conference '94 (DCC '94)

DCC '94, the fourth annual Data Compression Conference, was held at Snowbird, Utah March 29-31, 1994. Sponsored by the IEEE Computer Society Technical Committee on Computer Communications, the conference was organized by James Storer and Martin Cohen of Brandeis University with information distribution and registration assistance from the CESDIS administrative staff. The intent of the conference was to provide an international forum for current work on data compression and related areas.

Bulk mailings were distributed to the CESDIS university faculty database and non-university members of the IEEE Computer Society in October 1993 and January 1994. CESDIS staff members also prepared the confirmation and registration packets, conference badges, and registration lists. Robin Alford and Annemarie Murphy attended the multi-day conference to coordinate the on-site registration and organization activities for the conference, poster session, and associated workshops. Three hundred twenty individuals representing U. S., Canadian, European, and Asian university, government, and industry organizations attended the technical sessions or the workshops.

Copies of the call-for-papers announcement and conference program are included in Appendix B of this report. Also included is a list of the universities, companies, and other organizations represented at the conference. A copy of the conference proceedings, which includes the papers presented at the technical sessions and extended abstracts from the presenters at the poster session, is available from the IEEE, 445 Hoes Lane, Piscataway, NJ 08855. Phone orders may be placed to 800-678-4333 or 201-562-3800.

Future Earth Remote Sensing Missions Seminar Series

Organized by Staff Scientist Jacqueline Le Moigne and sponsored by CESDIS, this series of eight seminars was held during the fall of 1993 at Goddard Space Flight Center. Following the general introduction to Earth remote sensing provided by the CESDIS Spring 1993 seminar series, this new series focused on future Earth remote sensing missions. Directed mainly towards computer scientists and engineers, but of interest to a more general audience, these seminars provided up-to-date information about future satellites and sensors in preparation for the Mission to Planet Earth (MTPE) and the Earth Observing System (EOS) programs.

The series consisted of eight seminars. The first seminar provided an overview of future Earth remote sensing missions and instruments which will be flown for the next ten years as part of the EOS series or as separate MPTE missions. Goddard Space Flight Center, responsible for 17 of the 24 current flight missions, is especially involved in MODIS (Moderate-Resolution Imaging Spectroradiometer) for which 5 copies will be flown throughout the EOS program, GLAS (Geoscience Laser Altimeter System), and SeaWiFS (Sea-Viewing Wide Field Sensor). Subsequent talks focused on particular missions. For each mission, the scientific objectives were described, followed by the selected investigations and the proposed instrument(s) and overall project. The topics covered applications in geodynamics (LAGEOS III, Laser Geodynamics Satellite), studies of climate change and climate models (CLIMSAT), investigation of the energetics and dynamics of the mesosphere and lower thermosphere/ ionosphere (TIMED, Thermosphere Ionosphere Mesosphere Energetics and Dynamics), soil moisture measurements (ESTAR, Synthetic Aperture Radiometer Instrument), weather prediction and climate modeling (METSATS, Meteorological Satellites), and finally two planned EOS missions, EOS-CHEM measuring the atmospheric chemical composition, and EOS-ALT which will provide measurements about ocean circulation and ice sheet mass balance.

The average attendance for each talk of the series was about 30 people with a peak at 60 for the first seminar, which shows the wide existing interest in the future directions of Earth remote sensing within the science community. Video tapes were made for each of the talks and are part of the CESDIS video library facility, available for all researchers desiring to update their knowledge in one of the areas of remote sensing presented during the two 1993 seminar series. A copy of the series announcement, speaker abstracts, and a list of the companies and institutions represented are included in Appendix A of this report. A list of the attendees and copies of the presenter transparencies are included in CESDIS technical report TR-93-108, *Summary Report of the CESDIS Seminar Series on Future Earth Remote Sensing Missions*, prepared by Dr. Le Moigne.

Other CESDIS Seminars

1993

October 18 — Theodore Johnson , University of Florida. *UFMulti: A Distributed System for Processing Experimental Data.*

October 22 — Jaideep Srivastava, University of Minnesota, Computer Science Department. *PADMA: A Parallel Database Manager for Large-Scale Datasets.*

November 5 — Matthew O'Keefe, University of Minnesota, Department of Electrical Engineering and Army High Performance Computing Research Center. *I/O and Language Issues in High Performance Computing.*

1994

January 12 — John Corliss, Space Biospheres Ventures. *Biosphere 2: An Experimental Facility for Research on Ecological Systems.*

January 13 — Jack Schwartz, New York University, Multi-Media Center. *Multi-Media Initiatives at NYU.*

January 14 — Michael Stonebraker, University of California at Berkeley, Computer Science Division. *A Technical Agenda for Digital Library Research.*

January 24 — David Ebert, University of Maryland, Baltimore County, Computer Science Department. *Rendering, Animation, and Realistic Visualization of Gases.*

February 4 — Nabil Adam and Ramesh Subramanian, Rutgers University and the University of Alaska. *The Design and Implementation of an Expert Object-oriented Spatial Database Model.*

March 10 — Michael Keeler and Farzad Mahootian, Gonzaga College High School, Earth System Science Lab. *Ecologica: A Network-Enabled High School Course for Earth System Science.*

April 28 — David Van Buren, California Institute of Technology and Jet Propulsion Laboratory, Infrared Processing and Analysis Center. *AstroVR - A Collaborative Environment for Astronomy Research.*

June 8 — Thomas Ruwart, University of Minnesota, Graphics and Visualization Laboratory. *How to Implement High Performance Storage Server Architectures with Workstation Technology.*

June 9 — Matthew O'Keefe, University of Minnesota, Department of Electrical Engineering. *Comparing Cray T3D, Cray C90, And SGI Challenge Performance for an Ocean Circulation Application Code.*

CESDIS Science Council Meetings

Meetings of USRA's CESDIS Science Council were held at Goddard on January 10-11 and July 7, 1994. The function of the council is to act as a scientific board of directors for CESDIS and to provide a semiannual review of the performance of CESDIS to USRA's Board of Trustees. Presenters at the January session included Terrence Pratt (CESDIS Acting Director), Thomas Sterling and Jacqueline Le Moigne (CESDIS Staff Scientists), Alok Choudhary (Syracuse University), Andrew Grimshaw (University of Virginia), Diane Cook (University of Texas, Arlington), Theodore Johnson (University of Florida), and David DeWitt (University of Wisconsin). The last five are Principal Investigators under the CESDIS HPCC University Research Program in Parallel Computing. Two additional PIs attended but did not present. Discussions were organized and led by the Acting Director concerning the major CESDIS programs in HPCC, Digital Libraries, Applied Information Systems, and EOSDIS. Various NASA and CESDIS personnel participated and made presentations for these discussions.

The focus of the second meeting was the director search. Presentations were also made by Milt Halem on the digital libraries program, Thomas Sterling on the PetaFLOPS Workshop, and Bill Campbell on a new EOSDIS project.

Workshop (Task 41)

How to Produce a Multimedia Presentation

From March through June, 1994, CESDIS hosted a workshop at Goddard on the development and use of multimedia materials. The workshop, titled "How to Produce a Multimedia Presentation," was conducted by the Center for Digital Multimedia of New York University (NYU), which is headed by Dr. Jack Schwartz.

The workshop consisted of ten two-hour sessions. Dr. Schwartz led the first session on March 15. Later sessions were taught by Kerry O'Neil, NYU Multimedia Specialist. About 20 Goddard civil service and contractor employees participated, drawn primarily from the Earth Science (Code 900) and Space Science (Code 600) Directorates.

Presentations covered the following topics:

- Descriptions of available multimedia production tools, stressing readily available commercial tools.
- Multimedia materials integration, including text, sound, animation, and image data.
- Issues of design and style.
- Reviews of the best existing products.
- Issues of CD-ROM production.

Participants developed practice projects on six Apple Macintosh Quadra computers made available by the NYU Center during the workshop. A sample presentation module titled "Sea Ice in the Polar Regions" was designed and developed by participants from the EOS-PM project and the NYU staff. Other practice projects developed by workshop participants included "MODIS Project Overview," "Researching Climate Change," "Astrophysics Data to the People," and "Digital Library Technology Studio."

At the final session on June 10, Dr. Milt Halem, Chief of the Space Data and Computing Division, hosted a luncheon to present the workshop projects to Goddard officials, including Dr. John Klineberg, Center Director. The projects effectively showed the power of the multimedia technology as a means to communicate about Goddard programs to a broad audience.

Funding for the workshop was provided by the EOS-PM project under Dr. Claire Parkinson (Code 971) and the Digital Libraries Technology project under Dr. Nand Lal (Code 935).

Peer Review Support for NASA NRA-93-OSSA-09, *Applied Information Systems Research* (Task 32)

Two hundred twenty-five proposals were received in response to this new research announcement. They were distributed by CESDIS administrative personnel to 139 physical and computer scientists for a mail peer review, with evaluations scheduled for return by August 20, 1993. Tabulations began late in August and continued in September as last-minute evaluations were faxed by reviewers.

Sixty-one physical and computer scientists were recruited to serve as peer review panel members, 41 of whom also served as mail reviewers. Panels were convened on September 20-21 at the Holiday Inn Capitol in Washington, DC. All panelists met in a general session on the morning of September 20 to receive their instructions from Glenn Mucklow.

Panel chairs met with Glenn Mucklow on Wednesday, September 22 to present the proposal rankings and rationale for each panel and to develop a single list of proposals recommended for funding. As the final activity on this task, CESDIS personnel prepared a briefing book for the members of the Selection Committee and turned all materials over to Glenn Mucklow.

Minority University—Space Interdisciplinary Network (MU-SPIN) Support (Task 33)

MU-SPIN, the Minority University—Space Interdisciplinary Network is a networking and education initiative for historically black colleges and universities (HBCUs), minority universities (MUs), and other academic institutions with large minority student enrollments. The main goal of the MU-SPIN program is to transfer computer networking technology and its applications to HBCUs and MUs in support of collaborative interdisciplinary scientific research. The program offers technical and logistical assistance in training to faculty, staff, and students in a wide range of network-related issues.

During the period July 1993 to June 1994, CESDIS received funding to provide administrative support to the MU-SPIN Program Coordinator. The administrative assistant responsible for most of that support left CESDIS in August 1993. Since similar support was being provided by a student from Morgan State University through another contractor, it seemed an appropriate time to reduce the CESDIS support. Vendor and travel vouchers for the 1993 Users Working Group Conference were processed by CESDIS, but further involvement in the program was phased out by January 1994.

Earth Science Data Operations Facility Support (Task 36)

Robert Mack

Task Objective

Provide support in computational meteorology and computer sciences to the Code 902 Earth Science Data Operations Facility to include:

- Working with Tropical Rainfall Measuring Mission (TRMM) scientists and examining those science algorithms which require improvement in execution speeds and computation accuracy;
- Utilizing computer science or information science techniques to improve the computational performance of science algorithms;
- Providing advice concerning computational meteorology, computer science, and other innovative technologies applicable to data systems to the TRMM Science Data and Information System (TSDIS) Project;
- Providing advice concerning software engineering methodologies, implementation, and standards to TSDIS developers;
- Representing TSDIS Project and presenting results concerning computational meteorology, computer science, and data system technologies in meetings and conferences.

Robert Mack performed these functions as a USRA employee from mid-October until mid-June 1994 when he became a civil service employee. A representative sample of his activities includes the following:

- Reviewed drafts of TRMM plans and papers.
- Attended planning and review meetings for the configuration management plan, the software management plan, SeaWiFS software review, TSDIS system requirements, NASDA-TSDIS interface, ground validation algorithm developers, MOC-TSDIS interface, and TRMM-ECS interface.
- Worked on the refinement of the TSDIS-SDPF interface which included studying the generic ICD between SDPF and <mission>, the generic ICD between DDF Phase II and consumer system, and white papers on the SDPF produced by MTI.
- Evaluated the browse data system.

Digital Libraries (Tasks 26 and 39)

Thomas Hood

Attended meetings between Code 930 and Mitre Corporation personnel to discuss the role of the Digital Library Foundation and the future technical and strategic direction of the digital libraries project.

Participated in the weekly New York University (NYU) workshop on multimedia presentation preparation. A workshop project was developed on digital library technology. The results will be demonstrated in the Digital Library Technology Studio.
Presented the digital library technology multimedia presentation to former NASA Deputy Administrator Hans Mark, Goddard Director John Klineberg, Goddard Associate Director Jim Trainor, and Code 100 senior management. Also demonstrated the presentation to NIST officials.

Linked the digital library technology multimedia presentation to other multimedia projects developed in the NYU multimedia workshop for eventual recording onto compact disk for portability.

Technical Report Series

The technical report series inaugurated in September 1990 now contains 122 papers and reports contributed by the Director, Associate Director, Staff Scientists, Cray Fellowship recipients, and funded project personnel. A copy of the report abstracts and an order form are included as Appendix C.

Dissertation Series

Copies of dissertations written by Ph.D. students funded through CESDIS are now available for distribution. Abstracts and an order form are included in Appendix D.

APPENDIX A

Future Earth Remote Sensing Missions Seminar Series

October 5 - November 30, 1993

LIMITED SEATING AVAILABLE

PLEASE CALL OR FAX TO RESERVE YOUR SEAT!

Georgia Flanagan
CESDIS Code 930.5
NASA/GSFC
Greenbelt, MD 20771
Tel: 301-286-2080
Fax: 301-286-5152
(after October 1, the fax number will be 301-286-1777)

Name: _____

Affiliation: _____

Address: _____

_____

Phone: _____

Fax: _____

E-mail: _____

☐ I am a foreign national without a green card

CESDIS

CENTER OF EXCELLENCE IN SPACE DATA AND INFORMATION SCIENCES

Goddard Space Flight Center
Code 930.5
Greenbelt, MD 20771
301-286-2080

CESDIS

THE CENTER OF EXCELLENCE IN SPACE DATA AND INFORMATION SCIENCES

invites you to attend

Future Earth Remote Sensing Missions: Fall Seminar Series

every Tuesday
October 5 - November 23
2:00PM to 3:00PM

GSFC
Building 28
Room E210

Earth Remote Sensing Seminars in Cooperation with the IEEE Geoscience and Remote Sensing Society

Future Earth Remote Sensing Missions

The study of the Earth as an integrated system of interacting components is essential to assess the implications of global environmental changes, both natural and human-induced. To achieve such a global understanding of Earth evolution and global change, satellite remote sensing systems provide a major contribution through comprehensive and long-term global measurements.

Following the general introduction to Earth remote sensing provided by the CESDIS Spring '93 seminar series, this new series will focus on future Earth remote sensing missions. Directed mainly towards computer scientists and engineers, but of interest to a more general audience, the goal of these seminars is to provide up-to-date information about future satellites and sensors in preparation for the Mission to Planet Earth (MTPE) and the Earth Observing System (EOS) programs.

The first seminar will provide an overview of future Earth remote sensing missions. Subsequent talks, which will be introduced as part of the Global Change Research Program, will focus on particular missions. The eight seminars will be held weekly at Goddard Space Flight Center, Building 28, on Tuesdays at 2:00pm, starting October 5, 1993. The series, sponsored by CESDIS and NASA, is organized in cooperation with the Washington/Northern Virginia Chapter of the IEEE Geoscience and Remote Sensing Society.

Some seminars will be organized in cooperation with a Tea and Poster session held in the Atrium, Building 28, from 3:00PM till 4:00PM (call Dr. R. White @286-7802, for more information).

For more details, contact:

Dr. Jacqueline Le Moigne
CESDIS
NASA/GSFC
Greenbelt, MD 20771
Tel: (301) 286-8723
lemoigne@nibbles.gsfc.nasa.gov

All attendees who do not work on site at Goddard Space Flight Center are required to contact:

Georgia Flanagan
CESDIS
NASA/GSFC
Greenbelt, MD 20771
Tel: (301) 286-2080
Fax: (301) 286-5152
(286-1777 after October 1)
georgia@cesdis1.gsfc.nasa.gov

Foreign nationals with no green cards should contact Ms. Flanagan at least 2 weeks prior to the desired attendance date.

October 5	Dr. D. Zukor *Future Earth Remote Sensing Missions* (introduction by Dr. T. Pratt)
October 12	Dr. D. Rubincam *Lageos III: the Proposed Geodynamics Satellite* (introduction: to be announced)
October 19*	Dr. J. Hansen *Climsat* (introduction by Dr. M. Halem) **held in Skybox, Bldg. 28**
October 26	Dr. H. Mayr *TIMED, Thermosphere Ionosphere Mesosphere Energetics and Dynamics Mission* (introduction by Dr. R. Hartle)
November 2	Dr. D. LeVine *Synthetic Aperture Radiometer for Remote Sensing of Soil Moisture: ESTAR* (introduction by Dr. L.Thompson)
November 9	Dr. J. Susskind *NOAA Metsats* (introduction by Dr. F. Einaudi)
November 16	Dr. J. Zwally *Measurement of Polar Ice Changes with Satellite Altimetry* (introduction by Dr. C. Koblinsky)
November 23	Dr. J. Gleason *EOS-Chem* (introduction by Dr. M. Schoeberl)

EOS CHEMISTRY - The EOS Contribution to Our Understanding of the Atmosphere

Dr. James Gleason

NASA Goddard Space Flight Center, Laboratory for Atmospheres,
Atmospheric Chemistry and Dynamics Branch

EOS CHEMISTRY scheduled for a 2002 launch will provide global measurements of trace gases that are important for our understanding of the mechanisms that cause climate change and ozone depletion. Recent aircraft and satellite missions have shown that to advance our knowledge of the atmosphere, simultaneous measurements of several interacting species are required. EOS Chemistry will provide these measurements on a global scale.

The current configuration, building on the successful Upper Atmospheric Research Satellite, consists of several instruments that measure trace species and two instruments that monitor the solar input to the atmosphere.

The CHEM instruments are :

- *Microwave Limb Sounder (MLS): measures microwave trace gas emission from the limb*
- *High Resolution Dynamics Limb Sounder (HIRDLS): measures infrared trace gas emission from the limb*
- *Stratospheric Aerosol and Gas Experiment III (SAGE III): measures trace gases and aerosols using solar and lunar occultation*
- *Solar Stellar Irradiance Comparison Experiment II (Solstice II), UV Flux*
- *Active Cavity Radiometer Irradiance Monitor (ACRIM): Total Solar Irradiance*

The mission, the instruments and the science behind the measurements will be presented.

Dr. Gleason received his Ph.D. (Chemistry) in 1987 from the University of Colorado, Boulder. He did postdoctoral work at Brookhaven National Laboratory and then came to Goddard as an NRC Research Associate in the Astrochemistry Branch of the Laboratory for Extraterrestrial Physics. He moved to the Atmospheric Chemistry and Dynamics Branch in 1991, where his recent research has focused on remote sensing of atmospheric trace gases, primarily ozone.

CLIMSAT: Proposed Mission to Measure Global Climate Forcings and Feedbacks

Dr. James Hansen

NASA Goddard Institute for Space Studies

Empirical and theoretical evidence indicate that the Earth's climate system is sensitive to small global climate forcings. Precise measurements exist for the magnitude of certain climate forcing mechanisms, such as increasing atmospheric carbon dioxide, but other suspected forcings are measured crudely or are largely speculative. As a result we do not know the present net forcing of the climate system, which in turn means that it will be impossible to interpret quantitatively future climate change, even if substantial global changes are observed. This is a situation that should be of concern to policy makers.

The proposed CLIMSAT mission would address many of the data deficiencies with a pair of small satellites, one in a sun synchronous polar orbit and the other in an inclined precessing orbit. Each of the two satellites

would contain three instruments, and be launchable by a Pegasus-class vehicle. The instruments would measure both the reflected solar and emitted thermal spectra in manners optimized to yield information on climate forcings and radiative feedbacks, all of which operated by modifying the solar and/or thermal spectra.

Jim Hansen got his Ph.D. in space physics from the University of Iowa in 1967. He has been at GISS (Code 940) since then. He has carried out research in remote sensing of planetary atmospheres and in studies of climate change, including data analyses and the development of global climate models.

Synthetic Aperture Radiometer for Remote Sensing

Dr. D. M. Le Vine

NASA Goddard Space Flight Center,
Laboratory for Hydrospheric Processes, Microwave Sensors Branch

Soil moisture is a highly variable reservoir in the global hydrologic cycle that stores and distributes precipitation. Soil moisture is an important parameter in water budget studies, in agriculture and for understanding climate (e.g., a boundary condition for global circulation models). Passive microwave sensors at long wavelengths respond strongly to changes in the moisture content of soil because of the strong contrast between the dielectric constant of water (large) and that of dry soil (small). But long wavelengths mean large antennas in orbit which presents engineering challenges. To overcome such problems a novel remote sensing instrument, a synthetic aperture microwave radiometer called ESTAR, is being developed for remote sensing from space. An aircraft prototype operating at L-band (1.4 GHz) has been built, in collaboration with the University of Massachusetts, to demonstrate the concept. This instrument has been flown successfully in experiments conducted in cooperation with the USDA Agricultural Research Service to measure soil moisture, and preliminary designs have been developed for an "Earth Probe" to measure soil moisture from space.

Dr. Le Vine is a member of the Microwave Sensors Branch at NASA's Goddard Space Flight Center where his research has focused on the development of techniques for microwave sensing of the environment from space.

David M. Le Vine received his Ph.D. in electrical engineering from the University of Michigan (1968) and also earned his M.S. degree in physics and electrical engineering while at Michigan. Prior to coming to Goddard, Dr. Le Vine was Assistant Professor of Electrical Engineering at the University of Maryland and prior to that worked at the Radiation Laboratory, University of Michigan.

TIMED Mission–Thermosphere, Ionosphere, Mesosphere: Energetics and Dynamics

Dr. Hans G. Mayr

NASA Goddard Space Flight Center,
Planetary Atmospheres Branch, Laboratory for Atmospheres

The TIMED mission is the first of the new Solar Terrestrial Probes for NASA's Space Physics Division and is considered for launch before the end of the century. Spacecraft in high and low inclination orbits would carry instruments for remote sensing and in-situ measurements to investigate the energetics and dynamics of the mesosphere and lower thermosphere/ionosphere in the altitude range between about 50 and 200 km. An overview is given of the scientific objectives, the selected investigations and the proposed mission architecture.

Hans G. Mayr received his Ph.D. in physics from the University of Graz (Austria) and came to this country on a NRC fellowship. A staff member of the Laboratory for Atmospheres and the Project Scientist for TIMED, he has been involved in developing models of planetary ionospheres and upper atmospheres.

LAGEOS 3: The Proposed Geodynamics Satellite

Dr. David Rubincam

NASA Goddard Space Flight Center,
Laboratory for Terrestrial Physics, Geodynamics Branch

Goddard's own John A. O'Keefe first proposed using retroreflector satellites for geodetic purposes in the mid-1950's. However, he was too far ahead of his time. Only after the laser was invented did they become feasible and take shape in the form of LAGEOS, a dense, retroreflector-covered ball tracked with laser beams.

The LAGEOS satellites have proved to be remarkably useful in geophysics. LAGEOS 1, a U. S. satellite, was launched in May of 1976 in order to study the Earth's gravity field, tectonic plate motion, polar motion, and tides. It has succeeded spectacularly; so much so that LAGEOS 2 was built by the Italian space agency and launched in October of 1992 aboard the Space Shuttle and placed in orbit by an Italian booster. Now LAGEOS 3 is on the drawing boards, also to be built by Italy. The primary goal of LAGEOS 3 is to measure the Lense-Thirring effect, an effect predicted by Einstein's general theory of relativity but never before observed.

Dave Rubincam got his Ph.D. in physics at the University of Maryland in 1973. He has been a member of the Geodynamics Branch (Code 921) since 1978. He has published several papers dealing with non-conservative forces on LAGEOS. His other interests include possible secular obliquity changes on Mars and radiative forces on small asteroids.

NOAA Polar Orbiting Meteorological Satellites (METSATS)

Dr. Joel Susskind

NASA Goddard Space Flight Center,
Laboratory for Atmospheres, Satellite Data Utilization Office

NOAA maintains a 2 satellite (2:30 a.m., p.m. local time equator crossing and 7:30 a.m., p.m. local time equator crossing) suite of sun-synchronous polar orbiting satellites to provide data useful for aiding weather prediction, monitoring climate, and other purposes. The current satellite series started with TIROS-N in 1979, with similar instrumentation flown on subsequent NOAA satellites to the present, and scheduled to fly on NOAA J in 1994. An improvement will be made on NOAA K-N covering the period 1995-2005. Potential for major improvements in remote sensing capability exists for NOAA O, in 2005, and subsequent satellites, taking advantage of new instrumental developments such as AIRS and MODIS, to be first flown on the NASA EOS PM platform in 2000. Capabilities of the current satellite series will be shown, as well as options and projected improvements for an advanced sounding system.

Dr. Susskind is Head of the Satellite Data Utilization Office, METSAT Project Scientist, and a member of the EOS AIRS science team. He is currently involved in joint NASA/NOAA/DOD convergence studies for future satellite missions.

Future Earth Remote Sensing Missions

Dr. Dorothy J. Zukor

NASA Goddard Space Flight Center
Deputy Director, Earth Sciences Directorate

This talk will address some of the approved Earth science spacecraft planned for the 1990's and will describe some of the instruments currently under development as aircraft instruments. Some of the EOS instruments Goddard is sponsoring and some of the in-house development work will be described.

Dr. Zukor is presently the Deputy Director for the Earth Sciences Directorate of the Goddard Space Flight Center, NASA. She serves as the focal point for directorate activities involving Code 200 (Management Operations Directorate), Code 700 (Engineering Directorate), and Code 300 (Office of Flight Assurance). She is a member of the Engineering and Observing Systems Advisory Group which assesses the engineering capability of the Directorate and the Data System Advisory Committee which assesses the computational needs of the directorate. Both groups make recommendations to the Director of Earth Sciences.

Measurement of Polar Ice Changes with Satellite Altimetry

Dr. Jay Zwally

NASA Goddard Space Flight Center
Laboratory for Hydrospheric Processes, Oceans and Ice Branch

In today's climate, the Greenland and Antarctic ice sheets receive about 8 mm/year of the water that evaporates from the entire surface of the ocean. However, we don't know whether the snowfall on the ice sheets balances the amount of ice that melts, evaporates, and flows into the ocean in rivers and icebergs. Also, global sea level has been rising about 2 mm/yr, but the source of water for at least half of this rise is unknown. If climate warms, more ice would melt, but the ice sheets may also grow and take water out of the ocean as a warmer atmosphere in polar regions carries more moisture for snowfall.

The EOS ALT mission in July 2002 will acquire unique data for determining the mass balance of the polar ice sheets and for predicting future changes in global sea level. The Geoscience Laser Altimeter System (GLAS) on EOS ALT will measure changes in ice volume, seasonal and interannual changes in surface elevation caused by changes in precipitation and melting, and cloud heights and aerosol distributions in the atmosphere.

Dr. Jay Zwally is Deputy Project Scientist for the EOS ALT Mission and is on the GLAS Science Team. He received his Ph.D. in Physics from the University of Maryland and was Program Manager for Glaciology and Remote Sensing at the National Science Foundation before coming to Goddard in 1974.

COMPANIES AND INSTITUTIONS ATTENDING

GODDARD SPACE FLIGHT CENTER

Goddard Institute for Space Studies

Code 480	Code 738.3	Code 910.4	Code 930.1	Code 940
Code 504	Code 900	Code 914	Code 930.5	Code 943
Code 555.1	Code 902.2	Code 916	Code 932	Code 971
Code 563.3	Code 902.3	Code 921.0	Code 933	Code 974
Code 704	Code 910	Code 930	Code 934	Code 975

OTHER GOVERNMENT AGENCIES

Central Intelligence Agency
Naval Research Laboratory
NOAA/NESDIS/OSD
U.S. Army

UNIVERSITIES

University of Maryland

PRIVATE SECTOR

ADSYSTECH Inc.
Booz, Allen and Hamilton
CAELUM Research Corporation
Computer Science Corporation
COMSAT Labs
COMSO Inc.
Computer Technology Associates Inc.
DNA Springfield Research Facility
Fairchild Space Industry
General Sciences Corporation
Hughes STX
Kensington Systems Inc.
KT2 Inc.
MITRE Corporation
National Research Council
OMITRON
Planning Research Corporation Inc.
Research Data Corporation
Research and Data Systems Corporation
SAIC
Smart Inc.
Space Applications Corporation
Universal Systems and Technology Inc.
Universities Space Research Association

APPENDIX B

Data Compression Conference (DCC '94)

March 29 - 31, 1994

Snowbird, UT

Data Compression Conference (DCC'94)

(Sponsored by the IEEE Computer Society TCCC
in Cooperation with NASA/CESDIS)

Snowbird, Utah
March 29 - 31, 1994

GENERAL CHAIR: J. Storer, Brandeis U.
PROGRAM CHAIR: M. Cohn, Brandeis U.
PUBLICITY CHAIR: R. Miller, NASA

PROGRAM COMMITTEE:
J. Abrahams, *ONR*
A. Apostolico, *Purdue/Padova*
R. Arps, *IBM*
R. Baker, *PictureTel Inc.*
A. Bookstein, *U. Chicago*
M. Cohn, *Brandeis U.*
R. Gray, *Stanford U.*
D. Hirschberg, *UC Irvine*
G. Langdon, *UC Santa Cruz*
A. Lempel, *Technion*
B. Lucier, *Purdue U.*
J. Reif, *Duke U.*
D. Renner, *TRW*
E. Riskin, *U. Washington*
K. Rose, *UCSB*
D. Sheinwald, *IBM*
J. Storer, *Brandeis U.*
J. Tilton, *NASA*
M. Vetterli, *UC Berkeley*
J. Vitter, *Duke U.*
I. Witten, *U. Waikato*
X. Wu, *U. W. Ontario*
J. Ziv, *Technion*

SCHEDULE:

Monday Evening: Wine and Cheese Reception and Registration

Tuesday
Morning: Technical Sessions
Mid-Day: Invited Presentation
Afternoon: Technical Sessions

Wednesday
Morning: Technical Sessions
Mid-Day: Invited Presentation
Afternoon: Poster Session and Reception

Thursday
Morning: Technical Sessions
Mid-Day: Invited Presentation
Afternoon: Technical Sessions

Friday: Industry Workshop

Saturday: NASA Workshop

Advance Program

MONDAY, MARCH 28: Registration / Reception, 7:00 - 10:00pm, In the Golden Cliff Room

TUESDAY, MARCH 29

Welcome: 7:45am

Session 1: 8:00 - 10:05

8:00
"Variable Dimension Weighted Universal Vector Quantization and Noiseless Coding"
M. Effros, P.A. Chou, R.M. Gray

8:25
"Entropy-Constrained Tree-Structured Vector Quantizer Design by the Minimum Cross Entropy Principle"
K. Rose, D. Miller, A. Gersho

8:50
"A Subjective Distortion Measure for Vector Quantization"
X. Wu, K. Zhang

9:15
"Multidimensional Rotations for Quantization"
A. Hung, T. Meng

9:40
"A New Multiple Path Search Technique for Residual Vector Quantizers"
C.F. Barnes

Break: 10:05 - 10:30

Session 2: 10:30 - 12:35

10:30
"The Minimax Redundancy is a Lower Bound for Most Sources"
N. Merhav, M. Feder

10:55
"Differential State Quantization of High Order Gauss Markov Process"
A. Bist

11:20
"Data Compression Techniques for Stock Market Prediction"
S. Azhar, A. Glodjo, M. Kao

11:45
"Huffman-Type Codes for Infinite Source Distributions"
J. Abrahams

12:10
"On the Redundancy of Optimum Fixed-to-Variable Length Codes"
and
"Adaptive Variable-to-Variable Length Codes"
P. Stubley

Lunch Break: 12:35 - 2:00

Mid-Day Invited Presentation 2:00 - 3:30

"Data Compression Patents"
Wayne Barsky
Partner in the Law Firm of Irell and Manela

Session 3: 4:00 - 5:40

4:00
"Compression-Based Template Matching"
S. Inglis, I. Witten

4:25
"Markov Models for Clusters in Concordance Compression"
A. Bookstein, S. Klein, T. Raita

4:50
"Static Compression for Dynamic Texts"
A. Moffat, N. Sharman, J. Zobel

5:15
"Parsing Algorithms for Dictionary Compression on the PRAM"
D. Hirschberg, L. Stauffer

Break: 5:40 - 6:05

Session 4: 6:05 - 7:20

6:05
"Syntax-Constrained Encoder Optimization
Using Adaptive Quantization Thresholding for JPEG-MPEG Coders"
K. Ramchandran, M. Vetterli

6:30
"Customized JPEG Compression for Grayscale Printing"
R. Vander Kam, P. Wong

6:55
"Lossless Image Compression with Lossy Image
Using Adaptive Prediction and Arithmetic Coding"
S. Takamura, M. Takagi

WEDNESDAY, MARCH 30, 1994

Session 5: 8:00 - 10:05

8:00
"Explicit Bit Minimization for Motion-Compensated Video Coding"
D. Hoang, P. Long, J. Vitter

8:25
"Online Compression of Video Sequences Using Adaptive VQ Codebooks"
X. Wang, S.M. Shende, K. Sayood

8:50
"Compression of HDTV Signals for Low Bit-Rate Transmission
Using Motion Compensated Subband Transform Coding and a Self-Organization Neural Network"
R. Bhaskaran, S.C. Kwatra

9:15
"Differential Vector Quantization of Real-Time Video"
J. Fowler, S. Ahalt

9:40
"The MVP: A Highly-Integrated Video Compression Chip"
R. Gove

Break: 10:05 - 10:30

Session 6: 10:30 - 12:35

10:30
"A Hybrid Approach to Text Compression"
P. Gutmann, T. Bell

10:55
"Highly Efficient Universal Coding with Classifying to Subdictionaries for Text Compression"
Y. Nakano, H. Yahagi, Y. Okada, S. Yoshida

11:20
"Compression By Induction of Hierarchical Grammars"
C. Nevill-Manning, I. Witten, D. Maulsby

11:45
"Architectural Advances in the VLSI Implementation of Arithmetic Coding
for Binary Image Compression"
G. Feygin, P. Gulak, P. Chow

12:10
"Multiplication and Division Free Adaptive Arithmetic Coding Techniques for Bi-Level Images"
L. Huynh

Lunch Break: 12:35 - 2:00

Mid-Day Invited Presentation: 2:00 - 3:30

"Wavelets"
Prof. Bradley Lucier
Purdue University

POSTER SESSION AND RECEPTION: 4:00pm - 7:00pm, In the Golden Cliff Room

(Abstracts of each presentation appear in the conference proceedings.)

THURSDAY, MARCH 31, 1994:

Session 7: 8:00 - 10:05

8:00
"Bayes Risk Weighted Tree-Structured Vector Quantization
with Posterior Estimation"
K. Perlmutter, R.M. Gray, K.L. Oehler, R.A. Olshen

8:25
"Fast Bintree-Structured Image Coder for High Subjective Quality"
X. Wu, Y. Fang

8:50
"A High Performance Fixed Rate Compression Scheme
for Still Image Transmission"
M. Ruf

9:15
"A Nonlinear VQ-Based Predictive Lossless Image Coder"
M. Slyz, D. Neuhoff

9:40
"Band Ordering in Lossless Compression of Multispectral Images"
S. Tate

Break: 10:05 - 10:30

Session 8: 10:30 - 12:35

10:30
"Enhancement of Block Transform Coded Images
Using Residual Spectra Adaptive Postfiltering"
I. Linares, R. Mersereau, M. Smith

10:55
"Self-Similarity of the Multiresolutional Image/Video Decomposition:
Smart Expansion as Compression of Still and Moving Pictures"
O. Kiselyov, P. Fisher

11:20
"An Investigation of Wavelet-Based Image Coding
Using an Entropy-Constrained Quantization Framework"
M. Orchard, K. Ramchandran

11:45
"Filter Evaluation and Selection in Wavelet Image Compression"
J. Villasenor, B. Belzer, J. Liao

12:10
"Visibility of DCT Basis Functions: Effects of Contrast Masking"
J. Solomon, A. Watson, A. Ahumada
and
"Visibility of DCT Basis Functions: Effects of Display Resolution"
A. Watson, J. Solomon, A. Ahumada

Lunch Break: 12:35 - 4:00

Mid-Day Invited Presentation 2:00 - 3:30

"Audio Compression"
Prof. Allen Gersho
University California, Santa Barbara

Session 9: 4:00 - 5:40

4:00
"On Lattice Quantization Noise"
R. Zamir, M. Feder

4:25
"Vector Quantization of Contextual Information
for Lossless Image Compression"
X. Ginesta, S. Kim

4:50
"Bayes Risk Weighted VQ and Learning VQ"
R. Wesel, R.M. Gray

5:15
"Improved Techniques for Single-Pass Adaptive VQ"
C. Constantinescu and J. Storer

Break: 5:40 - 6:05

Session 10: 6:05 - 7:20

6:05
"Variable Dimension Vector Quantization of Speech Spectra
for Low Rate Vocoders"
A. Das, A.V. Rao, A. Gersho

6:30
"A Rapidly Adaptive Lossless Compression Algorithm
for High Fidelity Audio Coding"
T. Shamoon, C. Heegard

6:55
"Sharper Bounds on Occam Filters with Application to Digital Video"
B.K. Natarajan

FRIDAY, APRIL 1: Industry Workshop

SATURDAY, APRIL 2: NASA Workshop

Industry Workshop
"Image Compression - Applications and Innovations"
Friday April 1, 1994, Snowbird, Utah

(Held in conjunction with the IEEE Data Compression Conference March 29 - 31, 1994, Snowbird, Utah)

THEME: This workshop will focus on current industrial applications of data compression technology. Emerging compression-related technologies and products will shape the future of multi-media industries. Both hardware and software techniques and solutions will be discussed. Participants will include developers, designers, and users. The focus will be product related, with emphasis on innovative system solutions based on existing technologies, including implementation, tradeoffs, results, and future directions.

AGENDA:

Introduction/Welcome	8:00 - 8:15	*Coffee and refreshments*
Session One	8:15 - 10:15	**Multispectral Compression Techniques* *(4 Papers TBA)*

**The Multispectral Compression Session is co-sponsored by Defense LANDSAT Program Office (DLPO). The DLPO will share its vision for future LANDSAT BWC techniques.*

Session Two	10:45 - 12:15	*Multimedia Compression Procedures* *(3 Papers TBA)*
Session Three	2:00 - 3:30	*Varied Industry Applications* *(3 Papers TBA)*
Demonstrations	4:00 - 7:00	*Demonstration and Poster Session (located adjacent to the conference room) includes a hosted reception and Poster Session*

WORKSHOP COMMITTEE CHAIR:

Robert L. Renner
TRW R7/2014
One Space park
Redondo Beach, CA 90278
email: renner@spf.trw.com

PUBLICITY AND REGISTRATION:

Georgia Flanagan
CESDIS Code 930,5
NASA GSFC
Greenbelt, MD 20771
email: georgia@cesdis1.gsfc.nasa.gov

REGISTRATION and FEE: There is a $50 registration fee which includes refreshments, reception, and a copy of the Industry Workshop proceedings.

Name: _____ phone: _____

Affiliation: _____ email: _____

Address: _____

_____

Mail the registration form and fee to Georgia Flanagan, CESDIS, Code 930.5, NASA GSFC, Greenbelt, MD 20771. Payment is non-refundable and must be a check made out to "Industry Workshop" in U. S. dollars and drawn on a U. S. bank. Charge cards or purchase orders cannot be accepted.

ADVANCE PROGRAM: Space and Earth Science Data Compression Workshop

University of Utah, Salt Lake City, Utah, Saturday, April 2, 1994
(Sponsored by NASA/CESDIS, in Cooperation with DCC '94)

Registration and Continental Breakfast: 7:30-8:15am

Session 1: 8:15-10:15am

8:15 *An Image Assessment Study of Image Acceptability of the Galileo Low Gain Antenna Mission*
S. L. Chuang and T. Grant/Ames Research Center, R. F. Haines and Yaron Gold/Recom Technologies, and Kar Ming Cheung/Jet Propulsion Laboratory

8:45 *An Adaptive Vector Quantization and Rice Encoder Hybrid Algorithm*
C. Rex Reed and Scott E. Budge/Utah State University

9:15 *Image Compression Software for the SOHO LASCO and EIT Experiments*
Mitchell R. Grunes/Allied-Signal Technical Services, Russell A. Howard, Karl Hoppel and Stephen A. Mango/Naval Research Laboratory, and Dennis Wang/Interferometrics

9:45 *Comparison of the Lossy Image Data Compressions for the MESUR Pathfinder and for the Huygens Titan Probe*
P. Rüffer, F. Rabe, F. Gliem and H.-U. Keller/Technische Universität Braunschweig, Germany

10:15 Break

Session 2: 10:45am-12:15pm

10:45 *Advanced End-to-End Simulation for On-Board Processing (AESOP)*
Alan S. Mazer/Jet Propulsion Laboratory

11:15 *Performance Considerations for the Application of the Lossless Browse and Residual Model*
Walter D. Abbott III, Robert T. Kay and Ron J. Pieper/Naval Postgraduate School, Monterey, CA

11:45 *Image Quality Measures and Their Performance*
Ahmet M. Eskicioglu, Paul S. Fisher and S. Chen/University of North Texas

Buffet Lunch, Salt Air Room: 12:15-1:30pm

Session 3: 1:30-3:30pm

1:30 *Some Practical Aspects of Lossless and Nearly-Lossless Compression of AVHRR Imagery*
David Hogan/Atmospheric and Environmental Research, Inc., and Chris Miller, Than Lee Christensen and Raj Moorti/Martin Marietta Astro-Space, Princeton, NJ

2:00 *Vector Quantizer Designs for Joint Compression and Terrain Categorization of Multispectral Imagery*
John D. Gorman/Environmental Research Institute of Michigan, and Daniel F. Lyons/The Mitre Corporation

2:30 *Compression of Multispectral Landsat Imagery Using the Embedded Zerotree Wavelet (EZW) Algorithm*
Jerome M. Shapiro, Stephen A. Martucci and Martin Czigler/The David Sarnoff Research Center, Princeton, NJ

3:00 *Wavelet Compression Techniques for Hyperspectral Data*
Bruce Evans, Brian Ringer and Mathew Yeates/TRW Systems Integration Group, Redondo Beach, CA

3:30 Break

Session 4: 4:00-5:30pm

4:00 *A Comparison of Spectral Decorrelation Techniques and Performance Evaluation Metrics for a Wavelet-Based, Multispectral Data Compression Algorithm*
Roy M. Matic/Hughes Research Laboratories, Malibu, CA, and Judy I. Mosley/Santa Barbara Research Center, Goleta, CA

4:30 *A Comparative Study of SAR Data Compression Schemes*
C. Lambert-Nebout/CNES, O. Besson/ENSICA, and D. Massonnet/CNES - Toulouse, France

5:00 *Perceptual Compression of Magnitude-Detected Synthetic Aperture Radar Imagery*
John D. Gorman and Susan A. Werness/Environmental Research Institute of Michigan

5:30 Close

✂----------

REGISTRATION: Space and Earth Science Data Compression Workshop, April 2, 1994. Early registration fee (postmark by March 1, 1994) is $30.00. Late registration fee (after March 1, 1994) is $45.00. NOTE: You must register separately for DCC '94 and this Workshop. Registration includes proceedings, breakfast and buffet lunch.

Name: _____
Affiliation: _____
Address: _____
_____
Phone: _____ E-Mail: _____

Enclose the registration fee and mail to Ms. Georgia Flanagan, Mail Code 930.5, NASA GSFC, Greenbelt, MD 20771. Payment is non-refundable and must be a check made out to CESDIS *in U. S. dollars and drawn on a U. S. bank.* Charge cards or purchase orders cannot be accepted.

COMPANIES AND INSTITUTIONS ATTENDING

NASA HEADQUARTERS AND RESEARCH CENTERS

NASA Ames Research Center
NASA Langley Research Center
NASA Lewis Research Center

OTHER GOVERNMENT AGENCIES

Department of Defense
Department of the Navy
National Security Agency
Office of Naval Research

UNIVERSITIES

Arkansas State University
Brandeis University
Brigham Young University
California State University
CESDIS
City Polytechnic of Hong Kong
Colorado State University
Columbia University
Cornell University
Duke University
Dynetics Inc.
Eastern Connecticut State University
Florida Institute of Technology
Georgia Institute of Technology
King's College University of London
Kyushu Institute of Technology
Loughborouch University of Technology
Mississippi State University
National Chung Cheng University
NCSA
NEC Technologies Inc.
New Mexico Inst. of Mining & Technology
New Technology Inc.
Nichold Research Corporation
Northeastern University
Ohio State University
Polytechnic University of New York
Princeton University
Purdue University
Royal Melbourne Institute of Technology
Rutgers University
South Dakota State University
Stanford University
Stevens Institute of Technology
Syracuse University
Technical University of Braunschweig
Technical University of Denmark
Tel-Aviv University
Trinity College of Dublin
University of North Texas
University of Arizona
University of California
University of Cambridge
University of Canterbury
University of Chicago
University of Genoa
University of Hawaii
University of Illinois
University of Koblenz
University of Melbourne
University of Michigan
University of Nebraska
University of North Texas
University of Pennsylvania
University of Pittsburgh
University of Siegen
University of South Carolina
University of Southern California
University of Texas
University of Tokyo
University of Toledo
University of Toronto
University of Turku
University of Utah
University of Waikato
University of Washington
University of Waterloo
University of Western Ontario

UNIVERSITIES (continued)

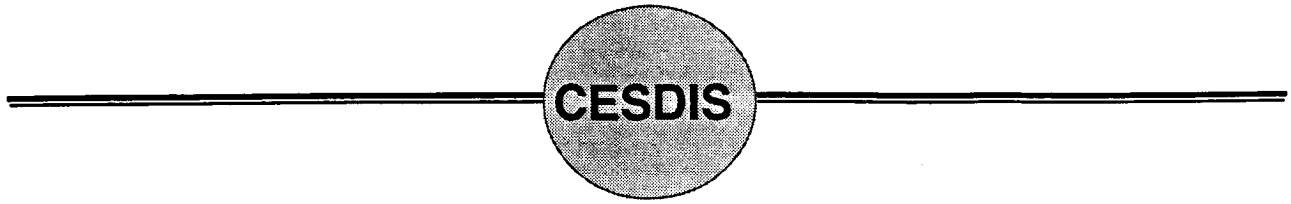
University of Wisconsin
Utah State University
Virginia Tech
York University

PRIVATE SECTOR

A. S. L. M.
Academia Sinica
ADTRAN
Advanced Hardware Architectures
Aerospace Corporation
ALADDIN Systems Inc.
Allied-Signal Technical Services
Argonne National Laboratory
AT&T Bell Laboratories
AT&T/GIS
Aware Inc.
Bell-Northern Research
Caelum Research Corporation
Chevron Petroleum Technology
CNES
Comsat Labs
David Sarnoff Research Center
Delta Information Systems
Digital Biometrics Inc.
DLR
Eastman Kodak Company
Environment Canada
ERIM
ETRI
Evans & Sutherland
Fujitsu Laboratories Ltd.
Hewlett-Packard Labs
Hughes Aircraft Company
Hughes STX
IBM Almaden Research Center
IBM Austin
IBM Corporation
IBM ISRAEL Scientific Center
IBM T. J. Watson Research Center
IEEE
Intel Corporation
Irell & Manella
Itek Optical Systems
IITRI
Jet Propulsion Laboratory
KCSL
Loral AeroSys
Los Alamos National Lab
MacDonald Detiwiler & Associates
MAGNALINK Communications
Martin Marietta Corporation
Microsoft Corporation
MIT/Motorola
MITRE Corporation
Motorola
Motorola Codex
National Science Foundation
Para Graph International
PictureTel Corporation
PRC Inc.
Research Triangle Institute
RICOH California Research Center
Rockwell International
Storage Technology Corporation
Technion
Telco Systems Inc.
Telecon Paris
Texas Instruments
TRW Defense Systems Division
TRW Systems Integration Group
Tubitak Ankara
UNISYS
VLSI Design Lab
Wang Laboratories
West Publishing Company
Xerox Corporation

APPENDIX C

Technical Report Series



CENTER OF EXCELLENCE IN SPACE DATA AND INFORMATION SCIENCES
CODE 930.5
NASA GODDARD SPACE FLIGHT CENTER
GREENBELT, MD 20771

TECHNICAL REPORT SERIES ABSTRACTS BY AUTHOR

301-286-4403 Internet: cas@cesdis1.gsfc.nasa.gov

James Anderson, University of Maryland

TR-92-91 Efficient Synchronization with Minimal Hardware Support — James H. Anderson — November 1992

We present an efficient algorithm for the implementation of spin locks in shared memory multiprocessors. Our algorithm does not require any hardware support other than atomic read and write operations and does not employ global spinning. When implementing combinable read-modify-write operations using our spin lock algorithm, the technique of software combining can be incorporated in order to enhance scalability. We present performance studies that show that our spin lock algorithm is comparable to queue-based locks that employ strong primitives such as compare-and-swap or fetch-and-add, and that our software combining algorithm is faster than other algorithms that employ hardware locking primitives. Our results suggest that synchronization can be accomplished with minimal hardware support on large-scale shared memory multiprocessors.

TR-92-92 A Fine-Grained Solution to the Mutual Exclusion Problem — James H. Anderson — August 1992

We present a "fine-grained" solution to the mutual exclusion problem. A program is *fine-grained* if it uses only single-reader, single-writer boolean variables and if each of its atomic operations has at most one occurrence of at most one shared variable. In contrast to other fine-grained solutions that have appeared in the literature, processes in our solution do not busy-wait, but wait on one another only by executing await statements. Such statements can be implemented in practice either by means of context switching or by means of "local" spinning. We show that our algorithm is correct even if shared variables are accessed nonatomically.

TR-92-93 On the Granularity of Conditional Operations — James H. Anderson — May 1992

We examine the "granularity" of statements of the form await *B Æ S*, where *B* is a boolean expression over program variables and *S* is a multiple-assignment. We consider two classes of such statements to have the same granularity if any statement of one class can be implemented without busy-waiting by using statements of the other class. Two key results are presented. First, we show that statements of the form await *B Æ S* can be implemented without busy-waiting by using simpler statements of the form await *X*, *X* := *y*, and *y* := *X*, where

y is a private boolean variable and *X* is a shared single-reader, multi-writer boolean variable. Second, we show that if busy-waiting is not allowed, then there is no general mechanism for implementing statements of the form await *B*, where *B* is an *N*-writer expression, using only assignment statements and statements of the form await *C*, where *C* is an (*N*-1)-writer expression. It follows from these results that the granularity of waiting depends primarily on the number of processes that may write each program variable. These results also show that, from a computational standpoint, operations that combine both waiting and assignment, such as the *P* semaphore primitive, are not fundamental.

CESDIS

TR-92-76 CESDIS Annual Report, Year 3
July 1990-June 1991

TR-92-94 CESDIS Annual Report, Year 4
July 1991 - June 1992

TR-94-110 CESDIS Annual Report, Year 5
July 1992 - June 1993

Alok Choudhary, Syracuse University

TR-94-119 Compiler and Runtime Support for Out-of-Core HPF Programs — Alok Choudhary — April 1994

This paper describes the design of a compiler which can translate out-of-core programs written in a data parallel language like HPF. Such a compiler is required for large scale scientific applications such as the *Grand Challenge* applications which deal with enormous quantities of data. We propose a framework by which a compiler together with appropriate runtime support can translate an out-of-core- HPF program to a message passing node program with explicit parallel I/O. We describe the basic model of the compiler and the various transformations made by the compiler. We also discuss the runtime routines used by the compiler for I/O and communication. In order to minimize I/O, the runtime support system can reuse data already fetched into memory. The working of the compiler is illustrated using two applications, namely a Laplace equation solver an LU Decomposition, together with performance results on the Intel Touchstone Delta.

James Coggins, University of North Carolina at Chapel Hill

TR-90-07 Designing C++ Libraries — James M. Coggins — 1990

Class libraries encapsulate useful implementation ideas. Designing libraries is difficult because of competing, conflicting objectives. Design criteria are needed to guide library architects toward good encapsulations. This paper argues that three strategies used in many object-oriented libraries are actually poor library design criteria, particularly in C++. A new criterion is defined and examples of its use are illustrated. The new criterion applies the common software engineering design principle "separation of concerns" in a specific way that leads to effective library designs. The approach leads to a model at collaborative development of class libraries customized for a family of applications.

TR-90-13 Interfacing Image Processing and Computer Graphics Systems Using an Artificial Visual System — James M. Coggins — 1986

An Artificial Visual System (AVS) has been developed to simplify three -dimensional microscope images for presentation and manipulation in an interactive computer graphics system. The AVS consists of several sets of spatial filters that decompose an image along three different measurement continua. A recombination algorithm processes the filter outputs to detect objects, to eliminate noise, and to map the detected objects into points in a multidimensional feature space. Recent discoveries regarding the geometry of the points in the feature space are described. One recent result simplifies the AVS by decreasing the number of filters required to obtain the same measurements. Not only are accurate measurements possible, but certain image distortions can be modeled and counteracted in the feature space.

TR-90-18 Anticipated Methodologies in Computer Vision — James M. Coggins — 1990

Over thirty years of computer vision research has yielded a multitude of diverse techniques for image processing and analysis, but no unified, scientific approach for comparing, evaluating, or applying those techniques. There remain many machine vision applications that should have simple solutions but that remain unsolved while the field devotes attention, effort, and dollars to engineering specific solutions to high-profile projects with little generalizability to other vision tasks.

Our research has approached the field with the specific objective of building a unified, scientific foundation for machine perception [Coggins 89]. This paper and the accompanying presentation will describe several machine vision applications that are driving our work and will present some of the insights that may herald a breakthrough toward a scientific discipline of machine perception.

Section 2 presents our approach to computer vision research by describing a series of applications ("driving problems") from which we induce principles to guide the unification of computer vision. Section 3 will summarize the key insights that are leading toward a scientific foundation for computer vision.

TR-90-22 Multiscale, Geometric Image Descriptions for Interactive Object Definition — James M. Coggins — 1989

A means is described of analyzing two- and three-dimensional images into a directed a cyclic graph of visually sensible, coherent regions and of using this DAG as the basis for interactive object definition. The image analysis is in terms of the geometry of the intensity surface via a multiscale approach with a focus on symmetry properties about ridges. The image analysis method, a system for interactive object definition, and results of their use on two-dimensional images are reported.

TR-90-26 Multiscale Vector Fields for Image Pattern Recognition — James M. Coggins — 1989

We propose a uniform processing framework for low-level vision computing in which a bank of spatial filters maps the image intensity structure at each pixel into an abstract feature space. Some properties of the filters and the feature space will be described. Local orientation is measured by a vector sum in the feature space as follows: each filter's preferred orientation along with the strength of the filter's output determine the orientation and the length of a vector in the feature space; the vectors for all filters are summed to yield a resultant vector for a particular pixel and scale. The orientation of the resultant vector indicates the local orientation, and the magnitude of the vector indicates the strength of the local orientation preference. Limitations of the vector sum method will be discussed. Our investigations show that the processing framework provides a useful, redundant representation of image structure across orientation and scale.

TR-90-30 Interactive Object Definition in Medical Images Using Multiscale, Geometric Image Descriptions — James M. Coggins — 1990

A promising approach to medical image object definition involves automatic computation of a region-based image description along with a region containment directed a cyclic graph (RCDAG) induced from the description via multiscale analysis of image structures [Pizer 1989]. The information resulting from this computation provides the basis for interactive object definition. During object definition, the human user inserts semantics into the image description through additions to and alteration of the automatically computed RCDAG. This paper describes the object definition method and a tool for interactive object definition. Design criteria and resulting design decisions for this tool are presented, followed by a discussion of preliminary image segmentation and object definition results.

TR-90-33 A Multiscale Description of Image Structure for Segmentation of Biomedical Images — James M. Coggins — 1990

A new representation of image intensity structure, the Multiscale Orientation Field, is described and its application to biomedical image segmentation is discussed. The MOF is composed of orientation vectors at every pixel and at multiple scales. The vectors are computed from the outputs of an Artificial Visual System composed of spatial filters whose design is described. A new abstract feature space for orientation measurement is defined and a filter sensitivity function that yields a desirable mapping into that feature space is described.

TR-90-35 Biomedical Image Segmentation Using Multiscale Orientation Fields — James M. Coggins — 1990

We present an algorithm for labeling image regions based on pixel-level statistical pattern recognition. The structure of multiscale regions about each pixel is measured using isotropic Gaussian filters and by a Multiscale Orientation Field. We create a redundant feature space representing several aspects of image structure across scale, orientation, and space. Our segmentation algorithm decides membership of pixels in regions using simple statistical pattern recognition methods such as distance measurement and thresholding. Feature vectors are examined locally to determine region membership; the features incorporate multiscale image structure information. Results of multiscale image segmentations on biomedical images are presented.

TR-90-37 Image Structure Analysis Supporting Interactive Object Definition — James M. Coggins — 1990

Multiscale geometric image structure analysis is used to produce a hierarchical labeling of image regions. The regions provide a language for fast, interactive object definition. The approach allows human analysts to quickly inject semantics into the image representation, enhancing rather than trying to replace the human operator's capabilities.

TR-91-49 Supervised Pixel Classification Using a Feature Space Derived from an Artificial Visual System — James M. Coggins

Image segmentation involves labeling pixels according to their membership in image regions. This requires that we understand what a *region* is. Using supervised pixel classification, we investigate how groups of pixels labeled manually according to perceived image semantics map onto the feature space created by an Artificial Visual System. We investigate multiscale structure of regions and show that pixels from clusters based on their geometric roles in the image intensity function, not by image semantics. A tentative abstract definition of a "region" is proposed based on this behavior.

David DeWitt, University of Wisconsin

TR-94-113 Paradise - A Parallel Geographic Information System — David De Witt — January 1994

The goal of the Paradise project is to design and implement a scalable, parallel geographic information system that is capable of storing and manipulating massive data sets. By applying object-oriented and parallel database technologies to the problem of storing and manipulating geographic information we hope to significantly advance the size and complexity of GIS data sets that can be successfully stored, browsed, and queried.

TR-94-117 Client-Server Paradise — David De Witt — March 1994

This paper describes the client-server version of Paradise, a new GIS under development at the University of Wisconsin. Paradise is being implemented as a SHORE value-added server. It provides the user an extended-relational data model with support for point, raster, polygon, and polyline ADTs/ An extended version of SQL is provided for formulating ad-hoc queries and a graphical user interface based on the Tk toolkit allows the user to query and browse graphically.

The target application for Paradise is NASA's EOSDIS project as well as other projects involving massive amounts of raster, satellite images, Paradise incorporates several performance optimizations including the transparent separation of raster images from their associated meta-data, division of raster images into chunks to minimize unnecessary I/O, and the automatic application of lossless compression /decompression on a chunk-by-chunk basis.

TR-94-121 Client-Server Paradise — David De Witt — June 1994

This paper describes the design and implementation of Paradise, a database system designed for handling GIS type of applications. The current version of Paradise, uses a client-server architecture and provides an extended-relational data model for modeling GIS applications. Paradise supports an extended version of SQL and provides a graphical user interface for querying and browsing the database. We also describe the results of benchmarking Paradise using the Sequoia 2000 storage benchmark.

Tarek El-Ghazawi, George Washington University

TR-94-111 I/O Performance of the MasPar MP-1 Testbed — Tarek A. El-Ghazawi — January 1994

Input/output speed continues to present a performance bottleneck for high performance computing systems as technology improves processor power, memory capacity, and disk capacity at a much higher-rate.

Developments in I/O architecture have been attempting to reduce this performance gap. The MasPar I/O architecture includes many interesting features. This work presents an experimental study of the dynamic characteristics of the MasPar parallel I/O. The results have revealed many strengths as well as areas for potential improvements, and are helpful to software developers in tuning the I/O activities of applications to the MasPar I/O Architecture.

James Foley, George Washington University

TR-91-65 Scientific Data Visualization Software: Trends and Directions — James D. Foley — 1990

Scientific data visualization has finally come of age as an important and accepted discipline. While scientists have been using computer graphics to visualize experimental data and computational results for at least 30 years, recent improvements in cost/performance of graphics workstations, more readily available software, and a new found identity have solidified the discipline. Many thousands of scientists regularly use visualizations as part of their work. My thesis is that scientists are forced to work too hard to create these visualizations, but that the evolving set of visualization tools can greatly reduce the requisite effort.

Eric Feigelson, Pennsylvania State University

TR-91-57 Study of Six Linear Least Squares Fits — Eric D. Feigelson

For many applications in physical sciences like astronomy, only a set of correlated data points (x_i, y_i) is available to fit a line. The underlying joint distribution is unknown, and it is not clear which variable is 'dependent' and which is 'independent'. In such cases, the choice of least-squares line is ambiguous. Astronomers have used a many as six different linear regression methods for this situation: the two ordinary least-squares (OLS) lines, orthogonal regression, the OLS bisector, the reduced major axis and the OLS-mean. The latter four methods treat the X and Y variables symmetrically. Relations between the six regression slopes are obtained. Estimates of their variances, derived using the delta method, show that the OLS bisector has the least variance among the symmetrical methods. Simulation studies confirm the accuracy of estimators for moderate and large samples.

Noah Friedland, University of Maryland

TR-92-83 A Markov Field/Accumulator Sampler Approach to the Atmospheric Temperature Inversion Problem — Noah Friedland

A new optimization method is proposed for the solution of the Atmospheric Temperature inversion problem. This problem involves obtaining an atmospheric temperature profile from infra-red radiances measured at the top of the atmosphere by satellites. This temperature profile is modeled as a 1-D Markov Random Field (MRF) through the definition of an energy function. This energy function incorporates observed infra-red radiance values along with some a priori assumptions regarding the desired profile's structure. The MRF is then driven to its most probable configuration, or equivalently its minimal energy value, using Simulated Annealing (SA). Three approaches are examined; SA with Random Uniform configuration sampling, SA with an approximated Gibbs Sampler and SA with an Accumulator Sampler (AS). The latter is a new method which overcomes the inefficiency of random sampling, without the computational overhead needed in calculating Gibbs distributions. These three SA sampling methods are implemented to recover 334 HIRS2 temperature profiles from synthetically generated infra-red radiances. Experimental results prove the Accumulator sampler's potential.

Hillel Gazit, Duke University

TR-91-56 A Randomized EREW Parallel Algorithm for Finding Connected Components in a Graph — Hillel Gazit, John Reif

We present a parallel *EREW* randomized algorithm for finding the connected components of an undirected graph. Our algorithm takes $T=O(\log(n)+\log(g))$ time an $P=O\left(\frac{n+m}{\log n}\right)$ p rocessor, where m = number of edges, n = number of vertices and g is the genus of the graph.

TR-91-66 Planar Separators and the Euclidean Norm — Hillel Gazit

In this paper we show that every 2-connected embedded planar graph with faces of sizes $d_1, \ldots, d_f$ has a simple cycle separator of size 1.58 $\sqrt{d_1^2+\cdots+d_f^2}$ and we give an almost linear time algorithm for finding these separators, $O(n\,\alpha(n,n))$. We show that the new upper bound expressed as a function of $|G| = \sqrt{d_1^2+\cdots+d_f^2}$ is no larger, up to a constant factor than previous bounds that were expressed in terms of $\sqrt{d \cdot v}$ where d is the maximum face size and v is the number of vertices and is much smaller for many graphs. The algorithms developed are simpler than earlier algorithms in that they work directly with the planar graph and its dual. They need not construct or work with the face-incidence graph as in [Mil86, GM87, GM].

TR-91-67 A Deterministic Parallel Algorithm for Planar Graphs Isomorphism — Hillel Gazit

We present a deterministc parallel algorithm to determine whether two planar graphs are isomorphic. The algorithm needs $O(\log(n))$ separators that have to be computed one after the other. The running time is $T = O(\log(n))$ time for finding separators, and the processors count is $\frac{n^{1.5}\cdot\log(n)}{T}$ —the same complexity as a deterministic single source BHS algorithm for planar graphs [22]. We also show that every planar graph has a separator such that $\sum_{v \in sep} d(v) = O(\sqrt{n})$ and give a parallel algorithm to find that separator.

TR-91-68 A Deterministic Parallel Algorithm for Finding a Separator in Planar Graphs — Hillel Gazit

We present a deterministic parallel algorithm for finding a simple cycle separator in a planar graph. The size of the separator is $O(\sqrt{n})$ and it separates the graph so that the largest part contains at most $\frac{2}{3}\cdot n$ vertices. Our algorithm takes, in the PRAM EREW model, $T=O(\log^3(n))$ time and uses $P = O(n^{1+\epsilon})$ processors where n is the number of vertices, f if the number of faces and ϵ is any positive constant. The algorithm is based on a randomized solution from 1987.

Using a variation of our algorithm we can construct a simple cycle separator of size $O(d \cdot \sqrt{f})$ where d is maximum face size. The running time is the same, but the number of processors we need is $P = O(n+f^{1+\epsilon})$.

TR-91-69 An Algorithm for Finding a $\frac{7}{3}\cdot\sqrt{n}$ Separator in Planar Graphs — Hillel Gazit

In this paper we present an optimal algorithm for finding a $\frac{7}{3}\cdot\sqrt{n}$ separator in a planar graph with n vertices. Similar to Lipton and Tarjan we try to use *BFS* layers to cut the graph into subgraphs with either small number of vertices or small radius. In our algorithm we look for a pair of layers that can form a smaller separator. We prove that our separator will have at most $\frac{7}{3}\sqrt{n}$ vertices, an improvement over previous works by Lipton and Tarjan $(\sqrt{8}\sqrt{n})$ and Djidjev $(\sqrt{6}\sqrt{n})$.

TR-91-70 Optimal EREW Parallel Algorithms for Connectivity Ear Decomposition and st-Numbering of Planar Graphs — Hillel Gazit

Parallel *EREW* deterministic algorithms for finding the connected components, ear decomposition and st-numbering of a planar graph are presented. The algorithms take time with $\frac{n+M}{\log(n)}$ processors. Previous results have the same complexity, but use the *CRCW* model.

The same algorithms can be used for graphs with low genus. Let g be the genus of the *minimal* embedding of the graph, n the number of vertices and m the number of edges. Our algorithm takes $T=O(\log(n) + \log^2(g))$ time and using optimal space and $P = O(\frac{n}{\log(n)})$ processors.

TR-91-71 An Optimal Randomized Parallel Algorithm for Finding Connected Components in a Graph — Hillel Gazit

We present a parallel randomized algorithm for finding the conneted components of an undirected graph. Our algorithm has an expected running time of with $P = O(\frac{m+n}{\log(n)})$ processors, where m is the number of edges and n is the number of vertices. The algorithm is *Optimal* in the sense that the product $P \cdot T$ is a linear function of the input size. The algorithm requires $O(m + n)$ space which is the input size, so it is *Optimal* in space as well.

Theodore Johnson, University of Florida

TR-94-115 2Q: A Low Overhead High Performance Buffer Management Replacement Algorithm — Theodore Johnson — March 1994

In a path-breaking paper last year Pat and Betty O'Neil and Gerhard Weikum proposed a self-tuning improvement to the Least Recently Used (LRU) buffer management algorithm. Their improvement is called LRU/k and advocates giving priority to buffer pages based on the kth most recent access. (The standard LRU algorithm is denoted LRU/1 according to this terminology.) If P1's kth most recent access is more recent than P2's, then P1 will be replaced after P2. Intuitively, LRU/k for k>1 is a good strategy, because it gives low priority to pages that have been scanned or to pages that belong to a big randomly accessed file (e.g., the account file in TPC/A). They found that LRU/2 achieves most of the advantage of their method.

The one problem of LRU/2 is the processor overhead to implement it. In contrast to LRU, each page access requires log N work to a priority queue where N is the number of pages in the buffer.

Question: is there an easier way (constant overhead per access as in LRU) with similar performance to LRU/2?

Answer: Yes.

Our "Two Queue" algorithm (hereafter 2Q) has constant time overhead, performs as well as LRU/2, and requires no tuning. These results hold for real (DB2 commercial) traces as well as simulated ones. Based on these experiments, we estimate that 2Q will provide a 5-10% improvement for most applications and more for some. We give some analytic reasons for this result.

TR-94-116 Sensitivity Analysis of Frequency Counting — Theodore Johnson — March 1994

Many database optimization activities, such as prefetching, data clustering and partitioning, and buffer allocation, depend on the detection of hot spots in access patterns. While a database designer can in some cases use special knowledge about the data and t he users to predict hot spots, in general one must use

information about past activity to predict future activity. However, algorithms that make use of hot spots pay little attention to the way in which hot spot information is gathered, or to the quality of counting. We present a numerical method for estimating the quality of the data, and a rule-of-thumb. We find that if b of the references are made to the hottest a of the N data items, then one should process $N\,a\,[\,(1 - a) / (b - a)^2\,]$ references.

Jacqueline Le Moigne, CESDIS

TR-93-95 Image Analysis by Integration of Disparate Information — Jacqueline Le Moigne — January 1993

Image analysis often starts with some preliminary segmentation which provides a representation of the scene needed for further interpretation. Segmentation can be performed in several ways, which are categorized as pixel-based. Each of these approaches are affected differently by various factors, and the final result may be improved by integrating several or all of these methods, thus taking advantage of their complementary nature.

In this paper, we propose an approach that integrates pixel-based and edge-based results by utilizing an iterative relaxation technique. This approach has been implemented on a massively parallel computer and tested on some remotely sensed imagery from the Landsat-Thematic Mapper (TM) sensor.

TR-93-104 Summary Report of the CESDIS Seminar Series on Earth Remote Sensing — Jacqueline Le Moigne — May 1993

The Earth Remote Sensing Seminar Series, sponsored by CESDIS, was held during the spring of 1993 at the Goddard Space Flight Center. The series consisted of 16 seminars, each of them being presented by an expert in the field. This report summarizes the series and includes copies of the speakers' presentations, as well as an attendee list.

TR-94-108 Summary Report of the CESDIS Seminar Series on Future Earth Remote Sensing Missions — Jacqueline Le Moigne — January 1994

The Earth Remote Sensing Seminar Series, sponsored by CESDIS< and organized in cooperation with the IEEE Geoscience and Remote Sensing Society was held during the fall of 1993 at the Goddard Space Flight Center. The series consisted of eight seminars, each presented by an expert in the field. This report summarizes the series and includes copies of the speakers' presentations, as well as an attendee list.

TR-94-112 Parallel Registration of Multi-Sensor Remotely Sensed Imagery Using Wavelet Coefficients — Jacqueline Le Moigne — January 1994

Due to the increasing amount and diversity of remotely sensed data, image registration is becoming one of the most important issues in remote sensing. In the near-future, remote sensing systems will provide large amounts of data representing multiple-time or simultaneous observations of the same features by different sensors. The combination of data from coarse-resolution viewing satellite sensors designed for large area survey and from finer resolution sensors for more detailed studies will allow better analysis of each type of data as well as validation of global low-resolution data analysis by the use of local high-resolution data analysis. This integration of information from multiple sources starts with the registration of the data. The most common approach to image registration is to choose, in both input image and reference image, some well defined ground control points (GCP's), and then to compute the parameters of a deformation model. The main difficulty lies in the choice of the GCP's. In our work, a parallel implementation of decomposition and reconstruction by wavelet transforms has been developed on a Single Instruction Multiple Data (SIMD) massively parallel computer, the MasPar MP-1. Utilizing this framework, we show how maxima of wavelet

coefficients, which can be used for finding ground control points of similar resolution remotely sensed data[8], can also form the basis of the registration of very different resolution data, such as data from the NOAA Advanced Very High Resolution Radiometer (AVHRR) and from the Landsat/Thematic Mapper (TM).

Tassos Markas, Duke University

TR-91-58 Fast Computations of Vector Quantisation Algorithms — Tassos Markas

In this document we present efficient vector quantization algorithms that can be used to compress images at low bit rates with low computational cost. Evaluated methods include the full-search VQ scheme, the Tree-Structured Vector Quantizer (TSVQ), and the fast computation of the TSVQ method. All these methods use the Mean Squared Error (MSE) distortion measure to identify the best-match between the original image blocks and a set of codebook vectors. We also present the tradeoff between speed and distortion when less complex distortion measures are used to compute the reconstructed image. The evaluated methods include the Mean Error (ME), and the Reduced size MSE (RMSE) distortion measures. The performance of these methods will be evaluated using the Peak Signal-to-Noise Ratio (PSNR) distortion measure.

TR-91-61 Image Compression Methods with Distortion Controlled Capabilities — Tassos Markas, John Reif

Traditionally, lossy compression schemes have focused on compressing data at fixed bit rates in order to communicate information over limited bandwidth communication channels, or to store information in a fixed-size storage media. In this paper we present a class of lossy data compression algorithms that are capable of encoding images so that the loss of information complies with certain distortion requirements. The developed algorithms are based on the Tree-Structured Vector Quantizers (TSVQ). The first distortion controlled algorithm uses variable-size image blocks, encoded on quad-tree data structures, to efficiently encode image areas with different information content. Another class of distortion controlled algorithms that are presented is based on recursive quantization of error image blocks that represent the difference between the current approximation and the original block. We will also describe the progressive compression properties of these algorithms. Finally, we will present their compression/distortion performance using satellite images provided by NASA, and we will show that they achieve better performance than the TSVQ algorithms at high bit rates.

Raymond Miller, University of Maryland

TR-90-01 Analyzing a CSMA/CD Protocol through a Systems of Communicating Machines Specification — Raymond E. Miller — January 1990

A model for the specification and analysis of communication protocols called *Systems of Communication Machines* is used to specify a CSMA/CD protocol, and to analyze it for safety and certain restricted liveness properties. The model uses a combination of finite state machines and variables in the specification of each machine, and the communication between machines is accomplished through shared variables. Enabling predicates and actions are associated with each transition; the enabling predicates determine when a transition may be taken, and the actions alter the variable values as the network progresses.

One of the advantages which this model has over most other formal description techniques is that simultaneous transitions are allowed. In this paper, simultaneous writes to a shared variable are used to model collisions. Another advantage is the use of shared variables rather than FIFO queues for communication between machines. This allows the modeling of the ethernet bus as a single variable shared by all communicating processes.

TR-90-08 Stabilization and Pseudo-Stabilization — Raymond E. Miller — May 1990

A *stabilizing* system is one which if started at any state is guaranteed to reach a state after which the system cannot deviate from its intended specification. In this paper, we propose a new variation of this notion, called pseudo-stabilization. A *pseudo-stabilizing* system is one which if started at any state is guaranteed to reach a state after which the system does not deviate from its intended specification. Thus, the difference between the two notions comes down to the difference between "cannot" and "does not" - a difference that hardly matters in many practical situations. As it happens, a number of well-known systems, for example the alternating-bit protocol, are pseudo-stabilizing but not stabilizing. We conclude that one should not try to make any such system stabilizing, especially if stabilization comes at a high price.

TR-90-14 Protocol Verification: The First Ten Years, The Next Ten Years; Some Personal Observations — Raymond E. Miller — June 1990

A number of formulations for protocol verification that have been developed over the past ten years are introduced and compared. Along with this the logical properties used for verifying that a protocol is correct are defined, and the techniques developed for seeing that these properties hold are discussed.

When questioning how well the current approaches for specification, verification and testing fit together to form a unified theory, a number of unpleasant gaps are found. We discuss the reasons for these gaps and what approaches might be used to alleviate the problems caused by these gaps. Both theoretical and practical questions arise. We discuss only a few of these with the hope of stimulating research into unifying the theoretical approaches to specification, verification and testing.

TR-90-19 Synthesizing a Protocol Converter from Executable Protocol Traces — Raymond E. Miller — June 1990

Communication Finite State Machines (CFSMs) with FIFO queues are used to model a protocol converter. A protocol conversion algorithm is developed and presented for the CFSM model of the protocols A and B. A converter H for protocols $A = (A_0, A_1)$ and $B = (B_0, B_1)$ is viewed as a black box such that H is between sender A_0 and receiver B_1. This gives a resulting protocol $X = (A_0, H, B_1)$. The conversion algorithm requires a specification of the *message relationships* between the messages of protocols A and B. For a class of protocol specifications it is shown that the message relationship *input specification* for the conversion algorithm can be derived from a systematic analysis of the given protocols. Is it assumed that protocols A and B have the *required progress properties.* The algorithm includes a *search* for related messages from the two protocols in a FIFO from a composite space formed by a cartesian cross-product of state space A_1 and B_0. The search produces finite length *traces* which are combined to form a state machine H which is examined for freedom from *unspecified receptions, deadlocks and livelocks. A bisynchronous to alternating bit* protocol conversion example demonstrates the applicability of the algorithm.

TR-90-23 Testing Protocol Implementation Based on a Formal Specification — Raymond E. Miller — July 1990

In this paper a procedure for generating test sequences for a formally specified protocol is given. The procedure is designed for a protocol specified as a *system of communicating machines*, which is a model for the specification and verification of network protocols. The procedure is then illustrated by the generation of a test sequence for a CSMA/CD protocol, previously specified by the model.

TR-90-27 Generating Minimal Length Test Sequences for Conformance Testing of Communication Protocols — Raymond E. Miller — 1990

(Superceded by TR-91-72)

A new technique of generating a test sequence for conformance testing of communication protocols is presented. This approach shows that it is possible to generate optimal length test sequences which include multiple unique input output (UIO) sequences and overlapping under certain conditions. In the absence of the above mentioned conditions, a heuristic technique is used to obtain sub-optimal solutions which show significant improvement over optimal solutions without overlapping. We also compare the computational complexity of our algorithm with that of existing techniques. Finally, a brief discussion of bounds on test sequence length is presented and our results are compared against these bounds.

TR-90-31 Two New Approaches to Conformance Testing of Communication Protocols — Raymond E. Miller — 1990

A new technique of generating minimal length test sequences for conformance testing of communication protocols using multiple Unique Input Output (UIO) sequences and *overlapping* is described and illustrated using several examples. The second approach shows how one can generate test sequences to give total fault coverage for certain classes of faults that are prescribed ahead of time.

TR-91-42 Generating Test Sequences with Guaranteed Fault Coverage for Conformance Testing of Communication Protocols — Raymond E. Miller — 1990

A technique for generating test sequences for conformance testing of communication protocols which provide guaranteed fault coverage is presented. Issues related to fault coverage are discussed and solutions to some of the problems are provided.

TR-91-43 Specification and Analysis of a Data Transfer Protocol Using Systems of Communicating Machines — Raymond E. Miller — December 1990

A model for communication protocols called *systems of communicating machines* is used to specify a data transfer protocol with variable window size (e.g., HDLC), which is an arbitrary nonnegative integer, and to analyze it for freedom from deadlocks. The model uses a combination of finite state machines and variables. This allows the size of the specification (i.e., number of states and variables) to be linear in the window size, a considerable reduction from the pure finite state machine model. A new type of analysis is demonstrated which we call *system state analysis*. This is similar to the *reachability analysis* used in the pure finite state model, but it provides substantial simplification by reducing the number of states generated. For example, with the protocol in this paper, if w is the window size, then the global analysis produces $O(w^5)$ states, while the system state analysis produces $O(w^3)$ states. The system state analysis is then combined with an inductive proof, extending the analysis to all nonnegative integers w.

TR-91-72 Modified Version of Generating Minimal Length Test Sequences for Conformance Testing of Communication Protocols — Raymond E. Miller

A new technique of generating test sequence for conformance testing of communication protocols is presented. This approach shows that it is possible to generate optimal length test sequences which include multiple unique input ouput (*UIO*) sequences and overlapping under certain conditions. In the absence of the above mentioned conditions, a heuristic technique is used to obtain sub-optimal solutions which show significant improvement over optimal solutions without overlapping. We also compare the computational complexity of our algorithm with that of existing techniques. Finally, a brief discussion of bounds on test sequence length is presented and our results are compared against these bounds.

TR-92-77 Generating Test Sequences with Guaranteed Fault Coverage for Conformance Testing of Communication Protocols — Raymond E. Miller — December 1991

A theoretical analysis of fault coverage of conformance test sequences for communication protocols specified as Finite State Machines (FSM's) is presented. Faults of different types are considered and their effect on testing is analyzed. The interaction between faults of different categories and the impact it has on conformance testing is investigated. Fault coverage is defined for testing of incompletely defined machines and also for testing of completely defined machines. An algorithm is presented to generate test sequences with maximal fault coverage for testing of incompletely specified machines. It is then augmented for testing of completely defined machines and finally a technique for generating test sequences which provide guaranteed maximal fault coverage for conformance testing of communication protocols is presented.

TR-92-78 On the Generation of Minimal Length Conformance Tests for Communication Protocols — Raymond E. Miller — November 1991

A new technique of generating a test sequence for conformance testing of communication protocols is presented. This approach shows that it is possible to generate optimal length test sequences which include multiple unique input output (*UIO*) sequences and overlapping under certain conditions. In the absence of the above mentioned conditions, a heuristic technique is used to obtain sub-optimal solutions which show significant improvement over optimal solutions without overlapping. We illustrate our technique on a practical example; the NBS Class 4 Transport Protocol (TP4). We also compare the computational complexity of our algorithm with that of existing techniques. Finally, a brief discussion of bounds on test sequence length is presented and our results are compared against these bounds.

TR-92-80 On Generating Test Sequences for Combined Control and Data Flow for Conformance Testing of Communication Protocols — Raymond E. Miller

Using a limited version of Estelle, from which a specification can be represented in terms of an Extended Finite State Machine (EFSM), we develop a technique to generate conformance tests which test both the data flow as well as the control flow. From the EFSM, we generate a Finite State Machine (FSM) with several transitions corresponding to a single transition of the EFSM. Moreover, the input and output parameters are also modified so that an "equivalent" FSM is obtained. The data flow graph (DFG) is constructed directly from the "equivalent" FSM. Test segments are obtained from the data flow graph as well as from the control flow graph and are combined "carefully" to generate an executable test sequence. Test data for the above sequence is chosen using a mutation technique to guarantee detection of specific kinds of faults in the data flow. Control flow is tested in the conventional way.

TR-92-85 Faults, Errors and Convergence in Conformance Testing of Communication Protocols — Raymond E. Miller

The objective of this paper is to focus on two important terms - fault and error and point out the imprecision in estimating fault coverage using simulation techniques. Based on the above discussion, error coverage of a test sequence is defined. In addition to that, another goal is to establish the significance of the notion of convergence in both specification and implementation and show how it plays an important role in determining the error coverage of an implementation.

TR-92-86 Research Issues for Communication Protocols — Raymond E. Miller

The keynote address for Twelfth International Conference on Distributed Computing Systems in Yokohama, Japan on June 10, 1992.

TR-92-89 Structural Analysis of a Protocol Specification and Generation of a Maximal Fault Coverage Conformance Test Sequence — Raymond E. Miller — July 1992

A theoretical analysis of fault coverage of conformance test sequences for communication protocols specified as Finite State Machines (FSM's) is presented. Faults of different types are considered and their effect on testing analyzed. The interaction between faults of different categories and the impact it has on conformance testing is investigated. Fault coverage is defined for testing of incompletely specified machines and also for testing of completely specified machines. An algorithm is presented to generate test sequences with maximal fault coverage for testing of incompletely specified machines. It is then augmented for testing of completely specified machines and finally a technique for generating test sequences which provide guaranteed maximal fault coverage for conformance testing of communication protocols is presented.

TR-93-97 Generating Maximal Fault Coverage Conformance Test Length for Communication Protocols — Raymond Miller Sanjoy — January 1993

This paper focuses on a technique to reduce the length of maximal fault coverage test sequences for communication protocols by removing "redundant" test segments. This approach conceptually begins with all the test segments needed for the generation of maximal fault coverage test sequences as in [MP92], analyzed the structure of the specified Finite State Machine (FSM) for the protocol and shows that certain segments in these tests are unnecessary to guarantee maximal fault coverage. From this analysis an algorithm is proposed for generating the reduced length sequences that still guarantee maximal fault coverage. We describe how these tests are in some sense minimal, or near minimal, length test sequences without losing fault coverage.

TR-93-98 Bounding the Performance of FDDI — Raymond E. Miller — February 1993

Increased demand on bandwidth combined with more reliable high speed network media, has resulted in the emergence of a new family of protocols, called High Speed protocols. The performance of these protocols is much studied by queuing models. Although useful in some cases, the complexity of queuing analysis obscures the fine details of a protocol. This paper is an attempt towards providing a better understanding of high speed protocols by using simple analysis techniques which complement information obtained from statistical studies.

In particular, we have selected FDDI as a representative of high speed protocols for the illustration of our technique. Our paper provides bounds on the performance of FDDI under various load conditions and traffic types. Media utilization and the station delay are used as performance measures. We show how cyclic behavior of FDDI changes under various conditions. The bounds found in this paper can be used to select FDDI parameters for the desired network behavior.

TR-93-106 DQDB Performance and Fairness as Related to Transmission Capacity — Raymond E. Miller — August 1993

Operational characteristics of the DQDB protocol have been studied mainly under overload conditions using simulations and some limited analyses. The unfairness in DQDB under overload conditions has been the main focus of these studies. Many variants of DQDB have been proposed to overcome this unfairness. This paper investigates a different aspect of DQDB behavior. We demonstrate through analysis and supporting simulations that the protocol behaves in a fair manner as long as the total load presented to the network does not exceed its full capacity.

Previous studies have shown that unfairness under overload conditions is dependent on the internode distances, the initial slot-patterns on the buses and the number of nodes. This paper shows that as long as the network does not enter the overload region, the protocol operations remains fair independent of these parameters. We also show that if the network moves from an overload condition to full or less load, the network again returns to fair operation. We have used node throughput and node delay as performance parameters. The results presented in this paper can serve as guidelines for operating a DQDB network when unfairness becomes significant. It suggests that additional capacity may be a better solution than other modifications to DQDB to obtain fairness.

TR-93-107 Deadlock Detection for Protocols Using Generalized Fair Reachability — Raymond E. Miller — September 1993 Cyclic Analysis

In this paper, the notion of fair reachability is generalized to cyclic protocols with $n > 2$ communicating finite state machines. An equivalence relation is established between the set of fair reachable states and the set of reachable global states with equal channel length. Based on this result, we show that the deadlock detection problem is decidable for cyclic protocols with finite fair reachability graphs. We also show that for any cyclic protocol with one bounded channel, its fair reachability graph is finite. As far as decidability of deadlock detection is concerned, our result extends the class of cyclic protocols studied in Peng and Purushothaman *A Unified Approach to the Deadlock Detection Problem in Networks of Communicating Finite State Machines.* Furthermore, our results are similar to those of Pachl's in an unpublished report *Reachability Problems for Communicating Finite State Machines*, but our approach is significantly different. Our decision procedure is much more straightforward and efficient, as compared to the ones in [6] and [8]. In this respect, we have improved the complexity of deadlock detection for the class of cyclic protocols with finite fair reachability graphs.

TR-94-109 Generalized Fair Reachability Analysis for Cyclic Protocols: Part 1 — Raymond E. Miller — January 1994

In this paper, the notion of fair reachability is generalized to cyclic protocols with $n>2$ communicating finite state machines. An equivalence is established between the set of fair reachable states and the set of reachable states with equal channel length. As a result, *deadlock detection* is decidable for cyclic protocols with finite far reachability graphs. The concept of *simultaneous unboundedness* is defined and the lack of it is shown to be a necessary and sufficient condition for a cyclic protocol to have a finite fair reachability graph. For the first time, we are able to exactly characterize the class of protocols studied by Peng & Purushothaman, and complements the one investigated by Pachl. More importantly, our decision procedure is much more straightforward and efficient, as compared to Pachl's and the one by Peng & Purushothaman. In this respect, we have improved the complexity of deadlock detection for the class of cyclic protocols with finite far

reachability graphs. To further demonstrate the strength of generalized fair reachability analysis, we also show that *livelock detection* is decidable for the class of cyclic protocols with finite fair reachability graphs.

Matthew O'Keefe, University of Minnesota

TR-94-118 Performance Characteristics of a 100 MegaByte/second Disk Array — Matthew T. O'Keefe — April 1994

Disk arrays offer performance, capacity, and reliability greater than that of a single disk drive by employing Redundant Array of Inexpensive Disks (RAID) techniques. In this paper we describe a hierarchical disk array configuration that is capable of sustaining a 100 MByte/second transfer rate. We have identified and measured overheads associated with virtual memory management, command setup, and the array controller and employed them to construct and verify a simple performance model. Unlike a previous study in which the backplane bus was a key bottleneck, our measurements showed that virtual memory management and striping granularity played the key roles in limiting performance. To reduce these overheads we propose a fine-grain interleaved striping driver to replace the traditional logical volume driver and detail the potential performance improvements possible with this new approach.

Kathleen Perez-Lopez, George Mason University

TR-94-120 Use of Subband Decomposition for Management of Scientific Image Databases — Kathleen G. Perez-Lopez — June 1994

Managing massive databases of scientific images requires new techniques that address indexing visual content, providing adequate capabilities, and facilitating querying by image content while representing the data in a concise, essentially lossless manner. This is particularly important where access is over widely distributed networks. Subband decomposition of image data using wavelet filters is offered as an aid to solving each of these problems. It is fundamental to a visual indexing scheme that constructs a pruned tree of significant subbands as a first level of index. Significance is determined by feature vectors including Gibbs random field statistics, in addition to more common measures of energy and entropy. Features are retained at the nodes of the pruned subband tree as a second level of index. Query image, indexed in the same manner as database images, are compared as closely as desired to database indexes. Browse images for matching images are transmitted to the user in the form of coefficients for selected subbands, which constitute the third level of index. The subband decomposition is compatible with an image compression schemes based on wavelet packet analysis, such as the FBI fingerprint image compression standard.

Terrence Pratt, CESDIS

TR-92-90 Kernel-Control Parallel Versus Data Parallel: A Technical Comparison — Terrence W. Pratt — September 1992

The *dusty deck* problem is the problem of how to automatically transform large existing serial/vector codes into a form suitable for efficient execution on new parallel architectures. *Kernel-control parallel* (KCP) methods provide a promising approach to solving the dusty deck problem for MIMD distributed memory machines. This paper is a status report on a project to develop this technology into a full-scale prototype for running large Fortran 77 codes on the Intel iPSC/860. Space limits preclude a full description of the methods. Instead, we sketch the approach and then use a series of questions to explore some of the key differences between this new technology and existing data parallel methods, which are similar but better known.

John Reif, Duke University

TR-90-04 On the Bit-Complexity of Discrete Solutions of PDEs: Compact Multigrid — John Reif — 1990

The topic of Partial Differential Equations (PDEs) is an interesting area where the techniques of discrete mathematics and combinatorial algorithms can be brought together to solve problems which would normally be considered more properly in the domain of continuous mathematics. We investigate the bit-complexity of discrete solutions to linear PDEs. We show that for a large class of PDEs, the solution of an N point discretization can be compressed to only a constant number of bits per discretization point, without loss of information or introducing errors beyond discretization error. We show that the bit-complexity of the compressed solution is O (N) for both storage space and the total number of operations. We also show that we can compute the compressed solution by a parallel algorithm using O (log N) time and N/log N bit-serial processors, provided that all the coefficients of the PDE are bounded integers of the magnitude O (1). The best previous bounds on the bit-complexity (for both sequential time and storage space) were at least N log N; furthermore, an order of N log N bit-serial processors were required to support the O (log N) parallel time in the known algorithms. We believe this is the first case where a linear or algebraic system can be provably compressed (i.e., the bit-complexity of storage of the compressed solution is less than the solution size) and also the first case where the use of data compression provably speeds up the time to solve the system (in the compressed form).

TR-91-44 An Exact Algorithm for Kinodynamic Planning in the Plane — John Reif

A long-standing open problem in robotics has been that of devising algorithms for generating time-optimal motions under kinodynamic constraints. This problem has been considered previously in the literature and approximation algorithms have been provided for the two and three dimensional cases [CDRX] but with the exception of the one-dimensional case, no exact algorithms have been given. In this paper, we provide the first exact algorithm for time-optimal kinodynamic motion planning in the two-dimensional case.

TR-91-51 BLITZEN: A Highly Integrated Massively Parallel Machine — John Reif

The goal of the BLITZEN project is to construct a physically small, massively parallel, high-performance machine. This paper present the architecture, organizations, and feature set of a highly integrated SIMD array processing chip which has been customer designed and fabricated for this purpose at the Microelectronics Center of North Carolina. The chip has 128 processing elements (PEs), each with 1K bits of static RAM. Unique local control features include the ability to modify the global memory address with data local to each PE, and complementary operations based on a condition register. With a 16K PE system (only 128 custom chips are needed for this), operating at 20 MHz, data I/O can take place at 10,240 megabytes per second through a new method using a 4-bit bus for each set of 16 PEs. A 16K PE system can perform IEEE standard 32-bit floating-point multiplication at a rate greater than 450 megaflops. Fixed point addition on 32-bit data exceeds the rate of three billion operations per second. Since the processors are bit-serial devices, performance rates improve with shorter word lengths. The BLITZEN chip is one of the first to incorporate over 1.1 million transistors on a single die. It was designed with 1.25-µm, two-level metal, CMSO design rules on an 11.0 by 11.7-mm die.

TR-91-52 Efficient Parallel Algorithms for Optical Computing with the DFT Primitive — John Reif

The optical computing technology offers new challenges to the algorithm designers since it can perform an *n*-point DFT computation in only unit time. Note that DFT is a non-trivial computation in the PRAM model. We

develop two new models, DFT-VLSIO and DFT-Circuit, to capture this characteristics of optical computing. We also provide two paradigms for developing parallel algorithms in these models. Efficient parallel algorithms for many problems including polynomial and matrix computations, sorting and string matching are presented. The sorting and string matching algorithms are particularly noteworthy. Almost all of these algorithms are within a polylog factor of the optical computing (VLSIO) lower bounds derived in [BR87] and [TR89].

TR-91-55 An Optimal Parallel Algorithm for Graph Planarity — Vijaya Ramachandran, John Reif — 1989

We present a parallel algorithm based on open ear decomposition which, given a graph G on n vertices, constructs an embedding of G onto the plane or reports that G is nonplanar. Our parallel algorithm runs on a CRCW PRAM in O(logn) time with the same processor bound as graph connectivity.

TR-91-56 A Randomized EREW Parallel Algorithm for Finding Connected Components in a Graph — Hillel Gazit, John Reif

We present a parallel *EREW* randomized algorithm for finding the connected components of an undirected graph. Our algorithm takes $T = O($ log (n)+log $(g))$ time an $P = O\ (\frac{n+m}{\log n})$ p rocessor, where m = number of edges, n = number of vertices and g is the genus of the graph.

TR-91-61 Image Compression Methods with Distortion Controlled Capabilities — Tassos Markas, John Reif

Traditionally, lossy compression schemes have focused on compressing data at fixed bit rates in order to communicate information over limited bandwidth communication channels, or to store information in a fixed-size storage media. In this paper we present a class of lossy data compression algorithms that are capable of encoding images so that the loss of information complies with certain distortion requirements. The developed algorithms are based on the Tree-Structured Vector Quantizers (TSVQ). The first distortion controlled algorithm uses variable-size image blocks, encoded on quad-tree data structures, to efficiently encode image areas with different information content. Another class of distortion controlled algorithms that are presented is based on recursive quantization of error image blocks that represent the difference between the current approximation and the original block. We will also describe the progressive compression properties of these algorithms. Finally, we will present their compression/distortion performance using satellite images provided by NASA, and we will show that they achieve better performance than the TSVQ algorithms at high bit rates.

Kenneth Salem, University of Maryland

TR-90-02 Altruistic Locking — Kenneth Salem — 1990

Long lived transactions (LLTs) hold on to database resources for relatively long periods of time, significantly delaying the completion of shorter and more common transactions. To alleviate this problem we proposed an extension to two-phase locking, called altruistic locking, whereby LLTs can release their locks early. Transactions that access this released data are said to run in the wake of the LLT and must follow special locking rules. Altruistic locking guarantees serializability and does not *a priori* specify an order in which database objects must be accessed.

TR-90-09 Coordinating Multi-Transaction Activities — Kenneth Salem — 1990

Data processing applications must often execute collections of related transactions. We propose a model for structuring and coordinating these multi-transaction activities. The model includes mechanisms for communication between transactions, for compensating transactions after an activity has failed, for dynamic creation and binding of activities, and for checkpointing the progress of an activity.

TR-91-46 Adaptive Prefetching for Disk Buffers — Kenneth Salem

Prefetching is a common technique for improving performance in memory hierarchy. Effective prefetching relies on accurate predictions of upcoming data requests. This report presents several prefetching techniques that attempt to learn to predict accurately by monitoring the stream of data requests. Their performance is evaluated using trace-driven simulations.

TR-91-59 Probabilistic Diagnosis of Hot Spots — Kenneth Salem

Commonly, a few objects in a database account for a large share of all database accesses. These objects are called hot spots. The ability to determine which objects are hot spots opens the door to a variety of performance improvements. Data reorganization, migration, and replication techniques can take advantage of knowledge of hot spots to improve performance at low cost. In this paper we present some techniques that can be used to identify those objects in the database that account for more than a specified percentage of database accesses. Identification is accomplished by analyzing a string of database references and collecting statistics. Depending on the length of the reference string and the amount of space available for the analysis, each technique will have a non-zero probability of false diagnosis, i.e., mistaking "cold" items for hot spots and vice versa. We compare the techniques analytically and show the tradeoffs among time, space and the probability of false diagnoses.

TR-91-62 Management of Partially-Safe Buffers — Kenneth Salem

Safe RAM is RAM which has been made as reliable as a disk. We consider the problem of buffer management in mixed buffers, i.e., buffers which contain both safe RAM and traditional volatile RAM. Mixed-buffer management techniques explicitly consider the safety of memory in deciding where to place recently read or written data. We present several such techniques, along with a simple model of a mixed buffer and its backing store. Using trace-driven simulations, we compare their effectiveness at reducing I/O to and from the backing store.

TR-91-63 Non-Deterministic Queue Operations — Kenneth Salem

Queues play a central role in transaction processing systems. We present a transaction model that allows significant concurrency improvements for extended queue operations such as non-blocking dequeue, priority dequeue, non-blocking enqueue, and others.

TR-91-75 Placing Replicated Data to Reduce Seek Delays — Kenneth Salem

In many environments, seek time is a major component of the disk access time. In this paper we introduce the ideas of replicating data on a disk to reduce the average seek time. Our focus is on the problem of placing

replicated data. We present several techniques for replica placement and evaluate their performance using trace-driven simulations.

TR-92-81 MR-CDF: Managing Multi-Resolution Scientific Data — Kenneth Salem

MR-CDF is a system for managing multi-resolution scientific data sets. It is an extension of the popular CDF (Common Data Format) system. MR-CDF provides a simple functional interface to client programs for storage and retrieval of data. Data is stored so that low-resolution versions of the data can be provided quickly. Higher resolutions are also available, but not so quickly. By managing data with MR-CDF, an application can be relieved of the low-level details of data management, and can easily trade data resolution for improved access time.

TR-92-82 Adaptive Block Rearrangement — Kenneth Salem

An adaptive technique for reducing disk seek times is described. The technique copies frequently-referenced blocks from their original locations to reserved space near the center of the disk. Reference frequencies need not be known in advance. Instead, they are estimating by monitoring the stream of arriving requests. Trace-driven simulations show that seek times can be cut in half by copying only a small number of blocks using this technique. It is designed to be implemented in a device driver or controller, and is independent of the file system or database manager that uses the disk.

TR-92-88 Summary Report of the CESDIS Workshop on Scientific Database Management — Kenneth Salem — September 1992

The CESDIS Workshop on Scientific Database Management was held in September, 1992 on the campus of the University of Maryland at College Park. The workshop consisted of invited presentations intended to promote discussion of research issues in scientific database management. This report is an overview of those presentations, including copies of the speakers' foils.

TR-93-101 Implementing Extended Transaction Models Using Transaction Groups — Kenneth Salem — April 1993

The last five years have witnessed the introduction of numerous extended transaction models. These models relax the ACID properties provided by transactions, replacing them with weaker guarantees. Despite their popularity, relatively little has appeared in the literature on implementing extended transactions. This paper describes *transaction groups*, a simple generic implementation of extended transactions which can be customized to provide the basic features of many different extended transaction models. Customization to several proposed models, including nested sagas, flexible transactions, and others, is illustrated. Transaction groups themselves can be implemented in a transaction processing system without modifying its existing transaction managers and resource managers. The proposed implementation exploits the well-documented ideas of recoverable storage, triggers, and exactly-once transaction execution. Transaction groups can also be implemented in federated systems.

TR-93-102 Adaptive Block Rearrangement Under UNIX — Kenneth Salem — April 1993

An adaptive UNIX disk device driver is described. The driver copies frequently-referenced blocks from their original locations to reserved space near the center of the disk to reduce seek times. Reference frequencies need not be known in advance. Instead, they are estimated by monitoring the stream of arriving requests.

Measurements show that the adaptive driver reduces seek time by more than half, and improves response times significantly.

TR-93-105 Space Efficient Hot Spot Estimation — Kenneth Salem — July 1993

This paper is concerned with the problem of identifying names which occur frequently in an ordered list of names. Such names are called hot spots. Hot spots can be identified easily by counting the occurrences of each name and then selecting those with large counts. However, this simple solution requires space proportional to the number of names that occur in the list. In this paper, we present and evaluate two *hot spot estimation* techniques. These techniques guess the frequently occurring names, while using less space than the simple solution. We have implemented and tested both techniques using several types of input traces. Our experiments show that very accurate guesses can be made using much less space than the simple solution would require.

Ramin Samadani, Stanford University

TR-90-06 Changes in Connectivity in Active Contour Models — Ramin Samadani — March 1989

One approach to motion detection is to track the positions of contours in an image sequence through time. In certain applications the objects both move and change their connectivity; cell division is one example. In this paper contours are extracted from objects moving in two dimensions whose motions is non-rigid and whose connectivity may change. A previously proposed solution for non-igid motion, which involves the use of simulated elastic curves to track objects, is extended to allow elastic materials to break, grow and connect open endpoints. The extensions allow object tracking even when the connectivity of the objects changes. An algorithm for tracking objects that divide in two is developed based on these extensions. The algorithm is tested using a computer generated image sequence simulating cell division.

TR-90-12 Model-Driven Image Analysis to to Augment Databases — Ramin Samadani — April 1989

In this paper we consider how information may be obtained from images. To search large image collections we need to search on secondary parameters. We may look for images containing certain types of objects, for images where the objects are of certain size or shape, or for images having certain features. Since we now have techniques to rapidly acquire and store many images, we need techniques for automatic image analysis to generate such parameters. This paper describes a promising category of image analysis, namely model-driven methods. Two examples, operating in very different domains, are presented.

TR-90-17 Evaluation of an Elastic Curve Technique for Automatically Finding the Auroral Oval from Satellite Images — Ramin Samadani — July 1989

The DE-1 satellite has gathered over 500,000 images of the Earth's aurora using ultraviolet and visible light photometers. About 600 new auroral images a day are expected from planned satellites. Manual methods are currently used for feature extraction. But, to allow the scientific use of more of these image, automated techniques for feature extraction must be used.

A feature of interest to geophysicists is the location and shape of the boundaries of the auroral oval. We have implemented a system for finding the inner boundary based on a recently proposed computer vision technique. The technique is analogous to solving the equations of motion for an elastic curve in a damped

medium, where the forces are provided by the image. The resulting equilibrium position of the elastic curve provides an automated method for finding the shape and location of the inner boundary of the auroral oval.

Two methods are used to evaluate the automatic boundary finding system. Both methods are based on consistency checks with manual measurements. It is found that the automatic method follows the general shape of the auroral oval very well. The method, however, smoothes some irregularities found with the manual method.

TR-90-21 Finding Curvilinear Features in Speckled Images — Ramin Samadani — July 1989

This paper describes a method for finding thin curves in digital images with speckle noise. The solution method differs from standard linear convolutions followed by thresholds in that it explicitly allows curvature in the features. Maximum aposteriori (MAP) estimation is used, together with statistical models for the speckle noise, and for the curve generation process, in order to find the most probable estimate of the feature, given the image data.

The estimation process is first described in general terms. Then, incorporation of the specific neighborhood system and a multiplicative noise model for speckle allows derivation of the solution, using dynamic programming, of the estimation problem. The detection of curvilinear features is considered separately. The detection results allow the determination of the minimal size of a detected feature. Finally, the estimation of linear features, followed by a detection step is shown for computer simulated images and for a synthetic aperture radar (SAR) image of sea ice.

TR-90-25 A Computer Vision System for Automatically Finding the Auroral Oval from Satellite Images — Ramin Samadani — February 1990

The DE-1 satellite has gathered over 500,000 images of the Earth's aurora. Finding the location and shape of the boundaries of the oval is of interest to geophysicists, but manual extraction of the boundaries is extremely time consuming. This paper describes a computer vision system that automatically provides an estimate of the inner auroral boundary for winter hemisphere scenes. The system performs automatic checks of its boundary estimate. If the boundary estimate is deemed inconsistent, the system does not output it. The performance of this system is evaluated using 44 DE-1 images. The system provides boundary estimates for 37 of the inputs. Of these 37 estimates, 31 are consistent with the corresponding manual estimates. At this level of performance, the supervised use of the system provides more than one order of magnitude increase in throughput compared to manual extraction of the boundaries.

TR-90-29 Evaluation of an Elastic Curve Technique for Automatically Finding the Auroral Oval from Satellite Images — Ramin Samadani — March 1990

The DE-1 satellite has gathered over 500,000 images of the Earth's aurora using ultraviolet and visible light photometers. A feature which has geophysical significance is the inner boundary of the auroral oval. Manual methods are currently used for feature extraction. We describe an automated algorithm for finding the inner boundary based on a recently proposed computer vision technique. The algorithm is analogous to solving the equations of motion for an elastic curve, where the forces are provided by the image. The resulting equilibrium position of the elastic curve provides an automated method for finding the shape and location of the inner boundary of the auroral oval.

Two methods, both based on comparisons with manual measurements, are developed for the evaluation of the automated algorithm. The first method compares the areas within the automated and the manual boundaries. The second method measures the overlap between the interiors of the two boundaries. The expected variation between two sets of manual measurements is used to set an upper bound to the allowed

discrepancy between the automated results and a single set of manual measurements. The algorithm, when tested with 71 satellite images, is found to perform best for those images without overlap between the aurora and the daylight hemisphere.

TR-91-48 Adaptive Control of Parameters for Active Contour Models — Ramin Samadani — 1991

The stability of active contour models or "snakes" is studied. It is shown that the modification of snake parameters using adaptive systems improves both the stability of the snakes and the results obtained. The adaptive snakes perform better with images of varying contrasts, noisy images and images with different curvatures along the boundaries. The computational costs at each iteration for the adaptive snakes is still of order *N*, where *N* is the number of points on the snakes. Comparisons re made between non-adaptive and adaptive snakes using computer simulations and satellite images. The additional costs for the adaptive snakes are modest.

TR-91-53 This Technical Report has been superseded by TR-92-87

A Minimization-Pruning Algorithm for Finding Elliptical Boundaries in Images with Non-Constant Background and with Missing Data — Ramin Samadani — 1991

TR-91-73 Adaptive Image Segmentation Applied to Extracting the Auroral Oval From Satellite Images — Ramin Samadani — 1991

There are over 500,000 global images of the earth's aurora available from the DE-1 satellite. About 1000 of such images per day are expected in the near future from planned satellites. Furthermore, finding the boundaries which delineate the aurora is necessary for scientific studies of the interaction of the sun's solar wind with the earth's plasma environment. Currently, the boundaries are extracted manually, which is time consuming. Therefore, the vast majority of the images are not analyzed. In order to make more of the existing and the expected data available for scientific studies, automated image analysis techniques are required. This paper describes an adaptive image segmentation algorithm which is applied to the difficult summertime scenes to extract inner auroral oval data.

TR-92-84 Adaptive Snakes: Control of Damping and Material Parameters — Ramin Samadani

The stability of active contour models or "snakes" is studied. It is shown that the modification of snake parameters using adaptive systems improves both the stability of the snakes and the boundaries obtained. The adaptive snakes perform better with images of varying contrasts, noisy images and images with different curvatures along the boundaries. The computational costs at each iteration for the adaptive snakes is still of order *N*, where *N* is the number of points on the snakes. Comparisons of the results for non-adaptive and adaptive snakes are shown using both computer simulations and satellite images.

TR-92-87 A Minimization-Pruning Algorithm for Finding Elliptical Boundaries in Images with Non-Constant Background and with Missing Data — Ramin Samadani — 1991

(Supersedes TR-91-53)

The DE-1 satellite has gathered over 500,000 images of the Earth's aurora using ultraviolet and visible light photometers. The extraction of the boundaries delimiting the aurora oval allows the computation of important parameters for the geophysical study of the phenomenon such as total area and total integrated magnetic field. This paper describes an unsupervised technique that we call "minimization-pruning" that finds the boundaries of the aurora oval. The technique is based on concepts that are relevant to a wide range of applications having characteristics similar to this application, namely images with variable background, high noise levels and missing data. Among the advantages of the new technique are the ability to find the object of interest even with intense interfering background noise, and the ability to find the outline of an object even if only a section of it is visible. The technique is based on the assumption that certain regions of the object are less obscured by the background, and hence the information provided by these regions is more important for finding the boundaries. The implementation of the technique consists of an iterative minimization-pruning algorithm, in which a fundamental part is a measure of the quality of the data for different regions along the boundary. Calculation of this measure is simplified by transforming the input image into polar coordinates. The technique has been applied to a set of more than 100 images of the aurora with good results. We also show examples of extraction of the inner and outer boundaries starting from the elliptical approximation and analyzing the image locally around that solution.

TR-94-114 Computer Assisted Analysis of Auroral Images Obtained from High Altitude Polar Satellites — Ramin Samadani — February 1994

We developed automatic techniques that allow the extraction of physically significant parameters from auroral images. This allows the processing of a much larger number of images than is currently possible with manual techniques. We applied our techniques to diverse auroral image datasets. We made these results available to geophysicists at NASA and at universities in the form of a software system that performs the image analysis. After some feedback from users, we transferred an upgraded system to NASA and to two universities. We demonstrated the feasibility of user-trained search and retrieval of large amounts of image data using our automatically derived parameter indeces. We developed and applied techniques based on classification and regression trees (CART) to broaden the types of images to which the automated search and retrieval may be applied. We tested our techniques with DE-1 auroral images.

Hikmet Senay, George Washington University

TR-90-05 Rules and Principles of Scientific Data Visualization — Hikmet Senay — May 1990

This report provides a set of rules and principles for scientific data visualization. These rules and principles have been acquired through informal discussions with data visualization experts and surveys of existing literature on graphics, data visualization, visual perception, exploratory data analysis, psychology, and human-computer interaction. Even though far from being complete and extensive, the set provided in this report forms a starting point for designing effective scientific data visualization techniques. Using these rules and principles, we are currently developing a visualization tool assistant (VISTA) which will advise scientists and engineers, who are not visualization experts, in selecting and creating effective data visualizations.

TR-90-11 VISTA: Visualization Tool Assistant for Viewing Scientific Data — Hikmet Senay — May 1990

Scientific data visualization has become an important field which combines computer graphics and data analysis techniques to allow scientists to visually analyze large data sets, like those typically found in the Earth and space sciences. Often, new facts and knowledge hidden within the data can be discovered through effective visualizations. However, not all visualization techniques for representing data are useful for all data sets. This means that the desired facts and knowledge can be made visible only if an appropriate visualization technique is applied to reveal the content of the given data set. The selection and creation of the right visualization technique to extract useful knowledge from data requires familiarity with the characteristics of the data as well as expertise in visualization. Scientists who are primarily concerned with the content of the data usually lack the knowledge of how the data is organized in a database or how it could be most effectively visualized using the available visualization tools and techniques. In this paper, we describe a knowledge-based visualization tool assistant (VISTA) which is being designed to assist scientists in the scientific data visualization process.

TR-90-16 Compositional Analysis and Synthesis of Scientific Data Visualization Techniques — Hikmet Senay — 1990

Scientific data visualization has become an important discipline which supports scientist in exploring data, looking for patterns and relationships, proving or disproving hypotheses, and discovering new phenomenon using graphical methods. Although visualization has such an important role in scientific discovery, it is still an art which requires significant knowledge in several fields, such as data management, computer graphics, and visual perception. However, those scientist who could benefit most from visualization tools and techniques usually lack the knowledge. Furthermore, there are very few guidelines for selecting and creating effective data visualization techniques. In order to reveal the knowledge that is essential for effective data visualization, the existing techniques that are known to be useful have been thoroughly analyzed. These analyses have led to the identification of several visualization primitives and a set of rules that can be used to design effective data visualization techniques. The results of the analyses further suggest a compositional approach to automating the synthesis of scientific data visualization techniques.

TR-91-60 Multi-Media Interaction with Virtual Worlds — Hikmet Senay

This paper is an outgrowth of a Visualization '90 panel which focused on interaction issues in visualization including requirements, techniques, and devices. Since multi-media interaction with visual representation of complex information spaces (virtual worlds) was the primary focus of discussion at the panel, the paper is devoted to further elaborating this issue. Starting with a brief technology description identifying significant characteristics of multi-media interaction techniques and devices, the paper describes how visual aspects of virtual worlds can be enhanced to provide additional cues by other media stimulating senses other than vision. Finally, several applications of this technology, illustrating multi-media interaction with visualization of complex information spaces, are presented.

TR-92-79 A Knowledge Based System for Scientific Data Visualization — Hikmet Senay

This paper describes a knowledge-based system, called visualization tool assistant (VISTA), which was developed to assist scientists in the design of scientific data visualization techniques. The system derives its knowledge from several sources which provide information about data characteristics, visualization primitives, and effective visual perception. The design methodology employed by the system is based on a sequence of transformations which decomposes a data set into a set of data partitions to visualization primitives, and combines these primitives into a composite visualization technique design. Although the primary function of the system is to generate an effective visualization technique design for a given data set by using principles of visual perception, the system also allows users to interactively modify the design, and renders the resulting image using a variety of rendering algorithms. The current version of the system primarily supports

visualization techniques having applicability in Earth and space sciences, although it may easily be extended to include other techniques useful in other disciplines such as computational fluid dynamics, finite-element analysis and medical imaging.

Douglas Smith, Carnegie Mellon University

TR-93-96 A Virtual Machine for High Performance Image Processing — Doug Smith — January 1993

In this report, I describe my work in developing a virtual architecture for high performance image processing. this introductory section provides the motivation for the research, and then briefly describes the contents of the remaining sections of the report.

Cynthia Starr, George Washington University

TR-91-74 Parallel Programming on the Silicon Graphics Workstation Using the Multiprocessing Library — Cynthia L. Starr — June 1991

This paper examines methods of explicit parallel programming on Silicon Graphics workstations. It presents techniques for designing parallel programs along with a brief review of circumstances that can result in incorrect computations. The parallel programming facilities provided by the Multiprocessing Library are examined in detail. Although parallel programming facilities provided by the operating system libraries are referred to during the course of the discussion, detailed coverage of the system calls is beyond the scope of this paper.

Thomas Sterling, CESDIS

TR-93-100 Fine Grain Dataflow without Tokens for Balanced Execution — Thomas Sterling — March 1993

Synchronization overhead and communication contention for fine-grain parallel processing are eliminated by the Associative Template Dataflow (ATD) architecture examined in this paper. Associative synchronization permits all elements of a distributed control state to be updated in a single cycle through a broadcast technique. A generalization of scoreboarding to include multiple memory elements as well as execution resources is employed to circumvent contention and respond to latency. Operations and their readiness state are maintained by a dataflow representation. A simulator of the ATD architecture has been developed and application programs have been written for experimentation. This paper presents the results of these experiments and demonstrates that high utilization of critical execution elements is achieved. Costs and programming mapping issues and methods are also considered.

TR-93-103 The Realities of High Performance Computing and Dataflow's Role in It: Lessons from the NASA HPCC Program — Thomas Sterling — January 1993

Dataflow addresses critical questions that would seem to be the inhibiting factors to successful exploitation of parallel processing. Yes, as the high performance computing community launches a major new National initiative to harness the potential of massively parallel processing and deliver useful Teraflops performance capacity into the hands of applications scientists, the immediate role of dataflow concepts and architecture appears uncertain. This paper reflects on the realities of current trends in supercomputing based on the

experience of the NASA High Performance Computing and Communications Program. Included are findings from the recent Pasadena Workshop on Systems Software and Tools for High Performance Computing Environments. From this perspective, the disparity between the perceived needs of the high performance computing community and primary contributions of the dataflow research community is presented. Consideration is given to ways in which the elegance of the dataflow model can be brought to bear on the real-world evolution of future generation supercomputing.

David Stotts, University of Maryland

TR-90-03 Modeling the Logical Structure of Flexible Manufacturing Systems with Petri-Nets — P. David Stotts — August 1989

Flexible Manufacturing Systems are computer-integrated systems which have many concurrent components, very complicated logical relations, and a distributed computer system structure. They have been increasingly and rapidly adopted for use in industry. Basic Petri-net definitions, both classical and extended for timing analysis, are reviewed in the paper. In addition, recent research using Petri-net theory as applied to the design and analysis of FMSs is reviewed. In particular, one of the most flexible of manufacturing line structures is discussed, the Robot Lattice Structure, which is analyzed using a new form of timed Petri-nets, termed Binary Timed Petri-Nets. A graphical modeling language for BTPNs is also briefly discussed.

TR-90-10 Bounding Procedure Execution Times in a Synchronous Petri Net Computation Model — P. David Stotts — 1990

A model of concurrent, time-dependent software systems is introduced that is based on a combination of hierarchical graph (h-graph) theory with timed Petri-nets. Termed the HG *model of time-dependent concurrent software systems,* it is intended to support a variety of performance studies of real-time concurrent computations. Concurrent control flow in the HG model is represented by a marked, timed Petri-net which is distinguished by its notion of place duration and its synchronous concurrent transition firing rule. This paper briefly presents the formal concepts on which the HG model is based, and illustrates its utility in calculating time bounds for procedure call executions. This computation uses the *synchronous concurrent reachability graph,* which is a restricted form of the classical Petri-net reachability graph.

TR-90-15 Coverability Graphs for a Class of Synchronously Executed Unbounded Petri Net — P. David Stotts and Terrence W. Pratt — 1990

Synchronous (or concurrent) transition firing rules for Petri nets are useful in modeling computations on real-time systems with multiple processors. A synchronous firing rule is one in which more than one transition may be fired to effect a single state change, allowing the physically concurrent operation of multiple hardware processors to be represented in the state sequence without including intermediate states that have no meaningful physical interpretation. A simple counter example illustrates that the standard method of generating a Petri net coverability graph is insufficient to represent the reachability set of a Petri net operating under synchronous firing rule. We describe a variant of one widely used concurrent execution rule (that of firing maximal subsets) in which the simultaneous firing of conflicting transitions is prohibited. An algorithm is then given for constructing the coverability graph of a net executed under this synchronous firing rule. The w insertion criteria in the algorithm are shown to be valid for any net on which it terminates; the set of nets for which the algorithm terminates is then shown to properly include the *conflict-free* class.

TR-90-20 YTRACC: An Interactive Debugger for YACC Grammars — P. David Stotts — 1990

We describe a program for the display and exploration of complex, domain-specific information: ytracc, an interactive grammar debugging tool for compiler writers. The ytracc system provides the designer of a yacc grammar a method of tracing a parser as it uses the grammar. Ytracc captures the states of the parse as it is carried out. The captured parse then can be replayed forwards or backwards, step-by step, or subtree-by-subtree, as defined by the nonterminals of the grammar. The tool has been successfully used by students as an assistant in an advanced undergraduate compiler construction class, and we use the tool in our everyday work.

TR-90-24 A Mills-Style Iteration Theorem for Nondeterministic Concurrent Programs — P. David Stotts — 1990

For Mills' program verification method, we prove a theorem for the meaning of a nondeterministic general iteration structure like that described by Dijkstra and Gries. This structure also allows expression of a class of concurrent algorithms, ones that can be viewed as the nondeterministic interleaving of a group of sequential computations. The theorem thus extends Mills' verification method into the domain of nondeterministic and concurrent programs.

TR-90-28 Generalizing Hypertext — P. David Stotts — 1990

All highly-interactive systems share features in common, and hypertext is no exception. As the underlying model of a hypertext system becomes generalized and the category of problems that can be solved broadens, the degree of overlap increases. In this paper, we describe our own general model of hypertext, called Trellis, and discuss the overlaps that we have observed. When fully generalized, a hypertext model can operate in multiple domains simultaneously.

TR-90-32 Increasing the Power of Hypertext Search with Relational Queries — P. David Stotts — 1990

We describe an SQL relational database schema for representing the objects in HyperCard, along with a technique for automatically populating this schema from a HyperCard stack using the facilities in HyperTalk with calls to the database manager. The standard relational database query language SQL can then be used to perform more general hypertext searches than are possible with the string search feature found in most hypertext browsing environments. Semiautomatic updates of the content of a hypertext are also possible using SQL updates on the object representations in the database to trigger corresponding HyperCard updates on the objects themselves. We describe a prototype implementation and present several example queries and updates to motivate this approach. These techniques, although demonstrated here specifically using HyperCard and Oracle for Macintosh, are generally applicable to a wide range of hypertext systems and relational databases.

TR-90-34 Temporal Hyperprogramming — P. David Stotts — 1990

The visual programming aspects of Trellis hypertext documents are described. A hypertext is a non-linearly organized, browsable information structure. The importance of browsing distinguishes hypertext from other network information systems. The possible experiences a user may have when interacting with a hypertext are as important as its form. Further, these *browsing semantics* should be an inherent characteristic of a

document, not of the implementation that allows browsing. In essence, a hypertext is an active entity that has a visible behavior, not a static entity that is manipulated by external means.

The Trellis model employs the dual nature of Petri nets to formally express both aspects of a hypertext in one structure. A Petri net is a bipartite graph, so it captures the linked structure of relationships among information elements. It is also an automaton, having an execution state and state transition rules, thereby formally capturing the interactions between reader and document. In this report, we define the temporal semantics of the Trellis model and illustrate them with a prototype hypertext system call *a*Trellis. This environment joins timed events and active computing engines into a dynamic, parallel browsing structure. In *a*Trellis, hypertext authoring is visual programming for a temporally-synchronized, visual outcome--temporal hyperprogramming.

TR-90-36 Programmable Browsing Semantics in Trellis — P. David Stotts — 1990

Different researchers have different ideas about how a hypertext should be navigated. Each new implementation of a hypertext browser works slightly differently from previous ones. This is due both to variations in personal taste and to discoveries of new, useful ways to organize and present information. In this report we outline a technique by which a hypertext system can offer flexible, programmable browsing behavior, or *browsing semantics.* Differences in the way documents are to be browsed can be specified by an author on a document-by document basis, or by a style designer for an entire class of documents. The ability to specify and modify how a browser presents information is an important and useful property in general. We first discuss the issues involved in programmable browsing semantics, and then we present one method of providing them within the context of the Trellis project at the University of Maryland.

TR-90-38 Separating Hypertext Content from Structure in Trellis — P. David Stotts — 1990.

We have defined a Petri-net-based model of hypertext (the Trellis hypertext model). This model distinguishes the hypertext's structure, content, and context. The structure of the hypertext describes the relationships that tie together elements of the content (e.g., the adjacencies or links). The hypertext's reader examines the content, which may be textual, graphical, or perhaps audible. The association of content with the structure defines the context in which the content is presented to the reader.

As we have investigated the Trellis model, and have designed and implemented prototypes based upon it, we have come to realize that separating the content from structure increases the flexibility of the system for the hypertext's author. However, the separation has also required inclusion of special mechanisms in the prototype implementation to ensure that the presentation of the context to the hypertext's reader is easy to comprehend. In this paper, we describe and discuss the implementation and implications separating content from structure as reflected in our hypertext model.

TR-90-39 Hierarchy, Composition, Scripting Languages, and Translators for Structured Hypertext — P. David Stotts — 1990

In this paper we describe a hypertext translator-generator system that uses *x*Ted, the visual Petri net editor from the *x*Trellis hypertext system, to specify the semantic component of a string-to-graph translation. *x*Ted-specified parsers convert general authoring notations into structured *x*Trellis documents for browsing. The operative mechanism is termed a *pair grammar,* in which a string grammar and a graph grammar are paired in a one-to-one correspondence. When an *x*Ted-specified parser reduces by one of its string grammar productions, the corresponding production in its graph grammar is used to generate a portion of the Petri net that implements that syntax. The use of pair grammars in conjunction with the Trellis model results in a general method of defining hypertext structure that is both *hierarchical* and *compositional.*

TR-90-40 Browsing Parallel Process Networks — P. David Stotts — 1990

A hypertext is a non-linearly organized, linked information structure, designed to be interactively browsed. In this report we demonstrate the use of a hypertext system for visualization and simulation of the parallel control flow and message network of concurrent programs. Instead of constructing code browsers as special-purpose window systems that understand the syntax and semantics of particular languages, we present a general approach that relies on the (usually extensive) browsing facilities of a hypertext system. Language-specific browsers can be realized by generating filters to convert program text into hypertext documents. We demonstrate our filter approach specifically with the parallel language CSP and the aTrellis hypertext system. aTrellis employs the dual nature of Petri nets to express formally in one structure both the linked information elements of a document and the reader/document interactions during browsing. aTrellis is especially appropriate for parallel program browsing because Petri nets are a natural concurrency model, and because it offers state-space analysis for deadlocks and other program properties. aTrellis-based program browsers for other parallel languages can be produced by designing appropriate Petri net translations.

TR-90-41 aTrellis: A System for Writing and Browsing Petri-Net-Based Hypertext — P. David Stotts — 1990

We have developed a new model of hypertext in pilot studies. The traditional hypertext model resembles a directed graph, representing information fragments and the relationships that tie the fragments together. Our model, based on Petri nets, also represents the hypertext's *browsing semantics* (i.e., how the information is to be visited). The Petri net model is a generalization of traditional directed graph models. It permits development of browsing and authoring systems that can incorporate the analytical techniques that have been developed for Petri nets and also incorporate the user interface designs that have been developed for hypertext systems. The Petri net basis also permits flexible specification of how a hypertext is to be browsed. New abilities include synchronization of simultaneous traversals of separate paths through a hypertext and incorporation of security/access control considerations into the linked structure of a hypertext (specifying nodes that can be proven accessible only to certain classes of browsers). In addition, different tailored versions can be generated from a single document structure with a Petri-net basis.

This report describes the Petri-net-based Trellis hypertext model, a prototype hypertext implementation call aTrellis, and an early version of an authoring language for Petri-net-based documents called *Alpha*.

TR-91-45 Place/Transition Nets with Debit Arcs — P. David Stotts

We add an extension called *debit arcs* to traditional place/transition nets (P/T-nets, also known as Petri nets). A debit arc allows its destination transition to fire whenever desired, but records a *debt* (or *antitoken*) in its source place if no token is there to be consumed. A normal token can annihilate with an antitoken, which can be thought of as "paying off" the debt. Two natural rules for token/antitoken annihilation (*instantaneous*, and *delayed*) are examined and are shown to create two distinct classes of automation in terms of language recognition power. Under instantaneous annihilation, nets with debit arcs are equivalent as a class to Turing machines, and so extend the modeling power of standard P/T-nets. Under delayed annihilation, nets with debit arcs are equivalents as a class to standard P/T-nets, and thus are only a notational convenience. Nets with debit arcs are shown to be a special case of colored nets.

TR-91-50 Structured Dynamic Behavior in Hypertext — P. David Stotts

In this paper we discuss a hypertext system that provides structure to the dynamic behavior inherent in reader/document interactions. We first describe what structure means for interactive information systems, and then show how timing and synchronization aspects, coupled with a client/server architecture, allow structure

to be expressed in a non-linear interactive document. This structure is created with a graphical editor, and can be interacted with by several different types of browsers simultaneously.

TR-91-54 A Functional Meta-Structure for Hypertext Models and Systems — P. David Stotts — December 1990

We describe hypertext "meta-structure"—one that provides an organization for the architecture of hypertext models and systems. The meta-structure was initially developed to help us understand the architecture of a specific hypertext model (the Trellis hypertext model). However, its organization seems generally applicable to a wide range of other models and systems as well. As such, the meta-structure is a good candidate for a high-level hypertext *reference model*, and so we refer to it as the *Trellis hypertext reference model*, or the *r-model*. The r-model represents a hypertext at five levels of abstraction—two abstract levels, two concrete levels, and one visible level. In this paper, we present the r-model, use it to classify four different hypertext (and hypertext-like) systems, and then discuss its relationship to various hypertext-defined concepts.

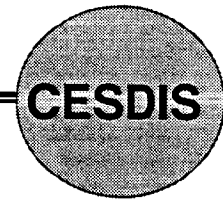
TR-91-64 Dynamic Adaptation of Hypertext Structure — P. David Stotts

A technique is described for adapting the apparent structure of a hypertext to the behavior and preferences exhibited by its users while browsing. Examples are given of the Trellis implementation of this technique, which employs the timing mechanism in Trellis; event durations in a document are altered without actually changing the links in the underlying Petri net. The two extreme of instantaneous events and infinite delays can be used to create apparent node and link deletions and additions, as well as to insert new tokens (loci of activity) into a document. Adaptation of these times is accomplished using a simple data state in which the event timings (and other document properties) are variables, called *attributes*. As a reader traverses hypertext links, author-supplied *adaptation agents* are invoked to collect information and possible change the values of the attributes. Agents encapsulate and effect the criteria for deciding when, and specifically how, a structure should be adapted. Several practical examples illustrated the conclusion of this report: sophisticated alterations do not require a complicated adaptation mechanism, that changing document constants into document variables provides flexibility to this mechanism, and that using a limited simple mechanism is the only hope for retaining analysis of the static and dynamic net properties.

Stephen Tate, Duke University

TR-93-99 Report on the Workshop on Data and Image Compression Needs and Uses in the Scientific Community — Stephen R. Tate — February 1993

On December 17, 1992, the Workshop on Data and Image Compression Needs and Uses in the Scientific Community was held at the Goddard Space Flight Center. Talks were given by both data compression researchers and applications scientists, and valuable communication was started between these two scientific communities.



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

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	Modelling 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)*

Number of Copies Requested	Report Number	Title
_____	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)*

Number of Copies Requested	Report Number	Title
_____	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)*
_____	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)*

Number of Copies Requested	Report Number	Title
_____	TR-91-57	Study of Six Linear Least Square Fits *(Eric Feigelson)*
_____	TR-91-58	Fast Computations of Vector Quantization Algorithsm *(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 $\frac{7}{3} \cdot \sqrt{n}$ Seperator 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 Segementation 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 Lenth Conformance Tests for Communication Protocols *(Raymond E. Miller)*

Number of Copies Requested	Report Number	Title
_____	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)*

Number of Copies Requested	Report Number	Title
_____	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-94-126. 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)*

Number of Copies Requested	Report Number	Title
_____	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)*

NAME _____

ADDRESS _____

_____

_____

_____

PHONE _____

E-MAIL _____

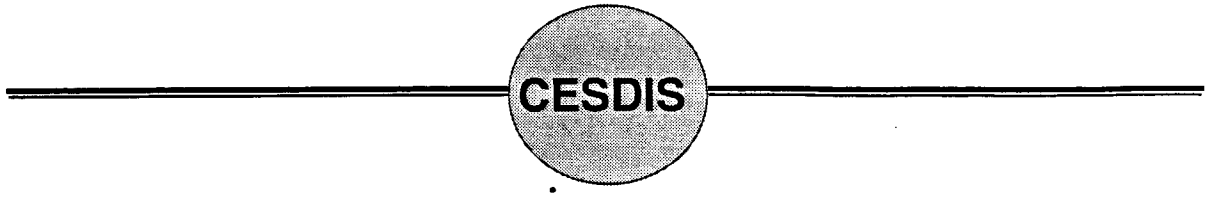
AREAS OF RESEARCH INTEREST_____

_____

—

APPENDIX D

Dissertation Series



CENTER OF EXCELLENCE IN SPACE DATA AND INFORMATION SCIENCES
CODE 930.5
NASA GODDARD SPACE FLIGHT CENTER
GREENBELT, MD 20771

301-286-4403 Internet: cas@cesdis1.gsfc.nasa.gov

DISSERTATION SERIES

Series Number	Title	Author	Date
DS-93-01	Data Compression: Algorithms and Architecutres	Tassos Markas	April 1993

Advisor: John Reif, Duke University, Computer Science Department

This dissertation presents the design of efficient lossy data compression for stand-still and multispectral images, as well as the design of high-performance parallel architectures suitable for compression of image and textual data. This dissertation is organized in three parts. The first part (chapter one) deals with some basic concepts of data compression; it describes the motivation factors that led me to pursue this work, it gives a brief description of the contributions of each research effort, and it concludes with an overview of existing distortion measures used to measure the loss of information in compressed images.

The second part is focused on the development of lossy compression algorithms for image data and it contains three independent research efforts. The first effort (chapter two) deals with a class of distortion controlled compression algorithms, where an image is encoded in such a way that the loss of information in the reconstructed image satisfies certain user requirements. An efficient encoding of the discrete wavelet transform, based on multidimensional bitmap trees, is presented in chapter three. This method has shown better compression/distortion performance compared to other existing compression methods, including the JPEG standard. Chapter four is devoted to the compression of multispectral images. A hybrid data compression algorithm that is based on histogram equalization and transform/subband coding has been designed. This algorithm takes into consideration the spectral and spatial redundancies found in multispectral images, and it outperforms existing methods. Its compression performance varies 20-30:1 for perceptually lossless quality, to ratios exceeding 100:1 suitable for browsing type applications.

The third part of this dissertation is devoted to the development of high-performance parallel architectures, suitable for real-time compression of image and text data. For image data, a shared-memory parallel architecture that implements a fast version of the tree-structured vector quantization algorithm will be presented in chapter five. The last chapter presents a systolic-type parallel architecture that is capable of compressing text data at high-speeds without any loss of information. This architecture implements a parallel version of the textual substitution algorithm, which is a variation of the compression algorithm found in UNIX systems.

Series Number	Title	Author	Date
DS-93-02	Feature Identification in Data Represented as Images	Domingo Mihovilovic	July 1992

Advisor: Ronald Bracewell, Stanford University, Department of Electrical Engineering

Enormous amounts of data for scientific purposes are recorded and archived, and only a very small fraction is studied. This trend will continue and become even more severe as the data are acquired at higher rates and mainly as images. For example, the "Mission to Planet Earth" program plans to send several petabytes (10^{15} bytes!) of data each day down to Earth. This amount of information cannot be processed without the assistance of automatic or semi-automatic tools. This dissertation covers two areas of research that deal with the automatic identification of features in data represented as images.

First, this dissertation describes a computer vision application in which the features of interest are boundaries of the auroral oval. The DE-1 satellite has gathered over 500,000 digital images of the Earth's aurora using ultraviolet and visible light photometers. The extraction of the boundaries delimiting the auroral oval allows the computation of important parameters for the geophysical study of the phenomenon such as total area and total integrated magnetic field. The system reads the data in the same format used by the satellite and after finding the aurora boundaries, it computes and outputs total area of the polar cap and total integrated magnetic field. This research introduces an unsupervised "minimization-pruning" technique that finds the boundaries of the auroral oval. The technique is based on the intuitive idea that certain regions of the object are less obscured by the background, and hence the information provided by these regions is more important. Among the advantages of the new technique are the ability to find the object of interest even with intense interfering background noise, and the ability to find the outline of an object even if only a section of it is visible.

Second, this dissertation describes a new technique for the analysis of non-stationary signals. In this case, the time-frequency plane is considered as an image and the features of interest are the time-frequency characteristics of the components of a signal. The dynamic spectrum (spectrogram, running window Fourier transform or Gabor information diagram) provides a natural way of showing the energy content as a function of time. It has been used in the analysis of many phenomena including seismic activity, speech, magnetospheric whistlers, underwater sound, and storm waves. The uncertainty principle sets the limit for both time and frequency resolution, defining a resolution cell as a rectangular box of sides Δt and Δf. Large time resolution (small Δt) implies small frequency resolution (large Δf) and vice versa.

The Gabor diagram is reviewed, and then a method of analysis based on a new type of elementary signal, called a chirplet, is introduced. The chirplet extends the Gabor diagram in two ways. First, it allows the aspect ratio of the cells to vary. Second, it allows for oblique cells by introducing a drift rate parameter. This type of cell can resolve structure that is oriented diagonally in the time-frequency plane, and that could not have been resolved by using the customary Gabor cells. Chirplet diagrams are compared to the customary Gabor technique and examples of magnetospheric whistlers will be shown for actual and computer generated data. Finally, a way of obtaining information from the data by analyzing the distribution of interference nulls in the time-frequency plane will be described.

DS-93-03	A Structural Analysis Approach for Designing Efficient and Effective Algorithms for Conformance Testing of Communication Protocols	Sanjoy Paul	December 1992

Advisor: Raymond E. Miller, University of Maryland, College Park, Department of Computer Science

This dissertation concerns generating test sequences for conformance testing of communication protocols. Two conflicting issues in the design of a test sequence are its length and its fault coverage. A shorter length test sequence is time efficient while a longer test sequence tends to give higher fault coverage.

In the first part of the thesis, a novel technique is proposed to generate *minimal length* test sequences. It utilizes the property of *overlapping* between test segments (subsequences). Next, protocol specification machines are *structurally analyzed* to understand the phenomena of *fault hiding* and *fault detection*. The information obtained from this theoretical analysis is used to generate test sequences with *maximal fault coverage*. The next part of the dissertation removes *redundancy* from the maximal fault coverage test sequence to obtain a *near minimal length* test sequence with *maximal fault coverage*.

A dimension of complexity is added to the problem of conformance testing by introducing extended finite state machines (EFSM's) to specify protocol entities. In this case, both the control flow and the data flow of the protocol entity can be tested. A technique is proposed in the dissertation to generate *executable* test sequences for *combined control flow and data flow* of a protocol entity specified as a single module of an EFSM. The test sequence has been shown to have a very high fault coverage with respect to control flow and pre-specified faults of data flow.

The problem of conformance testing becomes complicated when the notion of communication between finite state machines is introduced. Testing a protocol entity specified as a collection of communication finite state machines (CFSM's) is a challenging problem not only from a theoretical viewpoint but also from practical considerations. This dissertation provides some insight into this problem and suggests a heuristic technique which is practical and has a reasonably good fault coverage. The fault coverage can be further improved by dynamically altering a parameter during the testing process.

CENTER OF EXCELLENCE IN SPACE DATA AND INFORMATION SCIENCES
CODE 930.5
NASA GODDARD SPACE FLIGHT CENTER
GREENBELT, MD 20771

DISSERTATION SERIES ORDER FORM

301-286-4403 Internet: cas@cesdis1.gsfc.nasa.gov

Number of Copies Requested	Series Number	Title
_____	DS-93-01	Data Compression: Algorithms and Architectures *(Tassos Markas)*
_____	DS-93-02	Feature Identification in Data Represented as Images *(Domingo A. Mihovilovic')*
_____	DS-93-03	A Structural Analysis Approach for Designing Efficient and Effective Algorithms for Conformance Testing of Communication Protocols *(Sanjoy Paul)*

Name: _____

Affiliation: _____

Address: _____

_____

Phone: _____

Email: _____

APPENDIX E

CESDIS Personnel and Associates

(As of June 1994)

CESDIS ADMINISTRATIVE OFFICE
Voice: 301-286-4403
Fax: 301-286-1777
Internet: cas@cesdis1.gsfc.nasa.gov

All individual extensions will roll over to the main number if the party called does not answer by the fourth ring.

U.S. Mail Address

CESDIS
Code. 930.5
Goddard Space Flight Center
Greenbelt, MD 20771

Federal Express/UPS Address

CESDIS
Nimbus Road, Bldg. 28, Room W223
Goddard Space Flight Center
Greenbelt, MD 20771

ACTING DIRECTOR

Terrence Pratt 301-286-4108
E-mail ..... pratt@cesdis1.gsfc.nasa.gov

STAFF SCIENTISTS

Don Becker 301-286-0882
E-Mail ..... becker@cesdis1.gsfc.nasa.gov

Jacqueline Le Moigne 301-286-8723
E-Mail ..... lemoigne@nibbles.gsfc.nasa.gov

Robert Mack ....., 301-286-9595
E-Mail ..... mack@ame.gsfc.nasa.gov

Phillip Merkey 301-286-3805
E-Mail ..... merk@cesdis1.gsfc.nasa.gov

Thomas Sterling 301-286-2757
E-Mail ..... tron@cesdis.edu

TECHNICAL PERSONNEL

Thomas Hood 301-286-0879
E-Mail ..... hood@cesdis1.gsfc.nasa.gov

Michele O'Connell 301-286-8830
E-Mail ..... oconnell@cesdis.gsfc.nasa.gov

Larry Picha 301-286-0879
E-Mail ..... lpicha@cesdis.gsfc.nasa.gov

ADMINISTRATIVE STAFF

Nancy K. Campbell 301-286-4099
E-Mail ..... campbell@cesdis.usra.edu

Robin L. Alford 301-286-4403
E-Mail ..... robin@cesdis.usra.edu

Georgia Flanagan 301-286-2080
E-Mail ..... georgia@cesdis.usra.edu

Annemarie Murphy 301-286-8951
E-Mail ..... murphy@cesdis.usra.edu

UNIVERSITY PROJECT PERSONNEL

Boston University
Center for Space Physics
725 Commonwealth
Boson, MA 02215

Supriya Chakrabarti, Principal Investigator 617-353-5990
E-Mail ..... supc@veebs.bu.edu

Brandeis University
Computer Science Department
Waltham, MA 02254-9110

James Storer, Principal Investigator 617-736-2714
E-mail ..... storer@cs.brandeis.edu

California Institute of Technology
Henry M. Robinson Laboratory of Astrophysics 105-24
Pasadena, CA 91125

S. George Djorgovski, Principal Investigator 818-395-4415
E-mail ..... george@oracle.caltech.edu

Clemson University
Department of Electrical and Computer Engineering
Clemson, SC 29634-0915

Walter B. Ligon, III, Principal Investigator 803-656-1224
E-Mail ..... walt@eng.clemson.edu

Duke University
Department of Computer Science
207 North Building
Durham, NC 27706

John Reif, Principal Investigator 919-684-3048
E-mail ..... reif@cs.duke.edu

George Washington University
Department of Electrical Engineering and Computer Science
Washington, DC 20052

Tarek El-Ghazawi, Principal Investigator 202-994-5507
E-mail ..... tarek@seas.gwu.edu

Hughes Applied Information Systems
Colorado Engineering Laboratories
16800 E. Centre Tech Parkway
Aurora, CO 80011

Vance McCullough, Project Contact 303-344-6145
E-mail ..... mccollou@redwood.hac.com

Institute of Global Environment and Society (IGES)
4041 Powder Mill Road, Suite 302
Calverton, MD 20705

James Kinter, Principal Investigator 301-902-1247
E-mail ..... kinter@cola.iges.org

Massachusetts Institute of Technology
Artificial Intelligence Laboratory
Room 715
545 Technology Square
Cambridge, MA 02139

Berthold Horn, Principal Investigator 617-253-5863
E-mail ..... bkph@ai.mit.edu

National Center for Atmospheric Research (NCAR)
Scientific Computing Division
1850 Table Mesa Drive
Boulder, CO 80307-3000

Bill Buzbee, Principal Investigator 303-497-1206
E-mail ..... buzbee@bierstadt.scd.ucar.edu

Science Applications International Corporation (SAIC)
Laboratory for Atmospheric and Space Science
1710 Goodridge Drive
McLean, VA 22102

Edward Szuszczewicz, Principal Investigator 703-734-5516
E-mail ..... szusz@mclapo.saic.com

Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218

Glenn Miller, Principal Investigator 410-338-4700
E-mail ..... miller@scivax.stsci.edu

Syracuse University
Northeast Parallel Architectures Center
111 College Place
Syracuse, NY 13244-4100

Geoffrey Fox, Principal Investigator 315-443-1723
E-mail ..... gcf@nova.npac.syr.edu

Texas A&M University
Meteorology Department
College Station, TX 77843

Kenneth Bowman, Principal Investigator 409-862-4060
E-mail ..... bowman@csrp.tamu.edu

University of Colorado
Colorado Center for Astrodynamics Research (CCAR)
Aerospace Engineering Sciences
Campus Box 431
Boulder, CO 80309

William Emery, Principal Investigator 303-492-8591
E-mail ..... bemery@orbit.colorado.edu

University of Colorado
Laboratory for Atmospheric and Space Physics
Campus Box 392
Boulder, CO 80309-0010

Randal Davis, Principal Investigator 303-492-6867
E-mail ..... davis@aquila.colorado.edu

University of Colorado
Colorado Space Grant Consortium
Campus Box 10
Boulder, CO 80309-0010

Elaine Hansen, Principal Investigator 303-492-3141
E-Mail ..... ehansen@rembrandt.colorado.edu

University of Florida
Department of Computer and Information Science
Gainesville, FL 32611-2024

Theodore Johnson, Principal Investigator 904-392-1492
E-mail ..... ted@cis.ufl.edu

University of Illinois
Department of Computer Science
Urbana, IL 61801

Daniel A. Reed, Principal Investigator 217-333-3807
E-mail ..... reed@oboe.cs.uiuc.edu

University of Illinois
National Center for Supercomputing Applications
Computing Applications Building
605 East Springfield Avenue
Urbana, IL 61820

Robert Wilhelmson, Principal Investigator 217-244-6833
E-mail ..... bw@ncsa.uiuc.edu

University of Maryland
Center for Automation Research
Computer Vision Laboratory
College Park, MD 20742-3275

Nathan S. Netanyahu, Principal Investigator 301-286-4652
E-mail ..... nathan@cfar.umd.edu

University of Maryland
Department of Computer Science
A.V. Williams Building
College Park, MD 20742

Nicholas Roussopoulos, Principal Investigator 301-405-2687
E-mail ..... nick@cs.umd.edu

University of Maryland, Baltimore County
Department of Computer Science
5401 Wilkens Avenue
Baltimore, MD 21228

Timothy Finin, Principal Investigator 410-455-3522
E-mail ..... finin@cs.umbc.edu

University of Minnesota
Department of Electrical Engineering
4-174 EE/CSci Building
Minneapolis, MN 55455

Matthew O'Keefe, Principal Investigator 612-625-6306
E-mail ..... okeefe@ee.umn.edu

University of Nebraska
Department of Computer Science and Engineering
Lincoln, NE 68588-0115

Philip Romig, Graduate Research Assistant 402-472-5271
E-mail ..... romig@cse.unl.edu

University of Texas
Department of Computer Science Engineering
Box 19015
Arlington, TX 76019

Diane Cook, Principal Investigator 817-273-3606
E-mail ..... cook@cse.uta.edu

University of Virginia
Department of Computer Science
Thornton Hall
Charlottesville, VA 22903

James French, Principal Investigator 804-982-2213
E-mail ..... french@cs.virginia.edu

University of Washington
Department of Computer Science and Engineering, FR-35
Seattle, WA 98195

Linda Shapiro, Principal Investigator 206-543-2196
E-Mail ..... shapiro@cs.washington.edu

University of Wisconsin
Computer Sciences Department
1210 W. Dayton Street
Madison, WI 53706

David DeWitt, Principal Investigator 608-263-5489
E-mail ..... dewitt@cs.wisc.edu

University of Wisconsin
Space Science and Engineering Center
1225 West Dayton Street
Madison, WI 53706

Sanjay Limaye, Principal Investigator 608-262-6541
E-mail ..... limaye@vms3.macc.wisc.edu

CRAY FELLOWSHIP

Jonathan Bright
Johns Hopkins University
Computer Science Department
3127 Guilford Avenue
Baltimore, MD 21218
410-467-2571
bright@blaze.cs.jhu.edu

Kathleen Perez-Lopez
Department of Computer Science
School of Information Technology and Engineering
George Mason University
Fairfax, VA 22030-4444
703-993-1536
klopez@cs.gmu.edu

CESDIS ASSOCIATES

Brandeis University
Computer Science Department
Waltham, MA 02254-9110

James Storer 617-736-2714
E-Mail ..... storer@cs.brandeis.edu

University of California
Computer Engineering Department
Room 225 Applied Sciences
Santa Cruz, CA 95064

Glen Langdon 408-459-2212
E-Mail ..... langdon@cse.ucsc.edu

University of Maryland
Department of Computer Science
College Park, MD 20742

Joel Saltz 301-405-2684
E-Mail ..... saltz@hyena.umd.edu

University of North Carolina
Department of Computer Science
Campus Box 3175, Sitterson Hall
Chapel Hill, NC 27599-3175

David Stotts 919-962-1833
E-Mail ..... stotts@cs.unc.edu
Fax 919-962-1799

ALPHABETICAL DIRECTORY OF CESDIS PERSONNEL

Alford, Robin L	301-286-4403	robin@cesdis.usra.edu
Becker, Don	301-286-0882	becker@cesdis1.gsfc.nasa.gov
Bowman, Kenneth	409-862-4060	bowman@csrp.tamu.edu
Bright, Jonathan	410-467-2571	bright@blaze.cs.jhu.edu
Buzbee, Bill	303-497-1206	buzbee@bierstadt.scd.ucar.edu
Campbell, Nancy	301-286-4099	campbell@cesdis.usra.edu
Chakrabarti, Supriya	617-353-5990	supc@veebs.bu.edu
Cook, Diane	817-273-3606	cook@cse.uta.edu
Davis, Randal	303-492-6867	davis@aquila.colorado.edu
DeWitt, David	608-263-5489	dewitt@cs.wisc.edu
Djorgovski, George	818-395-4415	george@oracle.caltech.edu
El-Ghazawi, Tarek	202-994-5507	tarek@seas.gwu.edu
Emery, William	303-492-8591	bemery@orbit.colorado.edu
Finin, Timothy	410-455-3522	finin@cs.umbc.edu
Flanagan, Georgia	301-286-2080	georgia@cesdis.usra.edu
Fox, Geoffrey	315-443-1723	gcf@nova.npac.syr.edu
French, James	804-982-2213	french@cs.virginia.edu
Hansen, Elaine	303-492-3141	ehansen@rembrandt.colorado.edu
Hood, Thomas	301-286-0879	hood@cesdis1.gsfc.nasa.gov
Horn, Berthold	617-253-5863	bkph@ai.mit.edu
Johnson, Theodore	904-392-1492	ted@cis.ufl.edu
Kinter, James	301-902-1247	kinter@cola.iges.org
Langdon, Glen	408-459-2212	langdon@cse.ucsc.edu
Le Moigne, Jacqueline	301-286-8723	lemoigne@nibbles.gsfc.nasa.gov
Ligon, Walter	803-656-1224	walt@eng.clemson.edu
Limaye, Sanjay	608-262-9541	limaye@vms3.macc.wisc.edu

ALPHABETICAL DIRECTORY OF CESDIS PERSONNEL (continued)

Mack, Robert	301-286-9393	mack@ame.gsfc.nasa.gov
McCollough, Vance	303-344-6145	mccollou@redwood.hac.com
Merkey, Phillip	301-286-3805	merk@cesdis1.gsfc.nasa.gov
Miller, Glenn	410-338-4700	miller@scivax.stsci.edu
Murphy, Annemarie	301-286-8951	murphy@cesdis.usra.edu
Netanyahu, Nathan	301-286-4652	nathan@cfar.umd.edu
O'Connell, Michele	301-286-8830	oconnell@cesdis.gsfc.nasa.gov
O'Keefe, Matthew	612-625-6306	okeefe@ee.umn.edu
Perez-Lopez, Kathleen	703-993-1536	kplopez@cs.gmu.edu
Picha, Larry	301-286-0879	lpicha@cesdis.gsfc.nasa.gov
Pratt, Terry	301-286-4108	pratt@cesdis1.gsfc.nasa.gov
Reed, Daniel	217-333-3807	reed@oboe.cs.uiuc.edu
Reif, John	919-684-3048	reif@cs.duke.edu
Romig, Phillip	402-472-5271	romig@cse.unl.edu
Roussopoulos, Nick	301-405-2687	nick@cs.umd.edu
Saltz, Joel	301-405-2684	saltz@hyena.umd.edu
Shapiro, Linda	206-543-2196	shapiro@cs.washington.edu
Sterling, Thomas	301-286-2757	tron@cesdis.edu
Storer, James	617-736-2714	storer@cs.brandeis.edu
Stotts, P. David	919-962-1833	stotts@cs.unc.edu
Szuszczewicz, Edward	703-734-5516	szusz@mclapo.saic.com
Wilhelmson, Robert	217-244-6833	bw@ncsa.uiuc.edu