NASA

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
Fast I/O for Massively Parallel Applications

5. Report Date

6. Performing Organization Code

7. Author(s)
Matthew T. O'Keefe

8. Performing Organization Report No.

10. Work Unit No.

9. Performing Organization Name and Address
University of Minnesota
1100 Washington Avenue South
Minneapolis, MN 55415-1226

11. Contract or Grant No. NAS5-32337
USRA subcontract No. 5555-23

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Washington, DC 20546-0001
NASA Goddard Space Flight Center
Greenbelt, MD 20771

13. Type of Report and Period Covered Final
July 1993 - October 1996

14. Sponsoring Agency Code

15. Supplementary Notes
This work was performed under a subcontract issued by
Universities Space Research Association
10227 Wincopin Circle, Suite 212
Columbia, MD 21044 Task 23

16. Abstract

The two primary goals for this report were the design, contruction and modeling of parallel disk arrays for scientific visualization and animation, and a study of the IO requirements of highly parallel applications. In addition, further work in parallel display systems required to project and animate the very high-resolution frames resulting from our supercomputing simulations in ocean circulation and compressible gas dynamics.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
disk arrays and parallel systems	Unclassified--Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	1	

NASA Form 1626 Oct 86

Final Report

Fast I/O for Massively Parallel Applications

USRA Grant No. 5555-23
National Aeronautics and Space Administration
Program Director: Dr. Terence Pratt

Principal Investigator: Dr. Matthew T. O'Keefe

Overview

The two primary goals for this research were the design, construction and modeling of parallel disk arrays for scientific visualization and animation, and a study of the IO requirements of highly parallel applications. In addition, we pursued further work in parallel display systems required to project and animate the very high-resolution frames resulting from our supercomputing simulations in ocean circulation and compressible gas dynamics.

Results and Transitions

With major additional support from the Army Research Office, NSF, and our corporate sponsors we constructed, modeled and measured several large parallel disk arrays. These arrays consisted of Ciprico 6700 RAID-3 devices (8 data + 1 parity drive) combined together in a variety of configurations, from a group of 8 RAID-3 from which we achieved nearly 100 MBytes/second transfer speed to a 31 array system that achieved a record 500 MBytes/second. These large bandwidths are necessary to support the high-resolution frame rates we require for the 2400x3200 pixel PowerWall parallel display system.

In addition to constructing these disk systems and measuring their performance, we developed performance models that capture many of the performance-limiting effects, such as start-up delays on RAID devices, fragmentation, and virtual memory page management overhead for very large transfers. We developed new techniques for instrumenting the kernel for taking filesystem performance data.

Other projects including performance measurements and experiments with D2 Helical Scan tapes from Ampex Corporation. We verified tape performance exceeding 15 MBytes/second for large transfers using the Ampex DST 310 tape device. In addition, Thomas Ruwart collaborated with storage vendor MTI on the construction of a 1-Terabyte filesystem using a collection of MTI RAID arrays.

Using the high speed disk subsystems to supply the bandwidth, we constructed a 4 panel PowerWall display system in our NSF-support Laboratory for Computational Science and Engineering following our successful (and partially NASA-sponsored) prototype at the Supercomputing '94 conference. A critical component of this system is the software that allows parallel rendering across the separate but seamlessly connected panels. Russell Catellan was partially supported by NASA to construct this sofware, which includes a version of XRaz used for scientific animation and also a modified version of VIZ, a 3D volume renderer developed in Norway. The PowerWall has inspired a host of imitations throughout the HPC community, including NASA Goddard. It is useful for a variety of high-resolution display applications, including our primary mission of visualizing and analyzing datasets generated by our simulation software on supercomputers.

Finally, we developed a package for performing parallel IO on the Cray T3D machine that is used by our regular grid applications such as the Miami Isopycnic Coordinate Ocean Model. This software is portable to other platforms, including the SGI Challenge class machines.

NASA support has helped produce two MS students and approximately 8 techical papers, as well as a variety of software and other research products, such as movies used by other researchers.

Graduate Theses Supported

Student's Name	Date	Degree	Thesis Title
Steve Soltis	June 1995	Masters	Instrumenting a UNIX Kernel for Event Tracing
Derek Lee	Feb 1995	Masters	Scientific Animation
Jeff Stromberg	pending	Masters	Performance Effects of File Fragmentation

Research Products

[1] Digital Movies: MPEG movies from the calculation described in journal reference [11] are available on the WWW at URL address: *"http://www-mount.ee.umn.edu/~dereklee/micom_movies/micom_movies.html"*. These movies were recently reference by Semtner in his article on computer simulations of ocean circulation which appeared in the September issue of *Science*. As of November 16th, there have been 1557 accesses to this Web page. Actual data from our runs is also available at the Web site.

[2] *The PowerWall Project:* in collaboration with Paul Woodward's team and several computer vendors, including Silicon Graphics Inc., Ciprico Inc., and IBM, my group helped to construct and demonstrate a high-resolution display system for datasets resulting from supercomputer simulations, medical imaging, and others. My group helped in the control software, data preparation and processing, and the actual physical constrution. This system was demonstrated at the *Supercomputing '94* conference and was described in conference publication [21]. A PowerWall, funded through an NSF CISE grant and witrh partial support from NASA and additional equipment grants from SGI and others, is now in operation in IT's *Laboratory for Computational Science and Engineering.* See our Web page on the PowerWall at URL *"http:///www-mount.ee.umn.edu/~okeefe"*.

Software Developed

[1] *PowerWall Control Software.* NASA support helped further the development of the control software for our parallel display system known as the PowerWall. This scalable display allows high-resolution supercomputer simulations to be shown in their totality to both small and large audiences. The disk array systems constructed partly with NASA support provided the more than 300 MegaBytes per second data throughput required by the PowerWall. First constructed at Supercomputing '94, we have constructed a PowerWall with NSF support in our own laboratory.

[2] *UNIX kernel trace* and *fragmentation measurement* routines. These routines provide a means of measuring OS kernel performance and the effects of file fragmentation. Available on the Web at URL address: *"http:www-mount.ee.umn.edu/~soltis"*.

Papers Published

[1] Thomas M. Ruwart and Matthew T. O'Keefe, "Performance Characteristics of a 100 MegaByte/Second Disk Array," *Storage and Interfaces '94*, Santa Clara, CA, January 1994.

[2] Aaron C. Sawdey, Matthew T. O'Keefe, Rainer Bleck, and Robert W. Numrich, "The Design, Implementation, and Performance of a Parallel Ocean Circulation Model," *Proceedings of the Sixth ECMWF Workshop on the Use of Parallel Processors in Meteorology,* Reading, England, November 1994. Proceedings published by *World Scientific Publishers* (Singapore) in **Coming of Age**, edited by G-R. Hoffman and N. Kreitz, 1995.

[3] Paul R. Woodward, "Interactive Scientific Visualization of Fluid Flow," *IEEE Computer,* Oct. 1993, vol. 26, no. 10, pp. 13-26.

[4] Thomas M. Ruwart and Matthew T. O'Keefe, "A 500 MegaByte/Second Disk Array," *Proceedings of the Fourth NASA Goddard Conference on Mass Storage Systems and Technologies,* pp. 75-90, Greenbelt, MD, March 1995.

[5] Aaron Sawdey, Derek Lee, Thomas Ruwart, Paul Woodward and Matthew O'Keefe, and Rainer Bleck, "Interactive Smooth-Motion Animation of High Resolution Ocean Circulation Calculations," *OCEANS '95 MTS/IEEE Conference,* San Diego, October 1995.

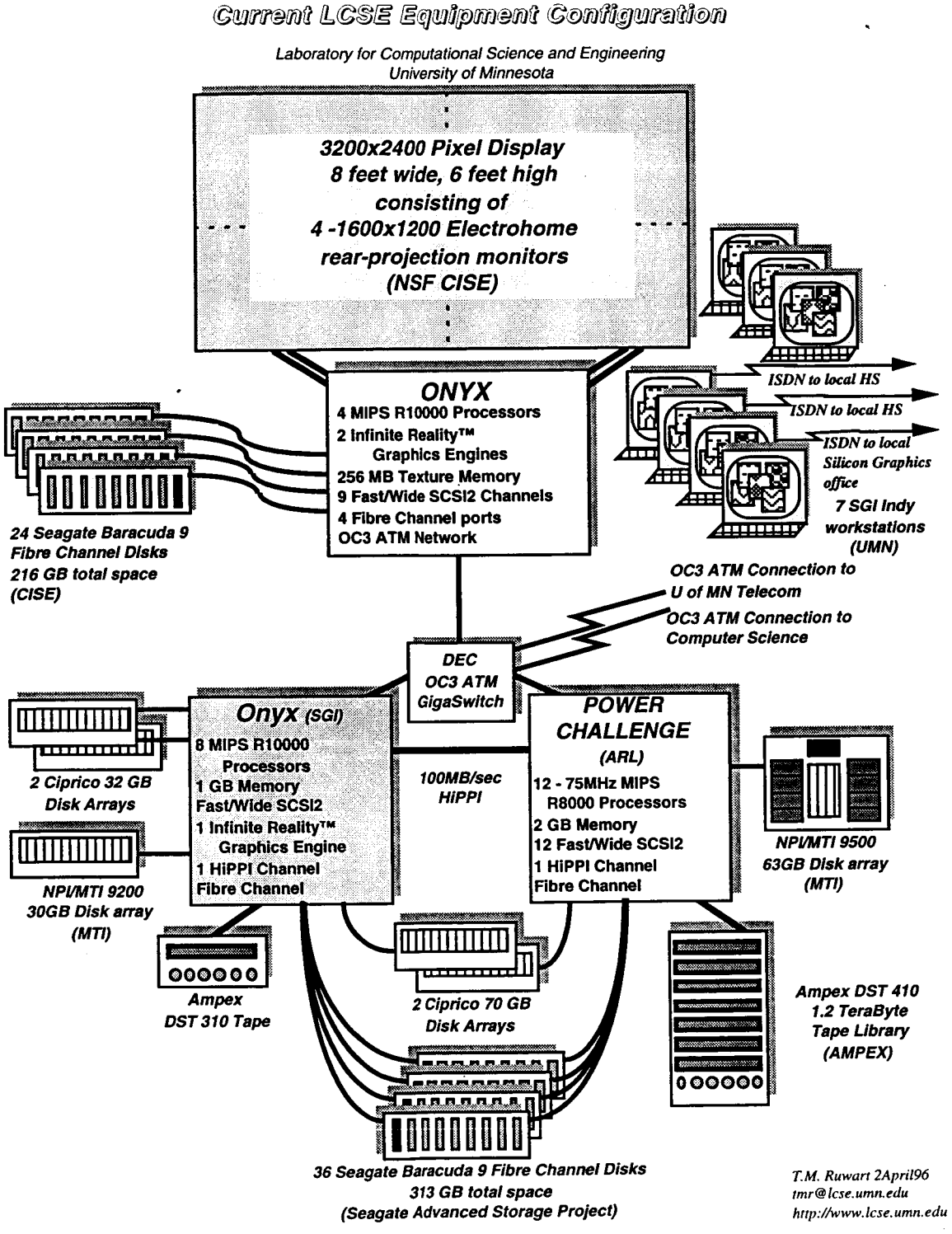
[6] Steve Soltis, Matthew O'Keefe, Thomas Ruwart and Ben Gribstad, "The Global File System (GFS)," to appear in the *Fifth NASA Goddard Conference on Mass Storage Systems and Technologies,* September 1996.

[7] Steven R. Soltis, Matthew T. O'Keefe and Thomas M. Ruwart, "Instrumenting a UNIX Kernel for Event Tracing," submitted to *Software: Practice and Experience*, 1995, under revision..

[8] Aaron C. Sawdey and Matthew T. O'Keefe, "A Software-level Cray T3D Emulation Package for SGI Shared-memory Multiprocessor Systems," submitted to *Software: Practice and Experience*, June 1995, under revision.

Technical Reports

[1] Aaron C. Sawdey, "Using the Parallel MICOM Code on the SGI Challenge Multiprocessor and the Cray T3D," technical report, University of Minnesota, available on the WWW at *http://www-mount.ee.umn.edu/~sawdey.*



Current LCSE Equipment Configuration including the PowerWall