Indiana University Shared University Research grants - Report on Accomplishments

June 20, 2000

Craig A. Stewart, Randall Bramley, Gerry Bernbom, Jon W. Dunn, Zdzislaw Meglicki, D.F. McMullen, David Hart, Mary Papakhian

Table of Contents

I. Executive Summary

Indiana University has a decades-old research relationship with IBM. This relationship has expanded and deepened significantly in the past two years, in large part thanks to the SUR grants awarded by IBM to IU, totaling $2.45M (retail value) in IBM high performance hardware. IU's accomplishments in SUR-funded research include the following:

- The creation of a unique, integrated system of supercomputing resources and massive data storage facilities.
- Pathbreaking research in component technologies in HPC through the development of the Common Component Architecture Toolkit. This research stands to revolutionize the way HPC systems are utilized.
- New development in use of nationally and internationally shared research instruments, aimed at providing "better-than-being-there" facilities for using these instruments and reducing time to discovery from weeks to hours or even minutes.
- Development of unique and innovative Mass Storage systems, including the creation of new code for HPSS that leverages IBM-developed software technology and expands its utility and performance. The development of this system also provided a cooperative testbed for IBM researchers in deploying innovative and unique installations of HPSS systems.
- Development of new Digital Libraries facilities, including extensive work with IBM in debugging new DL software and the independent development of significant Java-based extensions.

- Enhancement and distribution of an important community code in bioinformatics, including the demonstration of an international grid of RS/6000 SPs in attacking extremely large problems in evolutionary biology.
- The dissemination of GIS data in ways that are important to both government agencies and everyday citizens in Indiana, and the demonstration of the RS/6000 SP as a key component of this information delivery.
- The advancement of best practices in High Performance Computing, including the development of the facilities needed to support a large and diverse user community focused on the RS/6000 SP as the primary research computing environment at IU.

These accomplishments have been made possible by numerous ongoing and extensive collaborations between IU researchers and IBM researchers. We look forward to continued achievements and collaborations based on the existing SUR awards, as well as to future awards and even more dramatic research accomplishments.

II. Introduction

Indiana University and IBM have a long history of collaborative research. IBM support for IU researchers through hardware grants dates back to IBM awards that helped bring an IBM 3090 to IU in 1986. This hardware grant supported algorithm development by IU Distinguished Professor of Chemistry Ernest Davidson [1]. (IU's research history with IBM stretches farther back than that, however. One of the first 10 computer programs ever protected by copyright was a Fortran compiler for the IBM 709 called FASTRAN, by Frank Prosser, Steve Young, and Stan Hagstrom of IU. It was one of the key influences in the development of WATFOR. IBM never produced a FORTRAN compiler for the 709 that could match FASTRAN's performance.)

IU's relationship with IBM has in recent years entered into an unprecedented state of vigor and activity. SUR grants awarded to IU in the mid-90s helped establish the VARIATIONS program for digital music libraries [2], and brought to IU its first IBM RS/6000 SP. In the past two years IBM has awarded three SUR grants to IU, as follows:

- *VARIATIONS.* 1998. Jon Dunn. $200,000. This grant provided equipment including two IBM RS/6000 F50 AIX servers and an RS/6000 43P workstation, and IBM NT workstations. These systems formed the core hardware for continued development of the VARIATIONS digital music library program and other digital library efforts at IU.
- *Linear System Analyzer*. 1998. Randall Bramley, Craig Stewart, Christopher Peebles, D. F. (Rick) McMullen, Zdzislaw Meglicki, Juan Villacis, Thomas Stuckey. $1,250,000. This grant and an additional $670,000 purchase by IU resulted in an upgrade of our IBM RS/6000 SP to include 46 new 160MHz P2SC thin nodes and a switch router.
- *Infrastructure for deep computing: High Performance Computing component technologies and transparent, unlimited I/O in support of data-intensive research.* 1999. Randall Bramley, Michael A. McRobbie, Craig A. Stewart, Zdzislaw Meglicki, D.F. (Rick) McMullen, Gerry Bernbom. $1,000,000. This grant, together with an additional purchase of $1,032,000 by IU, added two new frames to our IBM RS/6000 SP. One frame was fully filled with 16 four-processor Power3+ Thin Nodes purchased under the Early Ship Program. The other frame included 13 two-processor Power3 Thin nodes.

IU has invested an additional $2M in IBM hardware to create the core of our massive data infrastructure, which includes a total of 2 TB in IBM spinning disk and 40 TB in IBM tape robots managed by HPSS.

As it stands today, IU's IBM RS/6000 SP includes a total of 139 processors for a peak theoretical capacity of 147 GFLOPS. It is currently the largest single supercomputer in the State of Indiana, one of the largest in the Midwest, and will hold an admirable position in the Top500 list when next revised in June. However, IU's relationship with IBM is based on much more than hardware. In the following sections we will outline some of the research and development activities and outcomes that have been facilitated by IBM SUR grants.

III. Component Architecture Research

III.a. Common Component Analysis Toolkit (CCAT) system development

A significant effort in the Computer Science Department at IU centers on the creation of software component methodologies for high-performance scientific computing. In spite of their great success in business and financial computing, component systems have not yet had a similar impact in computational science and engineering (CS&E). One reason for this is that CS&E computing involves objects that are under rapid development and change - making the mechanisms used in CORBA (IDL definition, stub and skeleton generation, filling-in of application-level codes to satisfy those stubs and skeletons) awkward and unwieldy. The primary reason component systems have not penetrated the CS&E world, however, is the need for high performance, end-to-end, in a component-based system.

To address these problems, IU Computer Science has been actively participating in the Common Component Architecture Forum, a group of Universities and national laboratories which is defining a set of minimal standards that will allow CS&E components to be shared among various frameworks such as PetSc, SCIRUN, POOMA, PAWS, and CCAT. A significant accomplishment by the CCA Forum in the past year has been the first release of the CCA standard. IU's computer science group had the first running, fully distributed, multi-language implementation, called the Common Component Architecture Toolkit, which was demonstrated at SC99. The design of CCAT is unique; rather than considering components as belonging to a component framework (the way that Java Beans belong to a beanbox), the services which all component systems need - component location, instantiation, connection, instance registration, events - are themselves implemented as CCA-compliant components, which any component can choose to instantiate and use. This extends the peer model of component systems so that framework services have no hierarchical control over the applications components.

To provide a universally accessible mechanism for describing components and their instances, XML is used for both. A component's DTD includes the location of its installations, the purpose of the component, author-related information, and descriptions of the component's ports. A running instance of a component also includes a start point (global reference to the component, which allows other components to get a handle and invoke methods on it), its current parameter settings, and its current port connections. This also allows a researcher to email an XML description of a running component to a collaborator, and that collaborator to drop that representation into his own framework and start connecting other ports to it, invoking its methods, etc.

Fuller descriptions of the accomplishments are available in the form of several publications. [3] describes the work in the context of problem-solving environments, and [4] shows a detailed application area (sparse linear systems of equations) for which CCAT has been used. Two papers in HPDC [5,6] give more details about the CCAT architecture, and [7] describes the mapping of the CCAT to the computational Grid infrastructure. The preliminary version of CCAT has been used to support large, industrial mold-filling applications [8], demonstrated at SC98 (http://www.extreme.indiana.edu/sc98/sc98.html). The home page for the CCAT project is

http://www.extreme.indiana.edu/ccat/index.html. Other key publications regarding this project include [9] through [12].

Current work on CCAT includes building a multi-protocol version which allows components to dynamically select a communications protocol, a long-term resource prediction and management system that provides recommendations to users and components about what machines and networks should be used when building a CCAT component application, and higher-layer tools that allow applications scientists to more rapidly put their work into a CCA framework.

III.b. Remote grid-based instrument control

Cutting-edge research in many sciences is predicated on the use of large nationally and internationally shared instruments, ranging from electron microscopes to particle accelerators. These instruments tend to have several characteristics in common: they produce large streams of data, they are often inconvenient to use locally, and time on these instruments is precious to the researchers. For example, to use the Advanced Photon Source at Argonne National Laboratory, scientists must wait months for beam time, travel to ANL, and then spend many days on site collecting data. Once data are collected, days or weeks may be spent on reconstruction. Hence, time to information is months rather than hours, and the duty cycle of the instruments themselves is needlessly poor. IU is engaged in research to provide facilities for remote use of such instruments, including the Xport crystallography testbed and geophysics data collection and analysis.

The Xport crystallography testbed project is aimed at revolutionary improvements in telepresence for major scientific instrumentation systems. Our goal is to exploit a combination of advanced networking technology, new middleware services, and remote instrumentation technologies to achieve interactive "better-than-being-there" capabilities for remote experiment planning, instrument operation, data reconstruction, and data analysis. These capabilities will be deployed and demonstrated at two brilliant X-ray source facilities, the Advanced Photon Source at Argonne National Laboratory and the Advanced Light Source at Lawrence Berkeley Laboratory. This partnership also includes researchers in the Common Component Architecture Forum [5] and in the Molecular Biology Collaborative Access Team [13].

The Xport project has three major goals aimed at improving the quality of research done at large instrument facilities, particularly high brilliance X-ray crystallography for macromolecular structure determination:

- To extend the Globus APS crystallography experiment to support a national-scale remote instrument testbed for macromolecular crystallography.
- To develop testbed-specific components using the Globus and CCAT toolkits for overall experiment resource allocation and configuration, including computer resources, instrument control, on-the-fly data reduction after each scan, real-time data streaming to mass storage, and real-time visualization using desktop and immersive systems.
- To develop extensions to the CCAT and Globus to support persistent object storage and retrieval for large objects related to diffraction data collection and reduction, and performance monitoring of network connections (bandwidth and latency), and time-stamp trace capabilities at the component level to allow performance analysis of component-based applications.

Diffserv QoS-enabled data acquisition and HPSS storage components are currently under development to support the real-time acquisition of large (18MB) detector images directly to HPSS servers at IU and ANL. The Department of Energy's Emerge QoS testbed network provides bandwidth that can be allocated on a scheduled basis to connect the APS and ALS "data sources"

to grid computing and storage resources. HPSS plays a major role in Xport as the central storage mechanism for the large detector images acquired at the APS and ALS. IU's RS/6000 SP is the platform of choice for the computationally intensive image analysis and structure refinement components of the Xport system.

Professor Gary Pavlis of IU has been a leader in development of seismic imaging as a tool to better understand the structure of the earth's mantle and crust [14,15]. Such imaging is important for commercial concerns such as oil exploration and for safety and planning needs in terms of understanding and predicting earthquakes. IU is involved in an NSF-funded project called USArray [16], whose objective is to understand the evolution and deeper geological structure of the North American continent. The array is composed of a fixed array of about 100 seismometers and a mobile portion of about 2000 seismometers. The mobile array will move around the continent and continental shelf over a 10-year period to provide complete coverage of the US and parts of Canada and Mexico. This will revolutionize our understanding of the earth by improving the imaging of the earth's upper mantle by at least two orders of magnitude. Prof. Pavlis is also employing a 50-element broadband array in the Tien Shan of Krygyzstan and China, to better understand the origins of this enigmatic mountain belt. Geophysical data collection and manipulation is interesting and challenging for a number of reasons. Seismic sensors are geographically distributed. Each sensor continuously transmits ground motion information, which must be aggregated with data from all other sensors, analyzed, and stored. Although each sensor produces only a small amount of data, the expected thousands of sensors will produce a tremendous quantity of data. The analysis of such seismic data is challenging because it places heavy burdens simultaneously on computational speed, I/O, and management of large data stores. Indiana University is developing new network protocols for real-time transmission and aggregation of large amounts of data using a grid computing and storage model. Furthermore, IU's efforts in massive data storage (see below) will greatly facilitate the I/O required for effective analysis of these data. Accomplishments to date include the development of software for storing and retrieving events to HPSS and several software packages (for monitoring data from an array in real time and analyzing stored data).

IV. Massive Data Storage

Massive data storage represents an important area of joint interest for IU and IBM [17]. During the past 16 months (Jan 1999 thru Apr 2000), IU has successfully implemented the basic elements of its high-capacity, distributed storage infrastructure: HPSS server with robotic tape storage and fast disk cache at IU-Bloomington, HPSS/DFS interface, HPSS remote mover with fast disk cache at IUPUI (Indianapolis campus), and the HPSS/GPFS "test" system at IU-Bloomington [18,19]. In May 2000 IU brought into service robotic tape storage (IBM 3494) for the remote HPSS mover at IUPUI, completing the basic infrastructure.

By combining massive data storage with major increases in supercomputing horsepower and high performance networking, IU is providing an unparalleled information technology infrastructure that enables IU scientists and scholars to move to new levels of research. IU's massive data storage implementation is distinctive in several respects:

- IU is the only site in the US with a single HPSS system distributed over a long-haul, wide-area network.
- IU is the only site running the HPSS/DFS interface and offering this as a primary user access method for massive data storage.
- IU is the only university site in the US using HPSS to provide massive data storage as a "baseline" service for all university researchers in all disciplines.

To take just one research initiative as an example, this massive data storage infrastructure plays an important role in IU's participation as a major collaborator in designing and building the next generation high-energy physics "data grid" that will be used by the ATLAS experiment at CERN and other high-energy physics experiments. Professor Robert Gardner of Indiana University's physics department was assigned a lead role to represent the US ATLAS team in this national endeavor (referred to as GriPhyN, the Grid Physics Network) [20]. Professors Randall Bramley and Dennis Gannon of IU's computer science department are co-investigators on this project. The massive data storage infrastructure is also an important element of several other research projects including those involving seismology, x-ray crystallography, and digital libraries cited elsewhere in this report.

IU technical and engineering staff have worked with IBM staff during the past year on development of a DMAPI-based interface between GPFS and HPSS, to permit transparent high-speed parallel data transfer from compute nodes to GPFS and from GPFS to HPSS. The HPSS "test" system will be a platform for trial and evaluation of this software, beginning in May 2000.

V. Digital Libraries and the VARIATIONS project

VARIATIONS is a digital library project which provides access to over 5500 titles of near CD-quality digital audio to users at computer workstations in the School of Music at Indiana University Bloomington [2, 21]. VARIATIONS serves as an essential component of music education at IU and as a testbed for multimedia digital libraries at Indiana University. This system first became operational in early 1996. The current IU-IBM SUR collaboration centers on several topics:

- Scaling up of VARIATIONS in terms of number of concurrent users, number of potential authorized users, and the size of the collection. IU adds up to eighty hours, or thirteen gigabytes, of compressed audio per week to the system. The current data stored totals 1.5 TB. The number of workstations able to access the VARIATIONS program now totals 150, all within the confines of the School of Music Library.
- Investigating the possibility of extending the VARIATIONS service to sites at IU outside the Music Library and to sites at other institutions.
- Investigating the use of IBM Digital Library (DL) and its accompanying products as a platform for a scaled-up version of VARIATIONS and for other digital library applications at IU.

Directly related to the scaling issues is a proposal that IU submitted in May 1999 to the NSF's Digital Library Initiatives round 2 (DLI-2) grant program. This proposal, accompanied by a letter of support from Peter Mandel of IBM Content Management, outlines the creation of a digital music library testbed at IU with access from seven satellite institutions located in the United States, Europe, and Japan, connected via Internet2/Abilene, TransPAC, and other advanced networks. If funded, this work will explore issues in the areas of network quality of service, the use of digital libraries in music pedagogy, the creation of middleware and metadata architectures to support distributed content and applications, user interface design, and intellectual property rights.

Since installing the SUR-granted equipment, IU has purchased licenses for IBM Digital Library version 2, including its Library Server, Object Server, Text Search Manager, Internet Connection, and Media Manager components. In addition, IU purchased IBM DL Collection Treasury, a set of software tools designed to assist in loading and providing access to digitized library or museum special collections. IU also has continued to work with VideoCharger, versions 1.1.1 and 2, serving as a beta test site for both the original VideoCharger 2 release and more recently for QuickTime support. In addition, IU has served as a trial customer for IBM's Z39.50 service

offerings for DL and has assisted in uncovering several bugs that were subsequently corrected by IBM. IU hosted a weeklong visit by Rich Dempsey and Tammy Parker of IBM in April 1999 for testing of IBM's Z39.50 support against various library automation software systems.

IU is now using IBM Digital Library 2.4.1 and the Digital Library Collection Treasury (DLCT). The first major project undertaken was to use components of DLCT to place online an historical photograph collection, the Frank M. Hohenberger Collection from IU's Lilly Library [22]. As part of this process, IU staff informed IBM of several limitations in DLCT subcomponents that make it difficult for collections that are more complex than a single image per item. The second major project to be implemented using IBM DL is the Hoagy Carmichael Collection [23, 24] consisting of digitized items from the songwriter's personal collection in a wide variety of formats. These formats include correspondence, photographs, manuscript music, typescripts, personal effects, and sound recordings. IU is developing its own Java-based loader and Java servlet and JSP-based Web access for the Charmichael Project as a result of the limitations encountered in DLCT and Net.Data / Dynamic Page Builder. Through work on this project, IU has assisted IBM in identifying a number of previously undiscovered bugs in DL for which we hope fixes will eventually be made available by IBM.

In February 2000, IU installed Enterprise Information Portal (EIP) 6.1 and a pre-release copy of IBM Content Manager (CM) 6.1, which are follow-on products to IBM Digital Library 2.4. The final version of IBM Content Manager 6.1 was installed in March. IU's Java-based access scripts and loader developed for the Hoagy Carmichael Collection have been successfully modified to work with CM/EIP. In addition, CM and EIP are being used as the basis for a prototype of a preprint / working paper archive to be used by scholars studying and writing about common pool resources, in conjunction with the Workshop in Political Theory and Policy Analysis, which is based at IU.

In Spring of 2000, we began storing newly-digitized VARIATIONS sound files in the HPSS-based massive data storage system and started a process of migrating sound files currently stored in ADSM to HPSS (approximately 6 TB of content including both compressed MPEG and uncompressed WAV files). This summer we plan to develop and test on-demand sound file retrieval into VideoCharger from HPSS, with the eventual goal of moving from ADSM to HPSS for digital library storage needs.

VI. Other Accomplishments

VI.a. Bioinformatics

The availability of large data sets of genome data make possible the use of statistical techniques for inferring evolutionary relationships among genes, gene products, organelles, and organisms. These statistical techniques include maximum likelihood methods for constructing evolutionary phylogenies. One of the more popular programs for performing maximum likelihood analysis is fastDNAml, a program produced by Olsen et al. [25]. With data sets that include hundreds of thousands of nucleotides, such comparisons can involve significant computations, and a grid-based approach to attacking these analyses has proven extremely valuable. Indiana University has ported fastDNAml to the RS/6000 SP using MPI as a communications library, and has modified the program for use on nationally and internationally distributed grids. IU has also created a facility for real-time viewing of the analysis as it progresses. The grid-enabled version of fastDNAml has been demonstrated at the 1998 IEEE SuperComputing conference and presented at IBM's CASCON 99 [26]. We are engaging in a partnership with fastDNAml author Gary Olsen of the University of Illinois, to continue enhancing and optimizing fastDNAml, particularly for the RS/6000 SP, and we will be distributing this code freely under the GNU General Public License.

VI.b. Environmental Sciences

The IU School of Public and Environmental Affairs GIS Lab is simultaneously engaged in research and public service using aerial imagery, satellite imagery, and other GIS data [27]. IU's RS/6000 SP is used to process and compress USGS digital orthophoto quadrangle (DOQ) aerial images. This makes it possible to deliver these very large images over the Web to researchers and the public [28]. The GIS lab is well on its way to making the DOQs for the entire State of Indiana available via the Web. The US Forest Service uses the DOQs for the annual Forest Health Monitoring and Inventory Assessment programs. The Health Monitoring program is a particularly high priority due to the advance of the gypsy moth in Indiana. Indiana's State Firefighting Coordinator recently used a DOQ, compressed on IU's RS/6000 SP and downloaded to his laptop via the Web, in planning strategy for extinguishing a forest fire in central Indiana. The availability of this data via the Web democratizes access to the data - making it as easily available to farmers and citizens concerned about the environment as it is to the State's researchers.

VI.c. Best practices in High Performance Computing

Many challenges face the HPC community today: demonstrating the societal value of the technology, leveraging IT investments both for advances in research and for economic growth and development, coping with and taking full advantage of the ongoing upheavals in the HPC industry, providing support for traditional HPC users, and reaching out to broaden the impact of HPC as a tool for advancing the state of human knowledge. IU has recently announced the $30-million Indiana Pervasive Computing Research (IPCRES) initiative, the chief goal of which is to build an advanced IT research capability in Indiana and use this as a means of state economic development [29]. IU is a member of the Coalition of Advanced Scientific Computing centers (CASC), which among other goals seeks to educate the US government on the value of HPC [30]. IU has also expressed its commitment to both traditional and new users of HPC through its Information Technology Strategic Plan [31], which outlines this strategy:

- "Provide the infrastructure needed by those disciplines that are already heavily invested in information technology and need to keep up and expand;
- "Encourage the exploration of new technology in disciplines where the cutting edge of research is intimately tied to advances in computing;
- "Promote the introduction of new technology in disciplines which would greatly benefit from it, but which may have previously not had access to, or made use of, information technology."

Indiana University is taking a direct approach to assuring that our investment in high performance computing and communications delivers benefit to IU researchers, IU students, and the residents of the State of Indiana whose taxes help support IU. IU takes a very inclusive view of our supercomputer user community. We actively recruit researchers in all disciplines - including the humanities and the arts - to make use of our RS/6000 SP. Users of this system range from beginning graduate students in introductory research methods classes to IU's most sophisticated research scientists. IU's RS/6000 SP is home to over 50 third-party software applications, as well as the most advanced grid middleware (including Globus and Legion). IU's RS/6000 SP is used by hundreds of its graduate students and faculty members. This is made possible by the robust reliability of the RS/6000 SP, its outstanding uniprocessor performance, and its superior parallel computing environment and facilities.

This broad use of the RS/6000 SP creates certain support challenges. Rather than serving a user community of a handful of advanced scientists in computationally advanced disciplines, our system serves hundreds of users with a varied experiences and backgrounds. As a result, IU has expanded its pathbreaking accomplishments in leveraged support [32] to include the delivery of

excellent online resources in support of the RS/6000 SP [33]. This makes it possible to focus our valuable human resources in two particular areas: in-depth work with advanced computational scientists, and outreach to scientists who could benefit from the application of parallel programming methodologies. The fact that the RS/6000 SP is the primary platform for all research computing at IU greatly facilitates the broader adoption of parallel programming methodologies. The general familiarity with the RS/6000 SP greatly eases the learning curve for parallel programming in this environment. This expanded use of IU's high performance computing resources has also enhanced the diversity of our user base (in particular increased numbers of women), thus helping us address the general lack of diversity in the HPC community today.

VI.d. Researcher-to-researcher relationships

The key value added for IBM and for IU as a result of our partnership is the collaboration and intellectual exchange that takes place between researchers of the two organizations. Success in this area has been a special strong point in the past two years. IU has made two visits to IBM's research laboratories in New York - both giving and receiving presentations about the newest technical developments in high performance computing. IU's technology has been featured in IBM displays at the ACM/IEEE SCxy Conferences, as IBM technology has been featured in IU's presentations. But most importantly, we have established a broad array of researcher-to-researcher collaborations. These are most notably strong between Randall Bramley's Laboratory and John Levesque's group. Strong relationships also exist with IBM groups or individual researchers in the following areas: mathematical functions, HPSS and massive data storage groups, and bioinformatics. Mutual interest has been established in the following areas, which may well lead to additional collaborations: GRAPE-based simulations, intelligent agents, pervasive computing, and Java-based tools. This broad engagement of collaborations across multiple areas of intellectual achievement creates a remarkable and highly productive relationship between IU and IBM.

VII. Selected citations

1. Davidson, E. "MELD: A Many Electron Description." Pp. 553-592 In: Modern Techniques in Computational Chemistry: MOTECCTM-90. E. Clementi, E. ESCOM Science Publishers, Leiden, 1990.

2. Indiana University Music Library. "VARIATIONS." 1995 http://www.dlib.indiana.edu/variations/ (4 June 2000)

3. Houstis, E. N., J. R. Rice, E. Gallopoulos, and R. Bramley, Enabling Technologies for Computational Science, Kluwer, Boston, 2000.

4. Gannon, D., R. Bramley, T. Stuckey, J. Villacis, J. Balasubramanian, E. Akman, F. Breg, S. Diwan, and M. Govindaraju, Enabling Technologies for Computational Science, Kluwer, Boston, 2000, Chapter 10: The Linear System Analyzer, pp. 123-134.

5. Villacis, J., M. Govindaraju, D. Stern, A. Whitaker, F. Breg, P. Deuskar, B. Temko, D. Gannon, and R. Bramley, "CAT: A high performance distributed component architecture toolkit for the GRID," in Proceedings of High Performance Distributed Computing Conference, 1999.

6. Bramley, R., K. Chiu, S. Diwan, D. Gannon, M. Govindaraju, N. Mukhi, B. Temko, and M. Yechuri, "A component based services architecture for building distributed applications," accepted for High Performance Distributed Computing Conference, 2000.

7. Bramley, R., D. Gannon, and J. Villacis, "Using the GRID to support software component systems," in Proceedings of SIAM Parallel Processing Conference, San Antonio TX, 1999, available on CD-ROM from SIAM Publishing, Philadelphia.

8. Ilinca, F., J.-F. Hetu, M. Audet, and R. Bramley, "Simulation of 3D mold-filling and solidification processes on distributed memory parallel architectures," in Proceedings of ASME International Mechanical Engineering Congress, 1998.

9. Breg, F., S. Diwan, J. Villacis, J. Balasubramanian, E. Akman, and D. Gannon, "Java RMI performance and object model interoperability: Experiments with Java/HPC++," Concurrency and Experience, (1998). Presented at 1998 ACM Workshop on Java for High-Performance Network Computing.

10. Gannon, D., R. Bramley, T. Stuckey, J. Villacis, J. Balasubramanian, E. Akman, F. Breg, S. Diwan, and M. Govindaraju, "Component architectures for distributed scientific problem solving." IEEE Computational Science and Engineering Magazine, 5(2):50-63, 1998.

11. Houstis, E., E. Gallopoulos, R. Bramley, and J. Rice, "Problem-Solving Environments for Computational Science," IEEE Computational Science and Engineering, 4 (1997), pp. 18-21.

12. Bramley, R., B. Char, D. Gannon, T. T. Hewett, C. Johnson, and J. R. Rice, "Enabling Technologies for Computational Science," Kluwer, Boston, 2000, Chapter 2: Workshop on Scientific Knowledge, Information, and Computing, pp. 19-32.

13. von Laszewski, G., Westbrook, M., Foster, I., Westbrook, E. and Barnes, C. "Using Computational Grid Capabilities to Enhance the Capability of an X-Ray Source for Structural Biology." To be published, IEEE Cluster Computing. http://www.mcs.anl.gov/~laszewsk/papers/laszewskiClusterComputing.ps (4 June 2000)

14. Neal, S.L. and Pavlis, G.L. "Imaging P to S Conversions With Multichannel Receiver Functions." Geophys. Res. Letters, 26, 2581, 1999.

15. Bear, L.K. and Pavlis, G.L. "Estimation of Slowness Vectors and Their Uncertainties Using Multiwavelet Seismic Array Processing." Bull. Seism. Soc. Amer., 87, 755, 1997.

16. Göran Ekström. "USArray." http://www.seismology.harvard.edu/projects/USArray/ 1998. (4 June 2000)

17. Distributed Storage Services Group. http://www.indiana.edu/~dssg/ (4 June 2000)

18. Meglicki, G. In review. "Towards Distributed Data Banks." Submitted to SC2000.

19. Meglicki, G. In review. "HPSS Data Migration Application for GPFS." Submitted to SC2000.

20. Gardner, Rob. "ATLAS GriPhyN Testbed." http://lexus.physics.indiana.edu/~rwg/griphyn/indiana.html (4 June 2000)

21. Dunn, J.W., and Mayer, C.A. "VARIATIONS: A Digital Music Library System at Indiana University." In DL '99: Proceedings of the Fourth ACM Conference on Digital Libraries, Berkeley, California, August 1999. Copy available at http://www.dlib.indiana.edu/variations/VARIATIONS-DL99.pdf. (4 June 2000)

22. Frank M. Hohenberger Photograph Collection. 1998 http://www.dlib.indiana.edu/collections/lilly/hohenberger/ (4 June 2000)

23. The Hoagy Carmichael Collection. 1998 http://www.dlib.indiana.edu/collections/hoagy/ (4 June 2000)

24. Brancolini, K.R., Dunn, J.W., and Walsh, J.A. "Digital Star Dust: The Hoagy Carmichael Collection at Indiana University." First Monday, Volume 5, Number 6, http://www.firstmonday.org/ (4 June 2000)

25. Olsen, G. J., Matsuda, H., Hagstrom, R., and Overbeek, R. "fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood." Comput. Appl. Biosci. 10: 41-48. 1994.

26. Stewart, C., et al. "Evolutionary biology and computational grids." http://www.indiana.edu/~rac/cascon/cascon.html (4 June 2000)

27. SPEA Geographic Information Systems. 4 June http://www.indiana.edu/~speagis/ (4 June 2000)

28. SPEA GIS. 2000 http://ovpit.ucs.indiana.edu/SPEAGIS/ (4 June 2000)

29. IPCRES: Indiana Pervasive Computing Research Initiative. 1999 http://www.indiana.edu/~ovpit/ipcres/ (4 June 2000)

30. Coalition for Advanced Scientific Computing. http://www.ncsc.org/casc/ (4 June 2000)

31. Information Technology Strategic Plan: Architecture for the 21st Century. 1998 http://www.indiana.edu/~ovpit/strategic/ (4 June 2000)

32. Voss, B. 1998 "IWIS Project." http://www.indiana.edu/~ucs/tlit/cni/ (4 June 2000)

33. The UITS Research SP system. http://sp-www.iu.edu/ (4 June 2000)