Acquisition of a High-Speed, High Capacity Storage System to Support Scientific Computing: The Data Capacitor Final Report. 2009

Final Report

NSF Award ID: ACI-0521433
Project Dates: 10/2007 to 09/2008
Principal Investigator: Stewart, C.A.
Co-Investigators: Simms, S., Pilachowski, C., Bramley, R.
Organization: Indiana University

Table of Contents

Final Report for Period: 10/2007 - 09/2008
Submitted on: 12/01/2008
Principal Investigator: Stewart, Craig A.
Award ID: 0521433
Organization: Indiana University
Submitted By:
Stewart, Craig - Principal Investigator
Title:
MRI: Acquisition of a High-Speed, High Capacity Storage System to Support Scientific Computing: The Data Capacitor

Project Participants

Senior Personnel

Name: Stewart, Craig
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Bramley, Randall
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Plale, Beth
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Hacker, Thomas
Worked for more than 160 Hours: No
Contribution to Project:
Thomas Hacker left Indiana University in Spring of 2006. We have requested that Thomas Hacker be removed from this project as a Co-PI and that Stephen Simms replace him as Co-PI.

Name: Pilachowski, Catherine
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Simms, Stephen
Worked for more than 160 Hours: Yes
Contribution to Project:

Other Collaborators or Contacts

Both Doug Balog from Pittsburgh Supercomputing Center and Greg Pike from Oak Ridge National Laboratory have been instrumental in IU's pursuit of making Lustre a viable WAN filesystem.

Matthias Mueller and his staff at the Center for Information Servies and High Performance Computing at Technische Universitaet Dresden (Germany) were insturmental in IU's work on Lustre-WAN over liong distances, including the IU/Dresden win in the Bandwidth Challenge at SC2007.

Activities and Findings

Research and Education Activities: (See PDF version submitted by PI at the end of the report)

See attached MS word file containing a final report.

Findings: (See PDF version submitted by PI at the end of the report)

See attached MS word file containing a final report.

Training and Development:

See attached MS word final report.

Outreach Activities:

See attached text file.

Journal Publications

Randall Bramley, Kenneth Chiu, Tharaka Devadithya, Nisha Gupta,
Charles Hart, John C. Huffman, Kianosh Huffman, Yu Ma, and Donald F. McMullen, "Instrument Monitoring, Data Sharing, and Archiving Using the
Common Instrument Middleware Architecture (CIMA)", Journal of Chemical Information and Modeling, p. 0, vol. March, (2006). Accepted,

Felipe Bertrand, Yongquan Yuan, Kenneth Chiu, Randall Bramley, "An Approach to Parallel MxN Communication", International Journal of High Performance
Computing Applications, p. 399, vol. 19, (2005). Published,

Douglass Post, Donald Batchelor, Randall Bramley, John R. Cary, Ronald
Cohen, Phillip Colella, Stephen Jardin, "Report of the Fusion Simulation Project Steering Committee", Journal of Fusion Energy, p. 1, vol. 23, (2005). Published,

Books or Other One-time Publications

Yu Ma and Randall Bramley, "A Composable Data Management Architecture
for Scientific Applications", (2005). conference paper, Published
Collection: Proceedings of Challenges of
Large Applications in Distributed Environments
Bibliography: Proceedings of Challenges of Large Applications in Distributed Environments (CLADE) Workshop, 2005.

Randall Bramley, Rob Armstrong, Lois McInnes, "High-Performance Component Software Systems", (2006). Book, Published
Editor(s): M. A. Heroux, P. Raghavan and H. D. Simon
Collection: SIAM series: Software, Environments, and Tools
Bibliography: Parallel Processing for Scientific Computing, Edited by Michael A. Heroux, Padma Raghavan, and Horst D. Simon 2006.

Felipe Bertrand, Randall Bramley, Kostadin B. Damevski, James A. Kohl, David E. Bernholdt, Jay W. Larson, and Alan Sussman, "Data Redistribution and Remote Method Invocation in Parallel Component Architectures", (2005). Book, Published
Collection: International Parallel and Distributed Processing Symposium
Bibliography: Felipe Bertrand, Randall Bramley, Alan Sussman, David E. Bernholdt, James Arthur Kohl, Jay W. Larson, Kostadin Damevski: Data Redistribution and Remote Method Invocation in Paralle

Gannon, Dennis, Beth Plale, marcus Christie, Yi Huang, Scott Jensen, Ning Liu, Suresh Marru, Sangmi Lee Pallickara, Srinath Perera, Satoshi Shirasunam Yogesh Simmhan, Aleksander Slominski, Yiming Sun, Nithya Vijayakumar., "Building Grid Portals for e-Science: A Services Oriented Architecture.", (2007). one-time publication, Accepted
Editor(s): Lucio Grandinetti
Bibliography: IOS Press, Amsterdam.

Plale, Beth., "Workload Characterization and Analysis of Storage and Bandwidth Neeeds of LEAD Workspace.", (2007). LEAD -technical paper, Published
Bibliography: LEAD TR001 V3.0

Simms, Stephen, Greogry Pike, Douglas Balog., "Wide Area Filesystem Performance using Lustre on the TeraGrid.", (2007). Proceedings, Published
Bibliography: Madison, WI.

Simms, Stephen, Scott Teige, Bret Hammond, Yu Ma, C. Westneat, Larry Simms, Douglas Balog., "Empowering Distributed Workflow with the Data Capacitor: Maximizing Lustre Performance Across the Wide Area Network.", (2007). Proceedings, Published
Bibliography: Monterey, CA.

Simms, Stephen, Matthew davy, Brett Hammon, Matt Lind, Craig Stewart, Randall Bramley, Beth Plale, Dennis Gannon, M-H Baik, Scott Teige, John Huffman, Donald McMullen, D. Balog and G. Pike., "All in a Day's Work: Advancing Data-Intensive Research with the Data Capacitor.", (2006). Proceedings, Published
Bibliography: ACM Press, New York, NY. Tamapa, FL, pg. 244 (http://doi.acm.org/10.1145/1188455.1188711)

Web/Internet Site

URL(s):
http://rac.uits.indiana.edu/spotlight.shtml
http://datacapacitor.org/
Description:

Contributions

Contributions within Discipline:
Progress 10/1/06 - 9/30/07

See text in uploaded report.

Contributions to Resources for Research and Education:

Progress 10/1/06 - 9/30/07

See text in uploaded report.

Contributions Beyond Science and Engineering:

Categories for which nothing is reported:

Organizational Partners

Any Product

Contributions: To Any Other Disciplines

Contributions: To Any Human Resource Development

Contributions: To Any Beyond Science and Engineering

Activities and Findings

The Data Capacitor facility for the short to mid-term storage of research data sets was purchased after a thorough evaluation of several vendor proposed solutions. After a period of rigorous testing, Indiana University formally accepted the Data Capacitor on October 6, 2006 and successfully brought the Data Capacitor into full production mode meeting a self-imposed deadline of April 1, 2007. At inception the Data Capacitor operated with capacity and capabilities greater than those outlined in the original MRI proposal to the NSF. At that time the Data Capacitor comprised:

- 52 Dell servers running Red Hat Enterprise Linux
- 24 Myricom 10 gigabit Ethernet cards
- 12 DataDirect Networks S2A9550 storage controllers
- 30 DataDirect Networks 48 bay SATA disk chassis
- 535 terabytes usable disk storage using the Lustre filesystem

The Data Capacitor is housed in six water-cooled Rittal racks that provide high quality controlled environments for the system equipment. At acceptance, the Data Capacitor demonstrated a measured 14.5 gigabyte per second aggregate transfer rate. The Data Capacitor is providing unprecedented capabilities in the rapid I/O of very large data files locally, and via high-speed networks throughout the US. As a result, the Data Capacitor is changing the way IU computer and discipline scientists think about and go about their research.

We have implemented high speed connections between the Data Capacitor and the following other systems and labs at IU:

- IU's HPSS mover nodes, which serve as a front end to IU's 2.8 petabyte massive data storage system
- IU's 30.7 TFLOPS Big Red supercomputer
- IU's 8.9 TFLOPS Quarry supercomputer
- A pool of GridFTP servers serving as data transfer nodes to resources located locally at IU and throughout the nation
- We have, as described in the original proposal, installed dedicated 10 Gigabit Ethernet links to IU's Computer Science, Astronomy, and Chemistry departments via a dedicated 10Gb fiber where both data acquisition and compute servers mount the filesystem.

Today, the Data Capacitor currently serves both local and national users by providing a large pool of high-speed storage to multiple distributed resources. While our major operational focus to date has been on serving local users, the Data Capacitor has generated considerable interest in the national research community. A recent survey of Data Capacitor users shows that 87.5% of respondents consider the Data Capacitor to

be important to their research activities. An unintentional, but highly advantageous result of the Data Capacitor project has been extensive development of Lustre as a Wide Area Filesystem. In 2006, IU performed experiments with the Pittsburgh Supercomputing Center demonstrating the possibilities of mounting the Lustre filesystem across great distance. In the past year the experiments have become stable with IU's development of a lightweight UID mapping scheme to keep filesystem permissions consistent across multiple resources. This breakthrough and the success of Data Capacitor project in general motivated IU to spend some of its own money to double the size of the Data Capacitor, expanding it to a petabyte of spinning disk. In April, IU dedicated roughly 380 TB of this new storage to create a full-blown wide area production service to assist projects in both the sciences and humanities.

Our efforts this year have focused on supporting our current users while expanding the number of remote users and sites that currently mount the Data Capacitor. We administered a survey to users of the Data Capacitor to get some feedback about how the project has been going and were pleased with the results.

On a scale where 1 is the least satisfied and 5 is the most satisfied; our users rated the Data Capacitor service a 4.38. The only negative complaint was about downtime. During the summer, Indiana suffered some power outages due to storms. While users did not lose data, they did lose some valuable compute time. Since those outages, IU has installed multiple generators to insure that the filesystem will remain up during power events. Users would like to see other TeraGrid sites mount the Data Capacitor. Our survey showed that over 16% of Data Capacitor users are either using the system from a remote mount point or using multiple distributed mount points. Those users would like to see the functionality we're providing at other large sites. Currently we're mounted in production at University of Kansas, University of Michigan, Mississippi State University, Technische Universitaet Dresden, Rochester Institute of Technology, and Pittsburgh Supercomputing Center.

On a scale where 1 is the least satisfied and 5 is the most satisfied; our users rated the Data Capacitor support and consulting a 4.71. We're very pleased and proud that that no one expressed any dissatisfaction with the support we have been providing for the service.

Contributions to Resources for Research and Education

The Indiana University Data Capacitor helps scientists across many disciplines meet the critical challenge of storing and manipulating massive data sets. Using the Data Capacitor, a team led by Indiana University was awarded first place in the SC07 Bandwidth Challenge, an annual competition that invites teams of technologists from the nation's most elite supercomputing facilities to push the limits of modern computer networks. The IU-led team demonstrated a bidirectional transfer rate of 18.2 Gigabits

per second (Gbps) out of a possible 20 Gbps, showing the Data Capacitor's tremendous capabilities for supporting data-intensive research across wide area networks.

As part of the Bandwidth Challenge project, the IU team and partners simultaneously supported the following data–centric scientific computing workflows:

- Modeling and analysis of the amyloid peptide thought to be the cause of Alzheimer's disease through the use IU's Big Red Supercomputer.
- Live acquisition of x-ray crystallography data.
- Digital preservation of ancient Sanskrit manuscripts.
- Performance analysis of a computational fluid dynamics application by the using the Vampir/VampirTrace software package.
- Simulations of high energy physics reactions.

A critical success in this demonstration was using Lustre-WAN across commodity networks moving data between Reno NV, Bloomington IN, Rochester NY, and Dresden Germany. Often today's demonstrations are tomorrow's production systems. In the case of Lustre-WAN, the Data Capacitor, and IU's Big Red Supercomputer, today's demonstrations are a reflection of today's production systems.

The work that IU has done across the WAN has inspired others to try the same strategy. For example, the University of Florida has been able to produce a Lustre storage system that serves their Tier-2 center, their Quantum Theory Project, and Florida International University simultaneously. The first two projects are served by their internal 10 Gb research network, not unlike the way the Data Capacitor is currently being used by Chemistry, Astronomy, and Computer Science at IU. The third is being served Lustre across the Florida Lambda Rail, following the work IU has been doing across the WAN. It's wonderful that our work has been useful to them in this way.

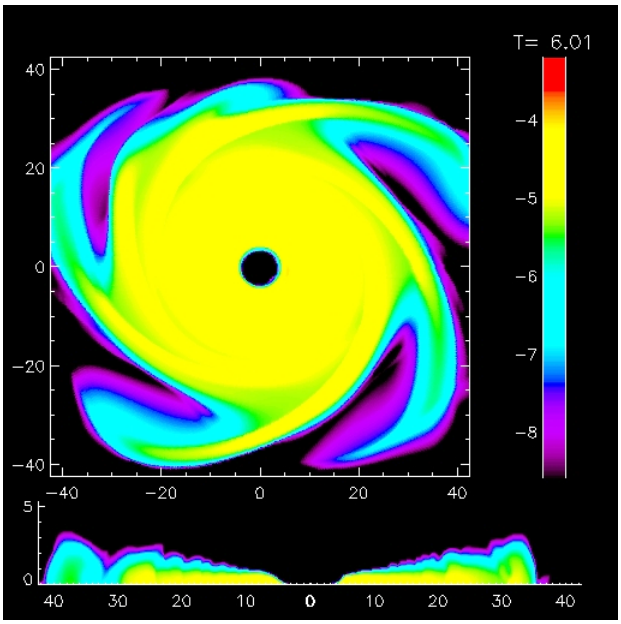
Data-Intensive Research

What makes Indiana University's cyberinfrastructure exciting is the many ways in which it is used for research. Researchers representing a wide range of fields are using and are making plans to use the Data Capacitor to accomplish their goals. Below are some projects making use of the Data Capacitor:

Origin of Gas Giants

IU astronomer Scott Michael has been working with Dr. Richard Durisen to understand the origin of gas giant planets. Their computer models run fastest and scale best on shared memory supercomputers, however, IU does not currently own a large shared memory machine. Instead of running their models locally, they have been running on the Altix 4700 machines at PSC and Technische Universitaet Dresden where we have mounted the Data Capacitor across great distance. The second part of Scott's workflow requires the IDL software package which the IU astronomy department has

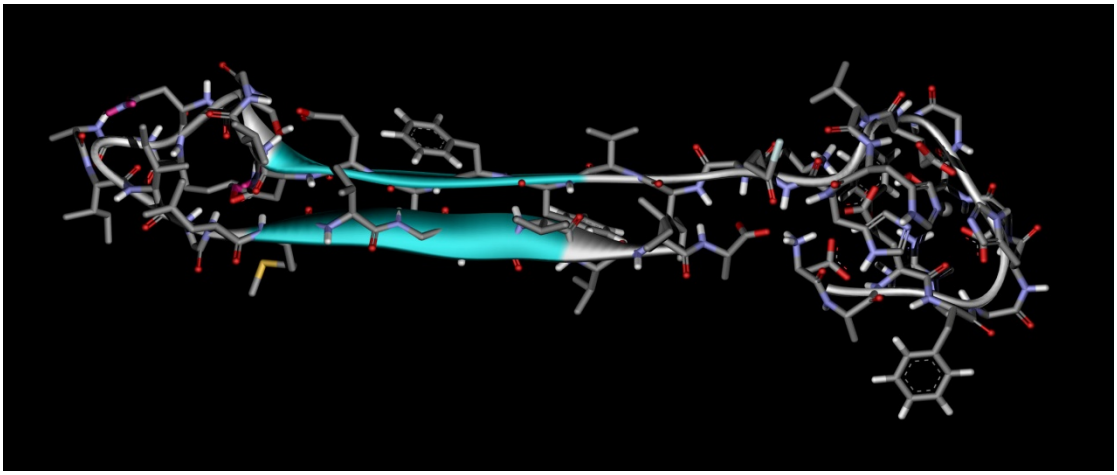
licensed and can't be run on Pople. So using the Data Capacitor, Scott is able to visualize his results in his office as they come from the simulation in real time. This real-time analysis allows real-time reactions and adjustments in response to run-time problems. Additionally preliminary analysis of the data allows for a faster setup time for subsequent simulations.



IU astronomer Scott Michael uses the Data Capacitor to manage and model data on the formation of gas giant planets.

Computational Chemistry in Alzheimer's Research

Living with Alzheimer's disease is commonly described as "going dark" in parts of the mind. Associate Professor Mu-Hyun Baik is using technology made available on the TeraGrid to do transformative research into Alzheimer's. His work concentrates on simulating the structure of the Amyloid-β protein, widely believed to be the cause of Alzheimer's disease. Only a better understanding of how Amyloid-β behaves chemically in the brain will lead to treatment and cure of the disease. This type of fundamental research requires enormous computational infrastructure and the ability to move a large amount of data quickly and easily, provided by IU's Big Red supercomputer and Data Capacitor respectively. Baik's research would not be possible without access to these TeraGrid resources.



The structure of Amyloid-β protein. The Data Capacitor was used to study the impact of changes in the protein in a search for ways to disrupt the creation of the fibrils that help cause Alzheimer's disease

Finite Element Methods in Three-Dimensional Flow

IUPUI Engineering School's Resat Payli is working to develop parallel software to use finite element and finite volume methods to solve three dimensional flow problems. Performing large-scale simulations with these codes and visualizing the obtained results requires a great deal of processing power, fast communication between the processors, and large storage systems.

Resat has been able to use the TeraGrid network and IU's Lustre WAN system to harness the power of two different geographically distributed systems to get his work done.

Like Scott Michael, Resat has been using the Data Capacitor's Lustre-WAN filesystem to tie IU systems and PSC together. Simulations are run on IU's BigRed cluster and the results are stored on the Data Capacitor which is also mounted on Pople, PSC's Altix supercomputer. Using 128 processors of Pople and the client/server mode of the open source ParaView visualization tool, simulation results get processed and rendered while the final images are sent to a client in Resat's laboratory.

Remote Sensing of Ice Sheets

The Center for the Remote Sensing of Ice Sheets (CReSIS) at the University of Kansas is a research group dedicated to the study, modeling, and analysis of the changes in the earth's many ice sheets. IU was approached by CReSIS to reanalyze a large chunk of instrument data that they had taken from the Greenland ice cap. It was their intention to copy their 20 Terabytes of data to USB disk drives and ship them to IU using a commercial carrier. However, by mounting the Data Capacitor across Internet2, researchers at University of Kansas were able to transfer their data faster than it would have taken them to copy the data to USB drives (let alone the cost and time that it would have taken to ship the drives).

Lustre across the WAN has become a reliable method for CReSIS to transfer data to IU for computation while functioning as a fast buffer to and from IU's HPSS system.

https://www.cresis.ku.edu/

EVIA Digital Archive Project

The EVIA project has been preserving hundreds of hours of videotaped ethnomusicological performances, with the intention of making them easily accessible for teaching and research.

Using the bandwidth capabilities of Internet2, IU has been able to mount the Data Capacitor at the University of Michigan where original material is digitized. This uncompressed video of performances is copied to the Data Capacitor where it can be examined, archived, and where compressed versions can be created for easy distribution.

http://www.indiana.edu/~eviada/

Preservation of Ancient Manuscripts

P.R. Mukund, Center for Preservation of Ancient Manuscripts (CPAM), Rochester Institute of Technology.

The Sarvamoola Granthas contain the philosophical teachings of Sri Madhvacharya, one of the most influential teachers in Hindu religious history. For centuries, the irreplaceable palm-leaf documents of the Sarvamoola Granthas were housed in Indian monasteries. With exposure to damaging atmospheric elements, over time the leaves became brittle, discolored and difficult to read. In the 1980s, with funding from the Smithsonian Institute, the documents were captured on microfilm, another medium susceptible to deterioration over time. Today, P.R. Mukund at the Center for Preservation of Ancient Manuscripts (CPAM) and his team are working to capture the document images in a digital format, which will better endure the passage of time and allows images from the damaged originals to be restored and once again made readable. The process of digitizing so many detailed images is no minor task, and resulted in massive amounts of data that the project team needed to move and manipulate easily. The team turned to the Data Capacitor, a system developed by Indiana University to temporarily store, transfer, and manipulate very large data sets.

Figure 1. Palm front images before and after digital restoration.

Books

- Gannon, Dennis, Beth Plale, Marcus Christie, Yi Huang, Scott Jensen, Ning Liu, Suresh Marru, Sangmi Lee Pallickara, Srinath Perera, Satoshi Shirasuna, Yogesh Simmhan, Aleksander Slominski, Yiming Sun, Nithya Vijayakumar, Building Grid Portals for e-Science: A Service Oriented Architecture To appear High Performance Computing and Grids in Action, IOS Press - Amsterdam, Lucio Grandinetti editor, 2007

- Plale, Beth. 2007. Workload Characterization and Analysis of Storage and Bandwidth Needs of LEAD Workspace, LEAD TR001 V3.0

- Stephen Simms, Gregory Pike, Douglas Balog, "Wide Area Filesystem Performance using Lustre on the TeraGrid." Proceedings of "TeraGrid '07 Conference", Madison, WI., 2007.

- Stephen Simms, S. Teige, Bret Hammond, Yu Ma, C. Westneat, Larry L. Simms, Douglas A. Balog, "Empowering Distributed Workflow with the Data Capacitor: Maximizing Lustre Performance across the Wide Area Network." Proceedings of "Workshop on Service-Oriented Computing Performance: Aspects, Issues, and Approaches", Monterey, CA, 2007.

- S.C. Simms, M. Davy, B. Hammond, M. Link, C. Stewart, R. Bramley, B. Plale, D. Gannon, M-H. Baik, S. Teige, J. Huffman, R. McMullen, D. Balog and G. Pike, All in a day's work: advancing data-intensive research with the data capacitor, SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing, ACM Press, New York, NY. Tampa, Florida, 2006, pp. 244. http://doi.acm.org/10.1145/1188455.1188711

- Yu Ma and Randall Bramley, "A Composable Data Management Architecture for Scientific Applications", (2005). conference paper, Published, Collection: Proceedings of Challenges of Large Applications in Distributed Environments

- Randall Bramley, Rob Armstrong, Lois McInnes, "High-Performance Component Software Systems", (2006). Book, Published, Editor(s): M. A. Heroux, P. Raghavan and H. D. Simon Collection: SIAM series: Software, Environments, and Tools

- Felipe Bertrand, Randall Bramley, Kostadin B. Damevski, James A. Kohl, David E. Bernholdt, Jay W. Larson, and Alan Sussman, "Data Redistribution and Remote Method Invocation in Parallel Component Architectures", (2005). Book, Published, Collection: International Parallel and Distributed Processing Symposium

Training and Development

WxChallenge. The Data Capacitor played an important role in the LEAD project in supporting the 2007 Spring WxChallenge (http://www.wxchallenge.com/), a national collegiate forecasting challenge. During the competition, over 1,000 students from 56 institutions used the LEAD Gateway to produce weather forecasts from meteorological data stored on the Data Capacitor. Terabytes of forecast and simulation data were written back to the Data Capacitor for visualization, analysis, and eventual storage on HPSS. Dr. Dennis Gannon and research scientist Suresh Marru supported this challenge.

Talks and Presentations

To promote the use and ideas underlying the creation of the Data Capacitor, invited talks and presentations were delivered to various national audiences. Delivered presentations are listed below:

- Stewart, C.A. 2006. IU's advanced cyberinfrastructure for life sciences computing. Computer Science Colloquium, Technische Universitaet Dresden, Dresden, Germany. April 22, 2006.

- Stewart, C.A. 2006. Advancing science with advanced cyberinfrastructure. IBM Briefing Center, San Diego, California.

- Simms, Stephen C. 2005. The Data Capacitor. SC|05, Seattle, Washington. November 18, 2005.

- Simms, Stephen C. 2006. The Data Capacitor Project and Digital Library Workflows. Digital Library Brown Bag, Bloomington, Indiana, February 22, 2006.

- Simms, Stephen C. 2006. The Data Capacitor Project. TeraGrid Site Visit, Indianapolis, Indiana, February 23, 2006.

- Simms, Stephen C. 2006. The Data Capacitor Project. ZIH Delegation Presentation, Indianapolis, Indiana March 13, 2006.

- Simms, Stephen C. 2006. The Data Capacitor: A High Speed, High Capacity, Storage System to Support Scientific Computing. Technische Universitaet Dresden, Dresden, Germany, April 7, 2006.
- Simms, Stephen C. 2006. The Data Capacitor Project. Lustre User Group Meeting, Hilton Head, South Carolina, April 19, 2006.
- Simms, Stephen C. 2006. Data Capacitor Demo. TeraGrid Æ06: Advancing Scientific Discovery, Indianapolis, Indiana, June 15, 2006.
- Simms, Stephen C. 2006. Lustre WAN Demo. TeraGrid Æ06: Advancing Scientific Discovery, Indianapolis, Indiana, June 15, 2006.
- Simms, Stephen C. 2006. The Data Capacitor. 9th HLRS Metacomputing and Grid Workshop, HLRS, Stuttgart, Germany, July 27, 2006.
- Simms, Stephen C. 2006. The Data Capacitor and Lustre WANdemonium. Lustre BOF, SC|06, Tampa, Florida, November 14, 2006.
- Simms, Stephen C. 2006. All in a Day's Work: Advancing Data-Intensive Research with the Data Capacitor. SC|06, Tampa, Florida, November 14, 2006.
- Simms, Stephen C. 2006. The Data Capacitor Project. SC|06, Tampa, Florida, November 15, 2006.
- Simms, Stephen C. 2006. All in a Day's Work: Advancing Data-Intensive Research with the Data Capacitor. SC|06, Tampa, Florida, November 15, 2006.
- Simms, Stephen C. 2006. An Introduction to Lustre and the Data Capacitor. SC|06, Tampa, Florida. November 16, 2006.

- Stewart, Craig A. 2006. On the road to petascale processing paradigms with IU's Big Red Supercomputer and IBM BladeCenter H. Indiana booth at SC|06, Tampa, FL. November 14 and 15, 2006.
- Simms, Stephen C. 2006. All in a Day's Work: Advancing Data-Intensive Research with the Data Capacitor. Indiana booth at SC|06, Tampa, FL. November 15, 2006.
- Simms, Stephen C. 2006. The Data Capacitor Project. Indiana booth at SC|06, Tampa, FL. November 15, 2006.
- "Adaptive weather forecasting using LEAD", B. Plale and D. Gannon, 2006 International Conference on High Performance Computing, Networking, Storage, and Analysis (SC06), November 2006.

- Simms, Stephen C. 2007. The Data Capacitor: Failover and Success. Data Search Institute Seminar, Bloomington, Indiana. February 6, 2007.
- Simms, Stephen C. 2007. The Data Capacitor Grant. Research Computing Grant Retreat, Mooresville, Indiana. March 28, 2007.
- Stewart, Craig A. 2007. Introduction to Research Computing. Ohio Supercomputer Center/ASC/MRSC. April 3, 2007.
- McCaulay, D. Scott. 2007. IU TeraGrid Quarterly Update Q1 2007. TeraGrid Quarterly Meeting, Breckenridge, CO. April 11, 2007.
- Simms, Stephen C. 2007. The Data Capacitor Project: Empowering Workflow and New Collaboration. Lustre User Group Meeting, Miami, Florida. April 24, 2007.
- Simms, Stephen C. 2007. Research Technology at IU. Biology Annotation Local Jamboree, Bloomington, Indiana, May 18, 2007.
- Simms, Stephen C. 2007. Wide Area Filesystem Performance Using Lustre on the TeraGrid. TeraGrid '07 Conference, Madison, WI. June 5, 2007.
- Simms, Stephen C. 2007. Wide Area Filesystem Performance using Lustre on the TeraGrid. Wright –Patterson Air Force Base, Dayton, Ohio. June 13, 2007.
- Stewart, Craig A. Implementation and experience with Big Red – a 20.4 TFLOPS IBM BladeCenter cluster. 2007. International Supercomputer Conference, Dresden Germany. June 26, 2007.
- Simms, Stephen C. 2007. Using the Data Capacitor for Remote Data Collection, Analysis, and Visualization. Indiana booth at SC|07, Reno, NV. November 14, 2007.
- Simms, Stephen C. 2008. IU Lustre WAN Update. Lustre User Group Meeting, Sonoma, California. April 29, 2008.

Journal Publications

- Randall Bramley, Kenneth Chiu, Tharaka Devadithya, Nisha Gupta, Charles Hart, John C. Huffman, Kianosh Huffman, Yu Ma, and Donald F. McMullen, "Instrument Monitoring, Data Sharing, and Archiving
- Using the Common Instrument Middleware Architecture (CIMA)", Journal of Chemical Information and Modeling, p. 0, vol. March, (2006). Accepted

- Felipe Bertrand, Yongquan Yuan, Kenneth Chiu, Randall Bramley, "An Approach to Parallel MxN Communication", International Journal of High Performance Computing Applications, p. 399, vol. 19, (2005). Published

- Douglass Post, Donald Batchelor, Randall Bramley, John R. Cary, Ronald Cohen, Phillip Colella, Stephen Jardin, "Report of the Fusion Simulation Project Steering Committee", Journal of Fusion Energy, p. 1, vol. 23, (2005). Published

Web Pages

http://datacapacitor.researchtechnologies.uits.iu.edu/
http://rtinfo.uits.iu.edu/cyberinfrastructure/resources.shtml
http://racinfo.indiana.edu/education_and_training/Cyberinfrastructure.pdf

Outreach Activities

We have engaged in three types of outreach activities: talks and presentations focused specifically on the Data Capacitor in outreach contexts (that is, given to groups other than traditional technical communities); Bandwidth Challenge entries at SC06 and SC07; and notices and articles in the high performance computing and grid computing press. These activities are described in greater detail below.

Talks specifically focused on outreach

- Simms, Stephen C. 2006. The Data Capacitor: Live from Dresden. IU Linuxfest 2006, Bloomington, Indiana via Technische Universitaet Dresden, Dresden, Germany, April 14, 2006.

- Simms, Stephen C. 2007. The Data Capacitor Project. Linuxfest 2007, Indianapolis, Indiana. April 11, 2007. (This talk was given in Indianapolis at IU's Linuxfest, an event held annually to educate students, faculty, staff, and the general public about the utility and availability of Linux.)

- Simms, Stephen C. 2007. The Data Capacitor. Rochester Institute of Technology, Rochester, New York. May 8, 2007. (This public talk was given at Rochester Institute of Technology to share with their community the ideas supporting and implementation of the Data Capacitor.)

Technical and Popular Press Articles

An important part of dissemination of information about any computer science research and development project is communication with the technical and lay communities. The following articles feature the Data Capacitor and research work done with it:

- "Indiana University Announces Lustre over WAN." HPCwire 29 April 2008. 31 October 2008. http://www.hpcwire.com/offthewire/IU_Announces_Lustre_over_WAN.html
- "IU Team Receives Award in Bandwidth Challenge." HPCwire 8 December 2006. 23 October 2007. http://www.hpcwire.com/hpc/1142351.html
- "IU Data Capacitor Reaches 977 MB/Second Across TeraGrid." HPCwire 8 June 2007. 23 October 2007. http://www.hpcwire.com/hpc/1601922.html
- "TechPoint Announces Mira Awards Nominations." Inside INdiana Business 3 April 2007. 23 October 2007. http://www.insideindianabusiness.com/newsitem.asp?ID=22604
- "Team Led by IU Wins Supercomputing Bandwidth Competition." HPCwire 16 November 2007. 31 October 2008. http://www.hpcwire.com/offthewire/17904409.html
- "TechPoint Announces Mira Awards Nominations." Inside INdiana Business 16 April 2008. 31 October 2008. http://www.insideindianabusiness.com/newsitem.asp?ID=28898

Research and Education Activities:

The Data Capacitor has been implemented and is in operation with capacity and capabilities greater than those outlined in the original MRI proposal to the NSF. The data capacitor comprises:

- 52 Dell servers running Red Hat Enterprise Linux
- 24 Myricom 10 gigabit Ethernet cards
- 12 DataDirect Networks S2A9550 storage controllers
- 30 DataDirect Networks 48 bay SATA disk chassis
- 535 terabytes usable disk storage using the Lustre filesystem

The Data Capacitor is housed in six water-cooled Rittal racks that provide high quality controlled environments for the system equipment. At peak performance, the Data Capacitor has a 14.5 gigabyte per second aggregate transfer rate per second. The Data Capacitor is providing unprecedented capabilities in the rapid I/O of very large data files locally, and via high-speed networks throughout the US. As a result, the Data Capacitor is changing the way IU computer and discipline scientists think about and go about their research.

Indiana University formally accepted the Data Capacitor on October 6, 2006 and successfully brought the Data Capacitor into full production mode meeting a self-imposed deadline of April 1, 2007.

We have implemented high speed connections between the Data Capacitor and the following other systems and labs at IU:

- IU's HPSS mover nodes, which serve as a front end to IU's 2.8 petabyte massive data storage system
- IU's 30.7 TFLOPS Big Red supercomputer
- A pool of GridFTP servers serving as data transfer nodes to resources located locally at IU and throughout the nation
- We have, as described in the original proposal, installed dedicated 10 Gigabit Ethernet links to IU's Computer Science, Astronomy, and Chemistry departments via a dedicated 10Gb fiber where both data acquisition and compute servers mount the filesystem.

Today, the Data Capacitor currently serves both local and national users by providing a large pool of high-speed storage to multiple distributed resources. While our major operational focus to date has been on serving local users, the Data Capacitor has generated considerable interest in the national research community. As a result, we are in the process (as of the writing of this annual report) of expanding the disk capacity of the Data Capacitor and making portions of the expanded disk space available to the national research community via the TeraGrid.

Our efforts this year have been focused in two areas: making the system fully functional and implementing a number of "capability-stretching" projects as proofs of concept of the value of the Data Capacitor to discipline scientists. In the former category, we have

focused on the highly nontrivial work of assembling and tuning the system and all of the code components, and then ensuring good functionality. IU has contributed important additions to the Lustre code base that have enabled secure use of Lustre as a Wide-Area Network file system. In particular, we have created a patch for the *lconf* command so that failover to a backup system (for fault tolerance) can, if so configured, automatically roll back to the intended primary production servers. In addition, we have been able to update and resubmit a patch to the MetaData Server code in versions 1.4 and 1.6 to enable what is referred to as "root squash" capabilities. This ensures that a super user on one component of the distributed Lustre-WAN installations will not have read/write privileges throughout user file space. Particularly important projects enabled by the Data Capacitor are described in the next section.

Contributions to Resources for Research and Education:

During the current reporting year the Data Capacitor has enabled several important (and in some cases, transformational) research projects, described below:

Common Instrument Middleware Architecture (CIMA). The Data Capacitor has served as a critical data holding place for research and development by the Common Instrument Middleware Architecture (CIMA) project (http://www.instrumentmiddleware.org/). CIMA has developed its initial implementations for management of crystallography data. In particular, the Data Capacitor has been used by the CIMA project team and the Indiana University Molecular Structure Center (IUMSC) to support the automatic tagging of diffractometer data with appropriate metadata as they are created, in real time. Their work continues as they stream live instrument data from multiple X-ray crystallography labs (representing three continents) onto the Data Capacitor, and the extremely fast I/O capabilities of the Data Capacitor are essential in enabling real-time data tagging.

Linked Environments for Atmospheric Discovery (LEAD). The Data Capacitor has facilitated research in several data management related areas in the Linked Environments for Atmospheric Discovery (LEAD) (http://portal.leadproject.org/) project, directed locally by Professors Beth Plale and Dennis Gannon. The Data Capacitor daily ingests gigabytes of observational and model generated atmospheric data, transforms the data to formats suitable for ingesting into a workflow system, and serves the data immediately for use in data driven computational workflows. Meteorology researchers want to feed current observational data from regional Doppler WSR-88d radars into their forecast models with minimal latency. The Indiana University architecture, where large amounts of fast disk are located near to supercomputing resources, makes it possible for LEAD to offer minimal latencies between severe storm detection by regional radars and the triggering of a weather forecast that will run at faster than real time to predict the storm development.

The Data Capacitor also hosts the LEAD Personal Workspace. The workspace is a managed disk space that houses the experimental workflow results for all LEAD users. With tools accessed from the LEAD gateway, users can browse and search their personal

workspaces. The Data Capacitor is key in the flexibility of this space because of its ability to support multiple protocols. For instance, meteorologists use the OPeNDAP protocol to access files, for instance, to pull into the Unidata IDV visualization tool. While OPeNDAP is limited to accessing specific data types, including it in a suite of protocols enhances the flexibility and usefulness of the personal workspace. A storage system that had limited protocol support, such as a mass store system, would have inhibited the evolution of this important tool developed by Dr. Beth Plale.

Semantic analysis of the Web. The Data Capacitor has provided Dr. Fil Menczer and his Informatics students with the enormous space required to process the semantic networks supporting the GiveALink project (http://www.givealink.org) or to perform textual analyses using all of Wikipedia as a data set.

Formation of gas-giant planets. Dr. Richard Durisen and graduate student Scott Michaels are developing simulations to better understand the formation of gas-giant planets. The Data Capacitor has provided the disk storage space they need in order to increase the resolution of their numerical simulations to permit better understanding of the formation of such planets.

Cancer research and multi-site data flows. Leveraging the Data Capacitor, Informatics professor Dr. Jake Chen recently garnered a prestigious bioinformatics research award from the Canary Foundation, whose slogan is "stopping cancer early." Dr. Chen will use the Data Capacitor to facilitate data sharing and multi-site workflows spanning Purdue University, IUPUI, and Indiana University Bloomington to further their work with the National Cancer Institute.

Reservation of ancient texts. Drs. P. R. Mukund and Roger Easton at Rochester Institute of Technology are using a Data Capacitor mount across the wide area network to support their efforts to preserve and archive scans of rare Sanskrit texts originally written on palm leaves in the 13th century.

Books or Other One-time Publications:

Gannon, Dennis, Beth Plale, Marcus Christie, Yi Huang, Scott Jensen, Ning Liu, Suresh Marru, Sangmi Lee Pallickara, Srinath Perera, Satoshi Shirasuna, Yogesh Simmhan, Aleksander Slominski, Yiming Sun, Nithya Vijayakumar, Building Grid Portals for e-Science: A Service Oriented Architecture To appear High Performance Computing and Grids in Action, IOS Press - Amsterdam, Lucio Grandinetti editor, 2007

Plale, Beth. 2007. Workload Characterization and Analysis of Storage and Bandwidth Needs of LEAD Workspace, LEAD TR001 V3.0

Stephen Simms, Gregory Pike, Douglas Balog, "Wide Area Filesystem Performance using Lustre on the TeraGrid." Proceedings of "TeraGrid '07 Conference", Madison, WI., 2007.

Stephen Simms, S. Teige, Bret Hammond, Yu Ma, C. Westneat, Larry L. Simms, Douglas A. Balog, "Empowering Distributed Workflow with the Data Capacitor: Maximizing Lustre Performance across the Wide Area Network." Proceedings of "Workshop on Service-Oriented Computing Performance: Aspects, Issues, and Approaches", Monterey, CA, 2007.

S.C. Simms, M. Davy, B. Hammond, M. Link, C. Stewart, R. Bramley, B. Plale, D. Gannon, M-H. Baik, S. Teige, J. Huffman, R. McMullen, D. Balog and G. Pike, All in a day's work: advancing data-intensive research with the data capacitor, SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing, ACM Press, New York, NY. Tampa, Florida, 2006, pp. 244. http://doi.acm.org/10.1145/1188455.1188711

Training and Development:

WxChallenge. The Data Capacitor played an important role in the LEAD project in supporting the 2007 Spring WxChallenge (http://www.wxchallenge.com/), a national collegiate forecasting challenge. During the competition, over 1,000 students from 56 institutions used the LEAD Gateway to produce weather forecasts from meteorological data stored on the Data Capacitor. Terabytes of forecast and simulation data were written back to the Data Capacitor for visualization, analysis, and eventual storage on HPSS. Dr. Dennis Gannon and research scientist Suresh Marru supported this challenge.

Talks and presentations. To promote the use and ideas underlying the creation of the Data Capacitor, invited talks and presentations were delivered to various national audiences. Delivered presentations are listed below:

Simms, Stephen C. 2006. The Data Capacitor and Lustre Wandemonium. SC|06 Lustre BOF, Tampa, FL. November 14, 2006.

Stewart, Craig A. 2006. On the road to petascale processing paradigms with IU's Big Red Supercomputer and IBM BladeCenter H. Indiana booth at SC|06, Tampa, FL. November 14 and 15, 2006.

Simms, Stephen C. 2006. All in a Day's Work: Advancing Data-Intensive Research with the Data Capacitor. Indiana booth at SC|06, Tampa, FL. November 15, 2006.

Simms, Stephen C. 2006. The Data Capacitor Project. Indiana booth at SC|06, Tampa, FL. November 15, 2006.

"Adaptive weather forecasting using LEAD", B. Plale and D. Gannon, 2006 International Conference on High Performance Computing, Networking, Storage, and Analysis (SC06), November 2006.

Simms, Stephen C. 2007. The Data Capacitor: Failover and Success. Data Search Institute Seminar, Bloomington, Indiana. February 6, 2007.

Simms, Stephen C. 2007. The Data Capacitor Grant. Research Computing Grant Retreat, Mooresville, Indiana. March 28, 2007.

Stewart, Craig A. 2007. Introduction to Research Computing. Ohio Supercomputer Center/ASC/MRSC. April 3, 2007.

McCaulay, D. Scott. 2007. IU TeraGrid Quarterly Update Q1 2007. TeraGrid Quarterly Meeting, Breckenridge, CO. April 11, 2007.

Simms, Stephen C. 2007. The Data Capacitor Project: Empowering Workflow and New Collaboration. Lustre User Group Meeting, Miami, Florida. April 24, 2007.

Simms, Stephen C. 2007. Research Technology at IU. Biology Annotation Local Jamboree, Bloomington, Indiana, May 18, 2007.

Simms, Stephen C. 2007. Wide Area Filesystem Performance Using Lustre on the TeraGrid. TeraGrid '07 Conference, Madison, WI. June 5, 2007.

Simms, Stephen C. 2007. Wide Area Filesystem Performance using Lustre on the TeraGrid. Wright –Patterson Air Force Base, Dayton, Ohio. June 13, 2007.

Stewart, Craig A. Implementation and experience with Big Red – a 20.4 TFLOPS IBM BladeCenter cluster. 2007. International Supercomputer Conference, Dresden Germany. June 26, 2007.

Outreach Activities:

We have engaged in three types of outreach activities: talks and presentations focused specifically on the Data Capacitor in outreach contexts (that is, given to groups other than traditional technical communities); a Bandwidth Challenge entry at SC06; and notices and articles in the high performance computing and grid computing press. These activities are described in greater detail below.

Talks specifically focused on outreach:

Simms, Stephen C. 2007. The Data Capacitor Project. Linuxfest 2007, Indianapolis, Indiana. April 11, 2007. (This talk was given in Indianapolis at IU's Linuxfest, an event held annually to educate students, faculty, staff, and the general public about the utility and availability of Linux.)

Simms, Stephen C. 2007. The Data Capacitor. Rochester Institute of Technology, Rochester, New York. May 8, 2007. (This public talk was given at Rochester Institute of

Technology to share with their community the ideas supporting and implementation of the Data Capacitor.)

SC06 Bandwidth Challenge:

The theme of the SC06 Bandwidth Challenge, "Bridging the Hero Gap," encouraged teams to show how their technology innovations allow researchers and scientists to manage and move large amounts of data without relying on "networking heroes" to help at every step. Participating teams were asked to fully use one 10 Gigabit (Gbit) path "end-to-end, disk-to-disk" from the competition venue in Tampa, Florida, back to each team's home institution using their actual production network. IU's effort titled "All in a Day's Work," emphasized how IU's advanced information technology infrastructure allows IU scientists and researchers to make use of leading edge data management and network services each and every day. The IU team demonstrated peak data transfer rate of 9.2 Gbit/second, with an approximate sustained average of 5.5 Gbit/second. IU's project featured four IU research teams running scientific analyses on IU's Data Capacitor. The IU team included Beth Plale, Suresh Marru, Chathura Herath, and AJ Ragusa on streaming CASA radar data for tornado prediction; Scott Teige and Stephen Simms on analysis of subatomic particles; Mookie Baik on analysis of disease-forming protein mutations; and Rick McMullen and John Huffman on capture of crystallographic data from multiple distributed sources.

Technical and Popular Press Articles:

An important part of dissemination of information about any computer science research and development project is communication with the technical and lay communities. The following articles feature the Data Capacitor and research work done with it:

“IU Team Receives Award in Bandwidth Challenge.” HPCwire 8 December 2006. 23 October 2007. <http://www.hpcwire.com/hpc/1142351.html>

“IU Data Capacitor Reaches 977 MB/Second Across TeraGrid.” HPCwire 8 June 2007. 23 October 2007. <http://www.hpcwire.com/hpc/1601922.html>

“TechPoint Announces Mira Awards Nominations.” Inside INdiana Business 3 April 2007. 23 October 2007.
<http://www.insideindianabusiness.com/newsitem.asp?ID=22604>

Web Pages:

http://datacapacitor.researchtechnologies.uits.iu.edu/
http://rtinfo.uits.iu.edu/cyberinfrastructure/resources.shtml
http://racinfo.indiana.edu/education_and_training/Cyberinfrastructure.pdf

References

Bailey, B.C., Fan, H., Huffman, J.C., Baik, M.-H. and Mindiola, D.J. *Intermolecular C-H Bond Activation of Hydrocarbons Promoted by Transient Titanium Alkylidynes. Synthesis, Reactivity, Kinetic and Theoretical Studies of the Ti°C Linkage*. Journal of the American Chemical Society, 2007. **129**: 8781–8793.

Bertrand, F., Bramley, R., Damevski, K.B., Kohl, J.A., Bernholdt, D.E., Larson, J.W. and Sussman, A. *Data Redistribution and Remote Method Invocation in Parallel Component Architectures*. Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05), IEEE Computer Society, 2005. **01**: 40–42.

Bertrand, F., Yuan, Y., Chiu, K. and Bramley, R. *An Approach to Parallel MxN Communication*. International Journal of High Performance Computing Applications, 2005. **19**: 399–407.

Bramley, R., Armstrong, R. and McInnes, L. *High-Performance Component Software Systems*. Parallel Processing for Scientific Computing, Heroux, M.A., Raghavan, P. and Simon, H.D., Society for Industrial and Applied Mathematics, Philadelphia, 2006.

Bramley, R., Chiu, K., Devadithya, T., Gupta, N., Hart, C., Huffman, J.C., Huffman, K., Ma, Y. and McMullen, D.F. *Instrument Monitoring, Data Sharing, and Archiving Using the Common Instrument Middleware Architecture (CIMA)*. Journal of Chemical Information and Modeling, 2006. **46**(3): 1017–1025.

Dede, Y., Zhang, X., Schlangen, M., Schwarz, Helmut and Baik, M.J. *A Redox Non-Innocent Ligand Controls the Life Time of a Reactive Quartet Excited State—An MCSCF Study of [Ni(H)(OH)]+*. Journal of the American Chemical Society, 2009.

Fan, H., Saleh, A., Clark, R., Adhikari, D., Zuno-Cruz, F., Sanchez-Cabrera, G., Huffman, J., Pink, M., Mindiola, D. and Baik, M. *Understanding Distorted T- vs. Y-Geometries for Neutral Chromous Complexes Supported by a Sterically Encumbering Beta-Diketiminate Ligand*. Journal of the American Chemical Society, 2008. **130**: 17351–17361.

Fullmer, B.C., Pink, M., Fan, H., Yang, X., Baik, M. and Caulton, K.G. *The Effect of One Valence Electron: Contrasting (PNP)Ni(CO) with (PNP)Ni(NO)*. Journal of Inorganic Chemistry, 2008. **47**: 3888–3892.

Gannon, D., Plale, B., Christie, M., Huang, Y., Jensen, S., Liu, N., Marru, S., Pallickara, S.L., Perera, S., Shirasunam, S., Simmhan, Y., Slominski, A., Sun, Y. and Vijayakumar, N. *Building Grid Portals for e-Science: A Services Oriented Architecture*. Grandinetti, L., Ed. , ISO Press, Amsterdam, 2007.

Guha, R., Wiggins, G., Wild, D., Baik, M., Pierce, M. and Fox, G. *Improving Usability and Accessibility of Cheminformatics Tools for Chemists Through Cyberinfrastructure and Education*. Cheminformatics, In Press, 2009. http://grids.ucs.indiana.edu/ptliupages/publications/cheminformatics-indiana.pdf.

Hossain, F., Rigsby, M., Duncan, C., Paul Milligan, J., Lord, R., Baik, M. and Schultz, F. *Synthesis, Structure and Properties of Low-Spin Manganese(III)-Poly(pyrazolyl)borate Complexes*. Inorganic Chemistry, 2007. **46**: 2596–2603.

Iyengar, S.S., Sumner, I. and Jakowski, J. *Hydrogen Tunneling in an Enzyme Active Site: A Quantum Wavepacket Dynamical Perspective*. Journal of Physical Chemistry, 2008. **B 112**: 7601–7613.

Jiang, X., Lim, Y., Zhang, B.J., Opsitnick, E., Baik, M. and Lee, D. *Dendritic Molecular Switch: Chiral Folding and Helicity Inversion*. Journal of the American Chemical Society, 2008. **130**: 16812–16822.

Ketterer, N., Fan, H., Blackmore, K., Yang, X., Ziller, J., Baik, M. and Heyduk, A. *Pi(dot)-Pi(dot) Bonding Interactions Generated by Halogen Oxidation of Zirconium(IV) Redox-Active Ligand*. Journal of the American Chemical Society, 2008. **130**: 4364–4374.

Kilgore, U., Sengelaub, C., Fan, H., Tomaszewski, J., Pink, M., Karty, J., Baik, M. and Mindiola, D. *A Transient V(III) Neopentylidene Complex. Redox Chemistry and Reactivity of the V=CHtBu Functionality*. Organometallics, 2009. **28**: 843–852.

Le, J., Lord, R., Noll, B., Baik, M., Schulz, C. and Scheidt, W. *Is Cyanide a Strong Field Ligand for Hemeprotein/Ferroheme?*, Angewandte Chemie Internationale, 2008. **47**.

Li, L., Liang, S., Pilcher, M.M., Standley, D.M. and Meroueh, S.O. *Molecular Design of Protein Interfaces with Flexible Backbone*. 2009.

Li, L., Uversky, V., Dunker, A. and Meroueh, S. *A Computational Investigation of Allostery in the Catabolite Activator Protein*. Journal of the American Chemical Society, 2007. **129**(50): 15668–15676.

Li, X., Moore, D. and Iyengar, S. *Insights from First Principles Molecular Dynamics Studies Towards Infra-Red Multiple-Photon and Single-Photon Action Spectroscopy: Case Study of the Proton-Bound Di-Methyl Ether Dimer*. Journal of Chemical Physics, 2008. **128**: 184308.

Li, X., Teige, V.E. and Iyengar, S.S. *Can the Four-Coordinated, Penta-Valent Oxygen in Hydroxide Water Clusters Be Detected Through Experimental Vibrational Spectroscopy?* Journal of Physical Chemistry, 2007. **A 111**: 4815–4820.

Ligand, S., Liwei, L., Hsu, W., Pilche, M., Uversky, V., Zhou, Y., Dunker, A. and Meroueh, S. *Exploring the Molecular Design of Protein Interaction Sites with Molecular Dynamics Simulations and Free Energy Calculations*. Biochemistry, 2009. **48**: 399–414.

Lord, R. and Baik, M. *Why Does Cyanide Pretend to be a Weak-Field Ligand in [Cr(CN)5]3-?*, Inorganic Chemistry, 2008. **47**: 4413–4420.

Lord, R., Schultz, F. and Baik, M. *Spin Crossover–Coupled Electron Transfer in [M(tacn)2]3+/2+ Complexes (tacn = 1,4,7-triazacyclononane; M = Cr, Mn, Fe, Co, Ni)*. Journal of American Chemical Society, 2009.

Ma, Y. and Bramley, R. *A Composable Data Management Architecture for Scientific Applications*. Proceedings of Challenges of Large Applications in Distributed Environments (CLADE) Workshop, 2005. 35–44.

Mantri, Y. and Baik, M. *Computational Models of Cisplatin*. Computational Inorganic and Bioinorganic Chemistry, Solomon, E., Ed., Wiley & Sons, Chichester, 2009.

Mantri, Y., Fioroni, M. and Baik, M. *Computational Study of the Binding of Cu(II) tp Alzheimer's Amyloid-beta Peptide—Do Abeta42 and Abeta40 Bind Copper in Identical Fashion?*, Journal of Biological and Inorganic Chemistry, 2008. **13**: 1197–1204.

Mantri, Y., Lippard, S. and Baik, M. *Bifunctional Binding of Cisplatin to DNA: Why Does Cisplatin Form 1,2-Intrastrand Crosslinks with AG, But Not with GA?* Journal of the American Chemical Society, 2007. **129**: 5023–5030.

Miao, Y. and Ortoleva, P. *Molecular Dynamics/Order Parameter eXtrapolation for Bionanosystem Simulations*. Journal of Computational Chemistry, 2008. **30**(3): 423–437.

Owston, R. and Abraham, J. *Flame Propagation in Stratified Hydrogen-Air Mixtures: Spark Placement Effects*. International Journal of Hydrogen Energy, 2009. **34**(15): 6532–6544.

Pitcock, W.H., Jr., Lord, R.L. and Baik, M.-H. *The Mechanism of the Rhodium(I)-Catalyzed [2 + 2 + 1] Carbocyclization Reaction of Dienes and CO—A Computational Study*. Journal of the American Chemical Society, 2008. **130**: 5821–5830.

Pordes, R. *Challenges Facing Production Grids*. High Performance Computing and Grids in Action, IOS Press, Amsterdam, 2008. 506–521.

Pordes, R., Altunay, M., Avery, P., Bejan, A., Blackburn, K., Blatecky, A., Gardner, R., Kramer, B., Livny, M., McGee, J., Potekhin, M., Quick, R., Olson, D., Roy, A., Sehgal, C., Wenaus, T., Wilde, M. and Würthwein, F. *New Science on the Open Science Grid*. Journal of Physics Conference Series, 2007. **125**: 012070.

Pordes, R., Petravick, D., Kramer, B., D. Olson , M.L., Roy, A., Avery, P., Blackburn, K., Wenaus, T., Würthwein, F., Foster, I., Gardner, R., Wilde, M., Blatecky, A., McGee, J. and Quick, R. *The Open Science Grid*. Journal of Physics Conference Series, 2007. **78**: 012057.

Pordes, R., Petravick, D., Kramer, B., Olson, D., Livny, M., Roy, A., Avery, P., Blackburn, K., Wenaus, T., Würthwein, F., Foster, I., Gardner, R., Wilde, M., Blatecky, A., McGee, J. and Quick, R. *The Open Science Grid--Its Status and Implementation Architecture*. JPCS Proceedings of International Conference on Computing in High Energy and Nuclear Physics (CHEP 07), 2007. http://cd-docdb.fnal.gov/cgi-bin/RetrieveFile?docid=2055&version=1&filename=OpenScienceGrid-CHEP-1.pdf

Post, D., Batchelor, D., Bramley, R., Cary, J.R., Cohen, R., Colella, P. and Jardin, S. *Report of the Fusion Simulation Project Steering Committee*. Journal of Fusion Energy, 2005. **23**(1): 1–26.

Risacher, S.L., Saykin, A.J., Shen, L., West, J.D., Kim, S., McDonald, B.C. and the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Volumetric, Cortical Thickness, and Grey Matter Density Changes on MRI Over 12 Months: Relationship to Conversion Status in the ADNI Cohort*. Alzheimer's and Dementia, 2009. **5**(4 Supplement 1): P200.

Risacher, S.L., Saykin, A.J., West, J.D., Shen, L., Firpi, H.A., McDonald, B.C. and the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Baseline MRI Predictors of Conversion from MCI to Probable AD in the ADNI Cohort*. Current Alzheimer Research, 2009. **6**(4): 347–361.

Saykin, A.J., West, J.D., Shen, L., Firpi, H.A., Wessels, A.M., Cannon, A., Risacher, S., McDonald, B.C. and the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Hippocampal and Distributed Gray Matter Abnormalities in MCI and Early AD in the ADNI Baseline 1.5T MRI Scans: Cross-Sectional Analysis and Relation to Verbal Learning*. Alzheimer's & Dementia, 2008. **4**(4 Supplement 2): T59.

Schultz, F., Lord, R., Yang, X. and Baik, M. *Inorganic Models for Two-Electron Redox Chemistry in Biological Systems: Ligand-Bridged Molybdenum and Tungsten Dimers*. Bioinorganic Chemistry, Chapter 10, Oxford University Press, ACS Symposium Series, 2009. **1012**: 151–166.

Scott, J., Fan, H., Wicker, B., Fout, A., Baik, M. and Mindiola, D. *Lewis Acid Stabilized Methylidene and Oxoscandium Complexes*. Journal of the American Chemical Society, 2008. **130**: 14438–14439.

Shen, L., Firpi, H.A., Saykin, A.J. and West, J.D. *Parametric Surface Modeling and Registration for Comparison of Manual and Automated Segmentation of the Hippocampus*. Hippocampus, 2009. **19**(6): 588–595.

Shen, L., Saykin, A.J., Firpi, H.A., West, J.D., McHugh, T.L., Wishart, H.A., Flashman, L.A., Wessels, A.M., McDonald, B.C. and Cannon, A. *Comparison of Manual and Automated Determination of Hippocampal Volumes in MCI and Older Adults with Cognitive Complaints*. Alzheimer's & Dementia, 2008. **4**(4 Supplement 2): T29–30.

Simms, S., Davy, M., Hammon, B., Lind, M., Stewart, C., Bramley, R., Plale, B., Gannon, D., Baik, M.-H., Teige, S., Huffman, J., McMullen, D., Balog, D. and Pike, G. *All in a Day's Work: Advancing Data-Intensive Research with the Data Capacitor*. Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, ACM Press, 2006. 244.

Sumner, I. and Iyengar, S.S. *Combining Quantum Wavepacket Ab Initio Molecular Dynamics (QWAIMD) with QM/MM and QM/QM Techniques: Implementation Blending ONIOM and Empirical Valence Bond Theory*. Journal of Chemical Physics, 2008. **129**: 054109.

Sumner, I. and Iyengar, S.S. *Quantum Wavepacket Ab Initio Molecular Dynamics: An Approach for Computing Dynamically Averaged Vibrational Spectra Including Critical Nuclear Quantum Effects*. Journal of Physical Chemistry, 2007. **A 111**: 10313–10324.

Vimal, D., Pacheco, A.B., Iyengar, S.S. and Stevens, P.S. *Experimental and Theoretical Study of the Kinetics of the OH +1,3-Butadiene Reaction Between 263 and 423 K at Low Pressure*. Journal of Physical Chemistry, 2008. **A 112**: 7227.

Wang, H., Sawyer, J.R., Evans, P.A. and Baik, M.-H. *Mechanistic Insight into the Diastereoselective Rhodium-Catalyzed Pauson-Khand Reaction: Role of Coordination Number in Stereocontrol*. Angewandte Chemie Internationale, 2007. **47**: 342–345.

Wang, Y., Saykin, A.J., Lin, C., Mosier, K.M., Kalnin, A.J. and Pfeuffer, J. *Within and Between Session Reproducibility of MR Perfusion using PICORE Q2TIPS at 3T*. International Society for Magnetic Resonance in Medicine 17th Scientific Meeting, Honolulu, HI, 2009, 2009. **17**: 1512.

Yang, X. and Baik, M.-H. *The Mechanism of Water Oxidation Catalysis Promoted by [tpyRu(IV)=O]2L3+—A Computational Study*. Journal of the American Chemical Society, 2008. **130**: 16231–16240.

Zhang, X., Schlangen, M., Dede, Y., Baik, M.-H. and Schwarz, H. *DFT-Studies on the Thermal Activation of Molecular Oxygen by Bare [Ni(H)(OH)]+*. Helvetica Chimica Acta, 2009. **92**: 151–164.