A SciDAC Institute for Computer Science and Data

Project Summary

The Department of Energy (DOE) has led the world in enabling large-scale high-performance computing as an indispensable tool for scientific discovery in a broad range of disciplines. Our newest generation of systems is not simply larger than past systems — it brings distinct and novel challenges. These challenges arise from ever-increasing core counts on nodes; the deepening of the memory hierarchy and inclusion of fast, nonvolatile storage within the HPC system; the widening ratio of the FLOPS to I/O bandwidth; and the increasing ubiquity of computation accelerators. The data produced by these simulations on these high-end systems has also increased by orders of magnitude in size and complexity. This continuing explosion of data mandates constant attention and improvement to the tools and techniques used to manage and analyze this data. Moreover, the breadth of science performed on DOE advanced computing resources is growing and new motifs of investigation are emerging, including those involving analysis of experimental and observational data.

In response to these challenges, we propose RAPIDS: a SciDAC Institute for Resource and Application Productivity through computation, Information, and Data Science. **The objective of RAPIDS is to assist Office of Science (SC) application teams in overcoming computer science and data challenges in the use of DOE supercomputing resources to achieve science breakthroughs.** To accomplish this objective, the Institute will solve computer science and data technical challenges for SciDAC and SC science teams, work directly with SC scientists and DOE facilities to adopt and support our technologies, and coordinate with other DOE computer science and applied mathematics activities to maximize impact on SC science. RAPIDS will be organized around four primary thrust areas. The cornerstone of the project is the application engagement and community outreach thrust area. A rigorous plan for engagement through direct interactions, as well as outreach to connect with the broader community, will allow us to have the greatest and most immediate impact. Software technologies for computation, information, and data science are central to the success of SciDAC and SC science on leadership computing platforms. Our three remaining technology thrust areas — data understanding, platform readiness, and scientific data management — focus on key challenges in these areas, bringing to bear mature technologies and enhancing these technologies to adapt to new science requirements and new platforms. Technical thrust area participants will engage directly with application teams to deploy these technologies and to glean insights on how to further improve these tools.

Rob Ross (ANL) will serve as the overall lead, with Lenny Oliker (LBNL) acting as deputy. Together they bring many years of extensive knowledge in successfully managing large, multi-institutional SciDAC projects and ASCR computer science teams. RAPIDS will bring together six DOE Laboratories and key industry and university partners with a broad range of relevant expertise, strong history of success in engagement with DOE scientists, and a strong record of software development and deployment at DOE facilities. The application engagements and technologies enabled and fostered under RAPIDS will enable science breakthroughs on current and upcoming DOE computing platforms and continue paying dividends for many years to come.

DOE Cover Page

Title: A SciDAC Institute for Computer Science and Data

Institution

Argonne National Laboratory
Mathematics and Computer Science Division
9700 S. Cass Ave, Building 240
Lemont, IL 60439

Administrative Point of Contact

Scott A. Ehling, Assistant Division Director
(T) 630-252-3338
sehling@mcs.anl.gov

Institute Director

Robert B. Ross, Senior Computer Scientist
(T) 630-252-4588
rross@mcs.anl.gov

DOE National Laboratory Announcement: LAB 17-1787

DOE/Office of Science Program Office: Advanced Scientific Computing Research

DOE Technical Contact

Dr. Ceren Susut
Ceren.Susut-Bennet@science.doe.gov

Research Area: Computer Science

Table 1: RAPIDS Team Members

Last Name	First Name	Title	Institution
Ross	Rob	Director, Principal Investigator	ANL
Balaprakash	Prasanna	Co-Lead: Data Understanding	ANL
Dubey	Anshu	Lead: Application Engagement	ANL
Guo	Hanqi	Senior Personnel	ANL
Harms	Kevin	ALCF Facility Liaison	ANL
Hovland	Paul	Co-Lead: Platform Readiness	ANL
Latham	Robert	Senior Personnel	ANL
Perarnau	Swann	Senior Personnel	ANL
Peterka	Thomas	Senior Personnel	ANL
Oliker	Lenny	Deputy Director, Principal Investigator	LBNL
Bethel	Wes	Lead: Data Understanding	LBNL
Buluc	Aydin	Senior Personnel	LBNL
Iancu	Costin	Senior Personnel	LBNL
Morozov	Dmitriy	Senior Personnel	LBNL
Ruebel	Oliver	Senior Personnel	LBNL
Sim	Alex	Senior Personnel	LBNL
Weber	Gunther	Senior Personnel	LBNL
Williams	Samuel	Co-Lead: Application Engagement, Math Institute liaison	LBNL
Wu	John	Co-Lead: Scientific Data Management	LBNL
de Supinski	Bronis	Principal Investigator	LLNL
Brugger	Eric	Senior Personnel	LLNL
Liao	Chunhua	Senior Personnel	LLNL
Mohror	Kathryn	Senior Personnel	LLNL
Ahrens	James	ECP Liaison, Principal Investigator	LANL
Grosset	Pascal	Senior Personnel	LANL
Patchett	John	Senior Personnel	LANL
Klasky	Scott	Lead: Scientific Data Management, Principal Investigator	ORNL
Bernholdt	David	Outreach Liaison	ORNL
Lee	Seyong	Senior Personnel	ORNL
Mehta	Kshitij	Senior Personnel	ORNL
Moore	Shirley	Senior Personnel	ORNL
Patton	Robert	Senior Personnel	ORNL
Podhorszki	Norbert	Senior Personnel	ORNL
Pugmire	Dave	Senior Personnel	ORNL
Roth	Philip	Senior Personnel	ORNL
Steed	Chad	Senior Personnel	ORNL
Vetter	Jeffrey	Lead: Platform Readiness	ORNL
Yoo	Shinjae	Principal Investigator	BNL
Liao	Wei-Keng	Principal Investigator	Northwestern
Choudhary	Alok	Senior Personnel	Northwestern
Parashar	Manish	Principal Investigator	Rutgers
Shen	Han-Wei	Principal Investigator	Ohio State

Continued on next page

Table 1 – continued from previous page

Last Name	First Name	Title	Institution
Siegel	Stephen	Principal Investigator	Delaware
Huck	Kevin	Principal Investigator	Oregon
Norris	Boyana	Senior Personnel	Oregon
Malony	Allen	Senior Personnel	Oregon
Hall	Mary	Principal Investigator	Utah
Geveci	Berk	Principal Investigator	Kitware
Fortner	Neil	Principal Investigator	HDF Group

Table 2: Collaborators of Team Members

Last Name	First Name	Institution Name
Abbasi	Hasan	Amazon
Abraham	Jacob A	University of Texas at Austin
Abram	Gregory	Texas Advanced Computing Center
Abramson	David	University of Queensland
Adve	Sarita	University of Illinois at Urbana-Champaign
Adve	Vikram	University of Illinois at Urbana-Champaign
Afibuzzaman	Md.	Michigan State University
Afsahi	Ahmad	Queen's University, Ontario
Agha	Gul	University of Arizona
Agnew	Grace	Amazon
Agrawal	A.	University of Illinois at Urbana-Champaign
Ahern	Sean	CEI, Inc
Aiken	A.	Stanford University
Aimone	James B.	Sandia National Laboratories
Akoglu	A.	Stanford University
Aktulga	H. Metin	Michigan State University
Aktulga	Metin	Michigan State
Alam	Sadaf	CSCS, Switzerland
Alexandrov	Vassil	Barcelona Supercomputing Center
Allen	Gabrielle	NCSA
Allen	Randall	Mentor Graphics
Altintas	Ilkay	SDSC
Amato	Nancy	Texas A&M University
Amit	Chourasia	UCSD
Amundson	James	FNAL
Antao	Samuel	IBM
Antoniu	Gabriel	INRIA
Apon	Amy	Clemson University
Arif	Mahwish	Queen's University, Belfast
Armstrong	R	UC Davis
Arnold	Dorian	Emory University
Arpaci-Dusseau	Andrea	University of Wisconsin Madison
Arpith	Jacob	IBM
Arunagiri	Sarala	Smart High Reliability Solutions
Asanovic	Krste	UC Berkeley
Aupy	Guillaume	INRIA
Ayguade	Eduard	Universitat Politicnica de Catalunya
Bader	David	Georgia Institute of Tech
Bae	Hansang	Intel
Bagchi	Saurabh	Purdue University
Baier	Christian	RWTH Aachen University
Bailey	Peter	Google
Balaras	Elias	GWU
Balduccini	Marcello	Drexel University

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Ballard	Grey	Wake Forest
Bangalore	Purushotham	U. of Alabama at Birmingham
Banicescu	Ioana	MSU - Mississippi State University
Barth	William	TACC
Bartlett	Rodney J.	Florida
Bartlett	Roscoe	Sandia National Laboratories
Batchelor	Donald	Retired
Baumann	Ronald	Texas Instruments, Germany
Bautista Gomez	Leonardo A.	Barcelona Supercomputing Center
Beck	M.	U. Tennessee
Beketayev	Kenes	SparcIt
Belli	Emily	General Atomics
Bennett	Janine	Sandia National Laboratories
Bent	John	EMC
Benzi	Michele	Emory University
Bercea	Gheorghe-Teodor	Imperial College
Bergman	Karen	Columbia University
Bergstrm	Christopher	Pathscale
Bernd	Sturmfels	UC Berkeley
Berry	Lee	Retired
Berry	Michael	University of Tennessee
Bertolli	Carlo	IBM
Bhargava	Bharat	Purdue University
Bhowmick	Sanjukta	University of Nebraska - Omaha
Biedron	Sandra	CSU
Birattari	Mauro	Universit Libre de Bruxelles, Belgium
Bischof	Christian	TU Darmstadt
Biswas	Rupak	NAS Ames Research
Blackmore	Robert	IBM
Blanco	Mark	Rensselaer Polytechnic Institute
Bland	Wesley	Intel
Blas	Francisco	U. Carlos III of Madrid
Blondel	Sophie	University of Tennessee
Blum	Volker	Duke University
Bonoli	Paul T.	MIT
Bordawekar	Rajesh	IBM
Bose	Pradip	IBM
Bosilca	George	University of Tennessee
Bosse	Torsten	Friedrich Schiller University Jena
Bouguerra	Mohamed-Slim	Yahoo
Brandt	Steve	Louisiana State University
Brennan	Dylan	Princeton
Brewer	Eric	Google / University of California , Berkeley
Bridges	Patrick	University of New Mexico
Brightwell	Ron	Sandia National Laboratories

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Bronevetsky	Greg	Google
Brooks	Richard	Clemson University
Brower	Richard	Brown University
Brown	Jed	University of Colorado Boulder
Brown	Jesse	UCSF
Browne	James	University of TX, Austin
Brunst	Holger	Technische Universitt Dresden
Buaba	Ruben	NC State U.
Bui	Hoang	Western Illinois University
Buntinas	Darius	Jump Trading
Burns	Randal	Johns Hopkins University
Calandra	Henri	Total E&P
Calder	Alan	Stony Brook University
Cameron	Kirk	Virginia Tech
Candy	Jeff	General Atomics
Carlson	Bill	IDA Center for Computing Sciences
Carlsson	Gunnar	Stanford University
Carothers	Christopher	Rensselaer Polytechnic Institute
Carpendale	Sheelagh	University of Calgary
Carr	Hamish	University of Leeds
Carretero	Jesus	U. Carlos III of Madrid
Carrington	Laura	San Diego Supercomputer Center
Carson	Erin	UC Berkeley
Carver	Jeffrey	Alabama
Cattaneo	Fausto	U Chicago
Cavazos	John	University of Delaware
Chaimov	Nicholas	ParaTools
Chame	Jacqueline	University of Southern California-ISI
Chan	Anthony	MathWorks
Chandrasekaran	Sunita	U. of Delaware
Chang	Choong-Seock	PPPL
Chang	Edward	UCSF
Chapel	Edward	NJEDge.Net
Chapman	Barbara	Stony Brook University
Chapman	Jarrod	N.A.
Chaudhuri	Abon	Walmart Labs
Chen	Chun-Ming	Google
Chen	Jacqueline	Sandia National Laboratories
Chen	Junjie	Texas Tech University
Chen	Roubing	Bosch
Chen	Yong	Texas Tech University
Chen	Zhengzhang	NEC
Chhugani	Jatin	HiPerform Inc.
Chien	Andrew	U Chicago
Ching	Avery	Facebook

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Chiu	Steve	Idaho State U.
Chivers	Daniel	Berkeley Applied Analytics
Choi	Jaesik	UNIST
Chopella	Venkatesh	IIIT-Hyderabad
Chou	Jerry	National Tsinghua Univ
Choudary	Chekuri S.	RNET
Chun	Miyoung	Kavli Foundation
Ciesko	Jan	Barcelona Supercomputing Center
Cisek	Daniel	Stony Brook University
Clark	Kate	NVIDIA
Cobb	Guy	Google
Coloma	Kenin	Ask.com
Conte	Tom	Georgia Institute of Tech
Coster	David	Max Planck Institute for Plasma Physics
Coteus	Paul	IBM
Couch	Sean	MSU
Craig	Anthony	UCAR
Crook	Russell	Surface Science Analytics, LLC
Crossno	Patricia	Sandia National Laboratories
Cui	Xuewen	Virginia Tech
Curry	Matthew	Sandia National Laboratories
Curtis	Anthony	Stony Brook University
Dai	Dong	Texas Tech University
Dai	Luogeng	Clemson University
Dalisay	Doralyn	Washington State University
Dally	Bill	NVIDIA Corp.
Danalis	Anthony	University of Tennessee
Das	Abhishek	Intel
Das	Sajal	Missouri University of Science and Technology
Davin	Laurance B.	Washington State University
Dayal	Jay	Intel
DeRose	Luiz	Cray Inc.
de Silva	Vin	Pomona College
Dean	Mark	University of Tennessee
Dean	Thomas	Google
Deattista	Kurt	Warwick, UK
Debusschere	Bert	Sandia National Laboratories
Deelman	Ewa	ISI
Demir	Dogan	Ouva
Demmel	James	UC Berkeley
Denker	Michael	Juelich Research Center
Di Fatta	Giuseppe	University of Reading
Diesmann	Markus	Juelich Research Center
Dillard	Scott	Google
Dinh	Minh Ngoc	University of Queensland

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Diniz	Pedro	University of Southern California-ISI
Dongarra	Jack	University of Tennessee
Donzis	Diego	Texas A&M University
Dorigo	Marco	Universit Libre de Bruxelles, Belgium
Dougherty	Max	UCSF
Dovrolis	Constantine	Georgia Institute of Tech
Driscoll	Michael	UC Berkeley
Duque	Earl	Intelligent Light
Duran	Alejandro	Intel
Durant	Marc	Trimble Navigation
Duriakova	Erika	University College Dublin
Dwyer	Matthew	Univ. of Nebraska Lincoln
Ebert	Achim	TU Kaiserslautern
Ebert	David	Purdue University
Edelsbrunner	Herbert	Institute of Science and Technology Austria
Edenhofner	John	Univ. of Delaware
Eichenberger	Alexandre	IBM
Eigenmann	Rudolf	Purdue University
Eisenhauer	Greg	Georgia Institute of Tech
Eliot	Jim	AWE
Endert	Alex	Georgia Tech University
Endrei	Mark	University of Queensland
Enz	Christian	Ecole Polytechnique Federale de Lausanne
Epifanovsky	Evgenny	Q-Chem
Erez	Mattan	University of Texas at Austin
Erler	Jochen	German Cancer Research Center
Ernst	D.	MIT
Estep	Donald	Colorado State University
Estlund	Karen	Pennsylvania State University
Ethier	Stephane	PPPL
Evans	Jenni	Pennsylvania State University
Evans	Noah	New Mexico Consortium
Fagan	Mike	Rice University
Favre	Jean	CSCS
Fazzari	Saverio	Booz Allen Hamilton Inc
Feibush	Eliot	PPPL
Feichtner-Kozlov	Dmitry	U Bremen
Felker	Kyle	Princeton University
Fenandez-Serra	Maria	Stony Brook University
Feng	Huan	Montclair University
Feng	Wu-chun	Virginia Tech
Ferreira	Kurt	Sandia National Laboratories
Fiala	David	Google
Figueiredo	Renato	University of Florida
Fischer	Curt R.	Stanford University

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Fisher	Brian	Simon Fraser University
Flocke	Norbert	U Chicago
Fogal	Tom	NVIDIA
Foley	Samantha S.	U Wisc.-La Crosse
Fong	Liana	IBM
Fortes	Jose	University of Florida
Fowler	Robert	RENCI
Fox	Armando	UC Berkeley
Fox	G.	Indiana Univ.
Freire	Juliana	NYU
Freudenthal	Eric	University of Texas at El Paso
Frings	Wolfgang	RWTH Aachen University
Frye	Richard	University of Arkansas
Fu	Song	University of North Texas
Fuerlinger	Karl	LMU-Munich, Germany
Fujita	Hajime	Intel
Gabriel	Edgar	University of Houston
Gainaru	Ana	University of Illinois at Urbana-Champaign
Gaither	Kelly	TACC
Ganapathysubramanian	Baskar	ISU - Iowa State University
Ganguly	Auroop	Northeastern University
Gao	Kui	Dassault Systemes Simulia Corp
Garth	Christoph	TU Kaiserslautern
Gavrilovska	Ada	Georgia Institute of Tech
Geimer	Markus	FZ Juelich, Germany
Georganas	Evangelos	Intel
Gerofi	Balazs	RIKEN
Gerrard	Mitchell	Univ. of Nebraska Lincoln
Gerstein	Mark	Yale University
Gibson	Garth	CMU
Gilbert	John	UCSB
Giles	C. Lee	Pennsylvania State University
Giles	Merle	NCSA
Gioiosa	Roberto	PNNL
Glassbrook	Richard	Capital Advanced Technology Services
Gobbert	Jens Henrik	Juelich Supercomputing Center
Goehner	Joshua	University of New Mexico
Goil	Sanjay	Intel
Gonzalez	Joseph	UC Berkeley
Goodell	David	Cisco
Gordon	Mark	Iowa State University/Ames Laboratory
Gorelenkov	Nikolai	PPPL
Gottschlich	Justin	Intel
Graham	Richard	Mellanox Technologies
Gramacy	Robert	Virginia Tech

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Grant	Ryan	Sandia National Laboratories
Gray	Genetha A.	Intel
Graziani	Carlo	U Chicago
Greenwald	Martin	MIT
Grieson	Brian	PPPL
Grigori	Laura	INRIA
Gropp	William	University of Illinois at Urbana-Champaign
Grout	Ray	NREL
Grover	Ronald	General Motors
Guha	Apala	IIT Delhi
Gurtowski	James	Target Rock Partners
Guzik	Stephen	Colorado State University
Habib	Salman	U Chicago
Haensell	Fabian	RWTH Aachen University
Hagen	Hans	Univ of Kaiserlautern
Hahnfeld	Jonas	RWTH Aachen University
Haider	Azzam	U. Tennessee
Hamann	Bernd	UC Davis
Hamilton	Stephen	Johns Hopkins University
Hammond	Karl	University of Missouri
Hammouda	Adam	Bridge Financial Technology
Han	Dianwei	Mount Sinai Medicine School
Hanrahan	Paul	Stanford University
Hao	Pengfei	University of Houston
Hariri	Salim	University of Arizona
Harris	Rachel	University of Arizona
He	Fei	Kansas State University
He	Xubin	Temple University
Healey	Christopher	NCSU
Heath	Michael	University of Illinois at Urbana-Champaign
Heikes	Ross	CSU
Heine	Christian	TU Kaiserslautern
Hendrix	William	U. South Florida
Henning	Meyerhenke	KIT
Herault	Thomas	University of Tennessee
Heroux	Michael	Sandia National Laboratories
Hesthaven	Jan	Brown University
Hilbrich	Tobias	Technische Universitt Dresden
Hlawitschka	Mario	HTWK Leipzig
Hoefler	Torsten	ETH, Switzerland
Hoffmann	Henry	University of Chicago
Hoisie	Adolfy	PNNL
Holland	Chris	UCSD
Hollingsworth	Jeffrey	University of Maryland
Holmes	Daniel	EPCC

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Homaifar	Abdollah	NC A&T State U.
Honavar	Vasant	Pennsylvania State University
Hong	Yang	Washington State University
Horton	Mitch	University of Tennessee
Houlberg	Wayne	Retired
House	Donald	Clemson University
Howar	Falk	Clausthal Univ. of Technology
Hu	Lin	UMass Amherst
Huang	Hao	GE Research
Huang	Jian	University of Tennessee
Huang	Min-Yu	VMware
Huang	Ping	Virginia Commonwealth University
Huansong	Fu	Florida State University
Huard	Guillaume	LIG, France
Hudson	Douglas L.	Retired
Hueckelheim	Jan	Imperial College London
Hugues	Maxime	Total E&P
Huntoon	Gwendolyn	KINBER
Hwu	Wen-Mei	University of Illinois at Urbana-Champaign
Ibrahim	Shadi	INRIA
Isaacs	Katherine	University of Arizona, Tucson
Isaila	Florin	University of Madrid
Ishikawa	Yutaka	RIKEN
Jablonowski	Christiane	Michigan
Jackson	Charles	Texas
Jacobsen	Doug	Intel
Jagode	Heike	University of Tennessee
Jana	Siddhartha	University of Houston
Jansen	Kenneth	UC Boulder
Jardin	Steve	PPPL
Jegelka	Stefanie	MIT
Jenko	F.	UC LosAngelos
Jernigan	Crystal	Leidos
Jessup	Elizabeth	University of Colorado, Boulder
Jiang	Jiming	UC Davis
Jin	Chao	University of Queensland/Monash University
Jin	Henry	NASA Ames Research
Johnson	Andrew	UIC
Johnson	Calvin	SDSU
Johnson	Fred	Leidos
Jones	Ian	Ocado, UK
Jordon	Kirk	IBM
Joy	Ken	UC Davis
Kaiser	Hartmut	Louisiana State University
Kale	Laxmikant	University of Illinois at Urbana-Champaign

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Kalukin	Andrew	Booz Allen Hamilton Inc
Kamil	Amir	U. Michigan
Kamil	Shoaib	Adobe
Kandemir	Mahmut	Pennsylvania State University
Kannan	Aswin	Oracle
Karakoyunlu	Cengiz	U. Connecticut
Karimabadi	Homa	U.C. San Diego
Karlsson	Sven	Technical University of Denmark
Kasen	Daniel	UC Berkeley
Kasiviswanathan	Shiva	Samsung Research
Katti	Amogh	University of Minnesota
Katz	Dan	University of Chicago
Kaufmann	Arie	SBU
Kelley	Tim	NCSU
Kelly	Paul	ICL
Kendall	Wes	Ambition LLC
Kepner	Jeremy	MIT-LL
Kerber	Michael	TU Graz
Kessel	Chuck	PPPL
Keum	Seung Hwan	General Motors
Khan	Kamil	McMaster University
Kim	Hyunjoo	Xerox
Kim	Jinoh	Texas A&M University
Kim	Kye Won	Washington State University
Kim	SeongJo	Pennsylvania State University
Kimpe	Dries	KCG Holdings
Klaassen	Withana	University of Texas at El Paso
Klemm	Michael	Intel
Klinvex	Alicia	Sandia National Laboratories
Klundt	Ruth	Sandia National Laboratories
Knight	Nick	UC Berkeley
Knio	Omar	Duke University
Knoll	Aaron	Texas Advanced Computing Center
Knudson	Kevin	U Florida
Koanantakool	Penporn	UC Berkeley
Koch	Christof	Allen Institute for Brain Science
Kocoloski	Brian	University of Pittsburgh
Koenning	Bjoern	Technical University Berlin
Kolla	Hemanth	Sandia National Laboratories
Kordenbrock	T.	Hewlett Packard
Krishnaiyer	Rakesh	Intel
Krishnamoorthy	Sriram	Pacific Northwest National Laboratory
Krishnamurthy	Venkat	Cray Inc.
Kritz	Arnold	Lehigh University
Krylov	Anna	USC - University of Southern California

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Ku	S.H.	PPPL
Kua	John	Berkeley Applied Analytics
Kubiatowicz	John	University of California , Berkeley
Kulkarni	Milind	Purdue University
Kumar	Mohit	Wayne State University
Kumar	Prabhat	4C Insights
Kumar	Vipin	UMN - University of Minnesota
Kumari	Sunita	Cold Spring Harbor Laboratory
Kurc	Tahsin	Stony Brook University
Kurtz	Rick	PNNL
Kwasniewski	Grezgorz	ETH, Switzerland
Kyrpides	Nikos	JGI
LaPre	Justin	Rensselaer Polytechnic Institute
Labarta	Jesus	Barcelona Supercomputing Center
Lagadapati	Mahesh	NVIDIA
Lam	Michael	University of Maryland
Lamb	Donald	U Chicago
Lange	Jack	University of Pittsburgh
Lao	Lang	General Atomics
Laros	James	Sandia National Laboratories
Lauer	Frank	University of Tennessee
Laura	Grigori	INRIA Saclay-Ile de France
Lecomber	David	Allinea
Lee	Choonseok	Washington State University
Lee	Dongeun	Texas A&M University
Lee	Dongwook	UC Santa Cruz
Lee	Hayan	Stanford University
Lee	Sang-Ik	Intel
Lee	Ten-Yok	MERL
Lee	W.	PPPL
Legrand	Arnaud	CNRS
Leigh	Jason	UIC
Leon-Carrion	Jose	Universidad de Sevilla
Leonidas	Guibas	Stanford University
Leung	Ruby	PNNL
Lewis	Norman	Washington State University
Lewis	Ryan	Stanford University
Li	Dong	UC Merced
Li	Jianwei	Bloomberg
Li	Kelvin	IBM
Li	Tonglin	IIT
Liang	Xu	University of Pittsburgh
Liao	Jianwei	Southwest University of China
Liberty	Dean	Advanced Micro Devices
Lifflander	Jonathan	University of Illinois at Urbana-Champaign

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Ligon	Walter	Clemson University
Lin	Pei-Hung	Oakland University
Lin	Zhihong	UC Irvine
Litvinova	Antonina	Gresham Computing, UK
Liu	Ruoqian	Uber
Liu	Ying	Chinese Academy of Sciences
Lofstead	Gerald	Sandia National Laboratories
Loh	Gabriel	Advanced Micro Devices
Lohrmann	Erich	Georgia Institute of Tech
Long	Rogelio	University of Texas at El Paso
Loren M.	Frank	UCSF
Lowenthal	David	University of Arizona
Lu	QiQi	Virginia Commonwealth University
Lucas	Robert	ISI/USC
Luchangco	Victor	Oracle
Lucia	Joe	Temple University
Ludascher	Bertram	UC Davis
Lugowski	Adam	UCSB
Lukic	Zarija	LBNL
Lumsdaine	Andrew	Indiana Univ.
Luo	Ziqing	Univ. of Delaware
Luu	Huong	University of Illinois at Urbana-Champaign
Ma	Kwan-Liu	UC Davis
Madduri	Kamesh	Pennsylvania State University
Mahajan	Manoj	National Geospatial-intelligence Agency
Mahoney	Michael	UC Berkeley
Maiterth	Matthias	Ludwig Maximilian University
Maltzahn	Carlos	UC Santa Cruz
Manne	Frederik	U. of Bergen
Marathe	Aniruddha	University of Arizona
Marcus	Shoshana	City University of New York
Marianiello	Andre	SevOne, Inc.
Marin	Gabriel	Google
Maris	Pieter	Iowa State
Maroudas	Dimitrios	UMass Amherst
Martorell	Xavier	Barcelona Supercomputing Center
Maruyama	Naoya	RIKEN
Maslov	Sergei	University of Illinois Urbana-Champaign
Mateo	Sergi	Barcelona Supercomputing Center
Matsuoka	Satoshi	Tokyo Institute of Technology
Mattson	ITimothy	ntel
Matzner	Richard	U. Texas
Maxwell	Reed	Colo. School Mines
McBryan	Oliver	UC Boulder
McCombie	Richard	Cold Spring Harbor Laboratory

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
McCraw	Heike	University of Tennessee
McElvain	Kenneth	UC Berkeley
McKee	Sally	Chalmers University of Technology
McMillan	Scott	CMU-SEI
Meijer	Jeroen	Univ. of Twente
Mellor-Crummey	John	Rice University
Meneghini	Orso	General Atomics
Meneveau	Charles	Johns Hopkins University
Meredith	Jeremy	Google
Messmer	Peter	NVIDIA
Metzler	Uwe	University of Applied Sciences Berlin
Mickelson	Sheri	UCAR
Midkiff	Samuel	Purdue University
Mikre	Henik	BLEN Corp
Miller	Barton	University of Wisconsin
Misra	Sanchit	Intel India
Mitra	Subhasish	Stanford University
Mohr	Bernd	FZ Juelich, Germany
Monga	Inder	Esnet
Moore	Shirley	University of Tennessee
More	Sachin	EMC
Moreira	IJose	BM
Moreland	Kenneth	Sandia National Laboratories
Moser	Robert	University of TX, Austin
Moussalem	Maysam	TACC
Mueller	Frank	NCSU
Mueller	Jens-Dominik	Queen Mary
Mueller	Matthias	RWTH Aachen University
Mukhanov	Lev	Queen's University, Belfast
Nagel	Wolfgang	TU Dresden, Germany
Najm	Habib	Sandia National Laboratories
Nakka	Nithin	University of Illinois at Urbana-Champaign
Nanjegowda	Ramachandra	University of Houston
Narahari	Bhagirath	GWU
Narayanamurthi	Mahesh	Virginia Tech
Narayanan	Ramanathan	Goldman Sachs
Naumann	Uwe	RWTH Aachen University
Navratil	Paul	TACC
Nelson	Thomas	University of Colorado
Newman	David	U Alaska, Fairbanks
Nguyen	Hoa	Microsoft
Ni	Lionel	University of Macau
Ni	Xiang	IBM
Nie	Bin	College of William & Mary
Nigmetov	Arnur	TU Graz

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Nikolopoulos	Dimitrios	Queen's University, Belfast
Nisar	Arifa	Salesforce
No	Jaechun	Sejong U.
North	Chris	Virginia Tech
Nunez	Filomena M.	Michigan State University
O'Brien	Kathryn	IBM
O'Connor	Mark	Allinea
O'Shea	Brian	MSU
Oesterling	Patrick	University of Leipzig
Oh	Sang-Yun	UCSB
Oldfield	Ron	Sandia National Laboratories
Olivier	Stephen	Sandia National Laboratories
Olschanowsky	Cathie	Boise State University
Otoo	E	U. of Witwatersrand
Owens	John	UC Davis
Ozisikyilmaz	Berkin	4C Insights
Pai	Vijay	Purdue University
Palsetia	Diana	4C Insights
Parashar	Manish	Rutgers Univerisity
Parasyris	Nikos	University of Edinburgh
Parchman	Zachary W.	Tennessee Tech.
Parker	Scott	University of Colorado Boulder
Pastore	Giovanni	INL
Patel	Janak	University of Illinois at Urbana-Champaign
Pattabiraman	Bharath	Google
Patterson	David	UC Berkeley
Patwary	Mostofa	Intel
Pavlos	Petoumenos	University of Edinburgh
Pawlowski	Roger	Sandia National Laboratories
Pawlowski	Steve	Micron Corp.
Peckham	Scott	University of Colorado Boulder
Pedretti	Kevin	Sandia National Laboratories
Peng	Zhenzhou	Snapchat Inc
Philippidis	Cesar	Mentor Graphics
Phillips	Carolyn L.	Neurensic
Pirkelbauer	Peter	University of Alabama at Birmingham
Pisharath	Jayaprakash	Intel
Pollonini	Luca	University of Houston
Porterfield	Allan	LogicBlox
Portluru	Vamsi	Comcast, VA
Pothen	Alex	Purdue University
Poxon	Heidi	Cray Inc.
Prasanna	Viktor	USC - University of Southern California
Prins	Jan	University of North Carolina
Protze	Joachim	RWTH Aachen University

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Qin	Hong	Stony Brook University
Quinlan	Michael	VISA
Radojkovic	Petar	Barcelona Supercomputing Center
Radulovic	Milan	Barcelona Supercomputing Center
Raicu	Ioan	IIT
Randall	David	CSU
Randles	Amanda	Duke University
Ranka	Sanjay	U. Florida
Rasquin	Michel	U Colorado Denver
Ravi	Vignesh	University of Arizona
Reddy	Narasimha	Texas A&M University
Reed	Daniel	University of Iowa
Renambot	Luc	UIC
Reteesh	Ranjan	GTech
Richard	Olivier	LIG, France
Ricker	Paul	University of Illinois at Urbana-Champaign
Rivera	Axel	Intel
Robert	Yves	INRIA
Robu	Bogdan	Gipsa, France
Roche	Kenny	Pacific Northwest National Laboratory
Rodrigues	Arun	Sandia National Laboratories
Rogers	Michael	Univ. of Nebraska Lincoln
Rokhsar	Daniel	JGI/UC Berkeley
Romps	David	UC Berkeley
Rosales-Fernandez	Carlos	TACC
Roscoe	Timothy	ETH, Switzerland
Rosner	Robert	U Chicago
Rossen	Mary Beth	Pennsylvania State University
Roth	G.	NVIDIA
Roy	Amit	drchrono
Ruiz	Daniel	Barcelona Supercomputing Center
Rupp	Karl	Argonne National Laboratory
Rutten	Eric	INRIA
Saad	Yousef	U. Minnesota
Sabin	Gerald	RNET
Safro	Ilya	Clemson University
8 Salinger	Andy	Sandia National Laboratories
Saltz	Joel	Stony Brook University
Samaddar	Debasmita	Culham Centre for Fusion Energy UK
Samanas	Evan	University of Wisconsin
Samatova	Nagiza	NCSU
Samsel	Francesca	U Texas Austin
Sanders	Peter	KIT
Sandin	Daniel	UIC emeritus
Sargsyan	Khachik	Sandia National Laboratories

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Sarkar	Vivek	Rice University
Sasidharan	Anjana	Amazon
Satheesh	Maheswaran	AWE
Savoie	Lee	University of Arizona
Schatz	Michael	Johns Hopkins University
Scheibe	Tim	PNNL
Scheuermann	Gerik	Leipzig University
Schmeisser	Andre	Fraunhofer ITWM
Schreiber	Rob	Hewlett Packard
Schulz	Christian	KIT
Schulz	Karl	TACC
Schulz	Martin	TUM
Schwan	Karsten	Georgia Institute of Tech
Schwartz	Oded	Hebrew University
Scott	Stephen	Tennessee Tech University
Sehrish	Saba	FNAL
Semazzi	Frederick	NCSU
Sen	Koushik	UC Berkeley
Sen	Mrinal	University of TX, Austin
Sewell	Christopher	Intuitive Surgical
Shaw	Kevin	Naval Research Laboratory
Shen	Xipeng	NCSU
Shephard	Mark	RPI - Rensselaer Polytechnic Institute
Shet	Aniruddha G.	Intel
Shi	Weisong	Wayne State University
Shin	Jaewook	NVIDIA
Shneiderman	Ben	University of Maryland
Siegel	Stephen	Delaware
Silva	Claudio	NYU
Singh	Mohit	Georgia Institute of Tech
Sinha	Rish	Microsoft
Sisneros	Robert	NCSA
Situ	Ning	University of Houston
Skjellum	Anthony	Auburn University
Smirni	Evgenia	College of William & Mary
Smirnov	Dmitriy	Pomona College
Smith	Brian E.	IBM
Smith	Garry	Freelance
Smith	Lauren	OSSS
Smith	Sterling	General Atomics
Snir	Marc	University of Illinois at Urbana-Champaign
Snyder	Philip	General Atomics
Sommer	Friedrich T.	UC Berkeley
Son	Seung Woo	UMass Lowell
Song	Sukhyun	Google

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Soto Cambres	Adrian	Stony Brook University
Sovinec	Carl	Wisconsin
Spafford	Kyle	Accelerate Diagnostics
Springel	V.	Max-Planck Institute
Squyres	Jeffrey	Cisco
Srinivasan	Sanjay	Pennsylvania State University
Stützle	Thomas	Université Libre de Bruxelles , Belgium
Stadel	Joachim	U. Zurich
Staebler	Gary	General Atomics
Stearley	Jon	Sandia National Laboratories
Steffen	Bernhard	TU Dortmund
Sterling	Thomas	Indiana Univ.
Stern	Eric	FNAL
Stone	Christopher P	Computational Science and Engineering LLC
Stotzer	Eric	Texas Instruments
Strnadova-Neeley	Veronika	UCSB
Strout	Michelle	University of Arizona
Su	Chun-Yi	Virginia Tech
Subasi	Omer	PNNL
Sugiyama	Linda	MIT
Suziki	Kenji	Univ. Chicago
Szalay	Alexander	Johns Hopkins University
Tang	Kun	Temple University
Tang	William	PPPL
Taufer	Michela	U.Dell.
Teeters	Jeff	UC Berkeley
Teller	Patricia	University of Texas at El Paso
Terboven	Christian	RWTH Aachen University
Terpstra	Daniel	Retired
Teruel	Xavier	Barcelona Supercomputing Center
Thapaliya	Sagar	Intel
Thottethodi	Mithuna	Purdue University
Tierney	Brian	ESnet
Tiwari	Ananta	San Diego Supercomputer Center
Tiwari	Devesh	Northeastern University
Toledo	Sivan	Tel Aviv
Tomov	Stan	U. Tennessee
Torrellas	Josep	University of Illinois at Urbana-Champaign
Trac	H.	Princeton
Trujillo	Andrew	UCSF
Trystram	Denis	LIG, France
Tufo	Henry	University of Colorado
Tuminaro	Ray	Sandia National Laboratories
Turton	Terece	U Texas Austin
Tynan	George	UCSD - University of California, San Diego

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Tzeferacos	Petros	U Chicago
Uhlemann	Kai	Coca Cola, Germany
Ullrich	Paul	UC Davis
Umar	Mariam	Virginia Tech
Unat	Didem	Koc University
Unsal	Osman	Barcelona Supercomputing Center
Utke	Jean	Allstate
Valeev	Ed	VT
van der Pas	Ruud	Oracle
van de Pol	Jaco	Univ. of Twente
Van Hensbergen	Eric	ARM Research
VanWyk	Eric	UMN - University of Minnesota
Vandierendonck	Hans	Queen's University, Belfast
Vanella	Marcos	GWU
Varela	Carlos	RPI - Rensselaer Polytechnic Institute
Varshney	Promad	Syracuse U.
Vary	James	Iowa State
Vasnier	Jean-Charles	CAPS
Vijaykumar	T.N.	Purdue University
Vivodtzev	Fabien	CEA
Vuduc	Richard	Georgia Institute of Tech
Wagner	Michael	TU Dresden
Wahib	Mohamed	Tokyo Tech
Walther	Andrea	Paderborn
Waltz	Ron	General Atomics
Wang	Bei	Princeton
Wang	Daifeng	Stony Brook University
Wang	Liqiang	University of Wyoming
Wang	Siwei	Hebrew University
Wang	Zheng	Lancaster University
Ward	Lee	Sandia National Laboratories
Ware	Colin	UNH
Ware	Doreen	Cold Spring Harbor Laboratory
Weaver	Vincent	University of Maine
Weber	Matthias	Technical University Dresden
Wei	Xie	Texas Tech University
Weide	Klaus	U Chicago
Weinberger	Shmuel	U Chicago
Weiner	Donald	Syracuse U.
Weissman	Jon	UMN - University of Minnesota
Welch	Aaron	University of Houston
Wheeler	Mary F.	University of TX, Austin
White	M.	UC Berkeley
Whitlock	Brad	Intelligent Light
Widener	Patrick	Sandia National Laboratories

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Willemberg	Jim	Sandia National Laboratories
Williams	Sean	UC Davis
Wilson	Paul	University of Wisconsin - Madison
Windus	Theresa	ISU - Iowa State University
Winslett	Marianne	University of Illinois at Urbana-Champaign
Wirth	Brian	University of Tennessee
Wisniewski	Robert	Intel
Wolf	Felix	TU Darmstadt, Germany
Wolf	Mathew	Georgia Institute of Tech
Wolfe	Glenn	RNET
Wolfe	Michael	NVIDIA
Wolfe	Noah	Rensselaer Polytechnic Institute
Wong	Michael	Codeplay
Worley	Patrick	Retired
Wu	Xingfu	Texas A&M University
Xie	Yusheng	Baidu
Xu	Donghua	Oregon State University
Xu	Jin	Stony Brook University
Xue	Ji	College of William & Mary
Yalamanchili	Sudhakar	Georgia Institute of Tech
Yamazaki	Ichitaro	UT Knoxville
Yan	Yonghong	Oakland University
Yang	Carl	UC Davis
Yang	Jiyan	Stanford University
Yao	Shun	Google
Yao	Yushu	Salesforce
Yeung	Pui-Kuen	Georgia Institute of Tech
Yi	Qing	University of Colorado at Colorado Springs
Yildrin	Esma	Louisiana State University
Ynnerman	Anders	Linköping University
You	Yang	UC Berkeley
Young	Jeffrey	Georgia Tech
Younkin	Timothy	University of Tennessee
Yu	Dantong	New Jersey Institute of Technology
Yu	Hongfeng	University of Nebraska-Lincoln
Yu	Weikuan	Florida State University
Yuan	Xin	University of Arizona
Zambreno	Joe	Iowa State
Zhang	Kunpeng	UIC
Zhang	Zhao	UC Berkeley
Zhao	Dongfang	IIT
Zheng	Fang	IBM
Zheng	Manchun	Pure Storage Company
Zheng	Yili	Google
Zhou	Bowen	Purdue University

Continued on next page

Table 2 – continued from previous page

Last Name	First Name	Institution Name
Zhou	Xiaobing	Hortonworks
Zhu	Qian	Accenture
Zhu	Zhigang	City University of New York
Zima	Hans	Independent
Zingale	Mike	Stony Brook University
Zirkel	Timothy	MITRE Corporation
Zivanovic	Darko	Barcelona Supercomputing Center
Zola	Jaric	University of Buffalo, NY
Zouridakis	George	University of Houston
Zubair	Mohamed	ODU

Table 3: Budget by Institution (in $K). Roles include Application Engagement (AE), Data Understanding (DU), Platform Readinesss (PR), Scientific Data Management (DM), and Project Management (PM). Note, all lab budgets are **inclusive** of any indirects associated with the administration of subcontracts. Subcontracts will be managed by the associated lab. [1]subcontracted by ANL. [2]subcontracted by LBNL. [3]subcontracted by ORNL

Institution	**Institutional PI**	**Role(s)**	**Budget**			
			Year 1	Year 2	Year 3	Total
ANL	Robert Ross	PM, AE, DU, PR, DM	1,156	1,180	1,228	3,564
LBNL	Leonid Oliker	PM, AE, DU, PR, DM	1,664	1,718	1,774	5,155
ORNL	Scott Klasky	PM, AE, DU, PR, DM	1,345	1,395	1,446	4,186
BNL	Shinjae Yoo	DU	150	150	155	455
LANL	James Ahrens	DU	300	309	318	927
LLNL	Bronis de Supinski	AE, DU, PR	400	418	436	1,254
Delaware[3]	Stephen Siegel	PR	80	82	84	246
HDF Group[2]	Neil Fortner	DM	141	141	141	424
Kitware[3]	Berk Geveci	DU	104	107	110	320
Northwestern[1]	Wei-keng Liao	DU, DM	120	123	127	371
Ohio State[3]	Han-Wei Shen	DU	96	100	103	299
Oregon[3]	Kevin Huck	AE, PR, DM	240	247	257	744
Rutgers[3]	Manish Parashar	DM	96	99	102	297
Utah[3]	Mary Hall	PR	109	110	112	332
Total			6,000	6,180	6,394	18,575

Table 4: Budget by Primary Topic Areas (in $K)

Primary Topic Area	**Area Lead**	**Budget**			
		Year 1	Year 2	Year 3	Total
Application Engagement	Anshu Dubey (ANL)	1,854	1,910	1,976	5,739
Data Understanding	E. Wes Bethel (LBNL)	1,837	1,892	1,958	5,687
Platform Readiness	Jeffrey Vetter (ORNL)	1,116	1,149	1,189	3,455
Scientific Data Management	Scott Klasky (ORNL)	980	1,009	1,044	3,034
Management		213	219	227	659
Total		6,000	6,180	6,394	18,575

Note, this table reflects a conservative estimate that 20% of personnel effort within the technology-focused primary topic areas (i.e., Data Understanding, Platform Readiness, and Scientific Data Management) will be spent performing Application Engagement. The model for this interaction is dicussed in Section 2.1.

Contents

1 Introduction

The Department of Energy (DOE) has led the world in enabling large-scale high-performance computing as an indispensable tool for scientific discovery in a broad range of disciplines. Our newest generation of systems is not simply larger than past systems - it brings distinct and novel challenges. These challenges arise from ever-increasing core counts on nodes; the deepening of the memory hierarchy and inclusion of fast, nonvolatile storage within the HPC system; the widening ratio of the FLOPS to I/O bandwidth; and the increasing ubiquity of computation accelerators. These advances in technology create new opportunities for science but also require adaptation of codes to capitalize on them. The data produced by the simulations on these high-end systems has also increased by orders of magnitude in size and complexity. This continuing explosion of data mandates constant attention and improvement to the tools and techniques used to manage and analyze this data. Moreover, the breadth of science performed on DOE advanced computing resources is growing well beyond that captured in the Seven Dwarfs [1] (i.e., dense and sparse linear algebra, spectral and n-body methods, structured and unstructured grids, Monte Carlo). In conjunction with traditional scientific computation, new motifs of investigation are emerging, including those involving analysis of experimental and observational data. To address these challenges, we propose RAPIDS: a SciDAC Institute for Resource and Application Productivity through computation, Information, and Data Science.

The objective of RAPIDS is to assist Office of Science (SC) application teams in overcoming computer science and data challenges in the use of DOE supercomputing resources to achieve science breakthroughs. To accomplish this objective, the Institute will:

- Solve computer science and data technical challenges for SciDAC and SC science teams,
- Work directly with SC scientists and DOE facilities to adopt and support our technologies, and
- Coordinate with other Institutes and DOE computer science and applied mathematics activities to maximize impact on SC science.

RAPIDS will be organized around four primary thrust areas. The cornerstone of the project is the application engagement and community outreach thrust area. A rigorous plan for engagement through direct interactions such as "tiger teams" as well as outreach to connect with the broader community will allow us to have the greatest and most immediate impact. Software technologies for computation, information, and data science are central to the success of SciDAC and SC science on leadership computing platforms. Our three remaining technology thrust areas — data understanding, platform readiness, and scientific data management — focus on key challenges in these areas, bringing to bear mature technologies and enhancing these technologies to adapt to new science requirements and new platforms. Technical thrust area participants will engage directly with application teams to leverage these technologies and to glean insights on how to further improve these tools.

Activities during the three-year duration of the Institute will focus on the current and upcoming generation of advanced computing platforms at the DOE facilities. We will build on our prior successes in fostering relationships with the facilities to further improve how we develop, test, and deploy software in these environments. State-of-the-art techniques in software development will be applied so that the software developed and deployed meets the high standards needed to ensure the correctness and performance of DOE science codes.

Our Institute brings together six DOE Laboratories and eight key industry and university partners with a broad range of relevant expertise and strong history of success in engagement with DOE scientists. The team is also tightly woven into the SciDAC community: Institute members are participants in thirty SciDAC4 Partnership proposals. Although the participating members in this Institute have each made significant contributions in specific areas of relevance to SciDAC and SC science, the team also brings particular and complementary strengths (e.g., performance engineering, data analytics, scientific data management). The collaborative structure of the Institute will encourage and enable the cross-fertilization and synergy

necessary to make possible the technological advances described in detail in the following sections.

Our Institute members have a significant track record of successful collaboration across large-scale SciDAC projects over the previous 15 years. Additionally, the team has demonstrated a strong record of software development, deployment, and support. Team members have played key development roles in many software packages used routinely by DOE scientists today, including the R&D 100 award-winning HDF5 (2002), VisIt (2005), FastBit (2008), and ADIOS (2013). Other notable examples such as PAPI, TAU, ROMIO, Paraview, Darshan, and VTK are all similarly integrated into the DOE HPC software ecosystem. The team will be proactive in modifying many of our tools to suit the needs of our application partners and to be ready for the LCF and NERSC platform changes occurring now and in the near future. We will also work together to combine technologies where complementary capabilities exist.

2 Institute Activities

Institute activities are centered around four primary thrust areas: application engagement and community outreach (§2.1), data understanding (§2.2), platform readiness (§2.3), and scientific data management (§2.4). The first focuses on how we interact with application teams, ensuring these interactions are methodical, thorough, and productive. The last three of these are *technology thrust areas*: they provide a toolbox of advanced computation, information, and data science technologies that address common challenges of strategic importance to the success of SciDAC and DOE science on HPC platforms. Table 5 provides additional details on the activities in all these areas. Of note, our data understanding thrust area includes machine learning technologies. Advances in that field and emerging needs in SciDAC and SC science motivate an increased investment in these tools.

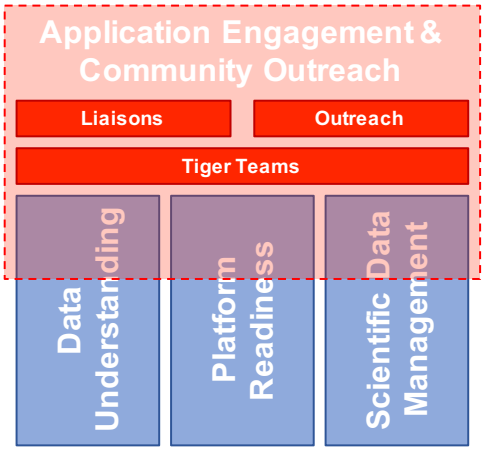


Figure 1: Vision of RAPIDS' engagement model showing matrixed thrust areas, tiger teams, liaisons, and community outreach.

We present these activities separately, but team members will routinely cross thrust area boundaries. In particular, team members in the technology thrust areas will engage with applications to deploy their technologies in service of science: *we anticipate members of the technology thrust areas will spend a minimum of 20% of their time in application engagement activities*. Figure 1 reflects this aspect of the organization.

2.1 Application Engagement and Community Outreach (Area Leads: Dubey, Williams)

Engagement with application teams is a cross-cutting activity across the entire Institute as shown in Figure 1. The guiding principle for engagements will be enabling access to expertise within RAPIDS to enable application teams to realize the highest possible scientific impact. We will holistically evaluate application pain points and work to address them by matching Institute expertise to the needs of the team. The level of engagement and resources allocated will be determined by a combination of factors, including application need, potential for impact, and available resources. An important part of this activity is also promoting what RAPIDS has to offer application teams. We will involve the technology thrust areas in information dissemination, evaluating, planning, and carrying out all engagements, including those with the broader community.

2.1.1 Application Engagement (Participants: Bernholdt, de Supinski, Dubey, Mohror, Roth, Williams)

We envision two modes for inter-Institute and application engagement: one where we "matrix" in members from RAPIDS' three technology thrust areas to collaborate with the application team; and another involving dedicated "tiger teams" of experts who will work actively with the application team developers. The engagements are expected to be short, 6–12 month collaborations with a specific objective[1]. Our application engagement measure of success is a demonstrable positive impact either on an application's performance or on an application development team's productivity. For brief engagement activities, this impact may manifest as a collection of recommendations regarding modifications to the application's implementation or its development practices, along with a prediction regarding the benefit to the application if those recommendations are adopted. For longer-term engagements that allow us to collaborate with the application team to realize our recommendations, performance and productivity measurements of the baseline and modified software and development practices will demonstrate our impact on the application project.

Matrixing RAPIDS Technologies and Applications: Tools and technologies have historically faced challenges in garnering application adoption as application stakeholders can become bewildered by the myriad of technologies and can lack sufficient knowledge on emerging tools to allow for rapid integration and evaluation. To avoid this pitfall, developers in the RAPIDS technology thrusts will be interfaced ("matrixed") with applications in order to integrate our technologies into the target application. These interactions will also provide feedback on RAPIDS technologies, allowing the team to identify new opportunities for improvement that may lead to refinements of the Institute development plans. Additionally, RAPIDS researchers may engage with application teams to investigate new programming models for their codes and ensure a robust dialog between the users and developers of these tools. Many of these technologies — including ADIOS, FastBit, HDF5, PAPI, PnetCDF, ROMIO, Paraview, TAU, VisIt, and VTK — are part of the programming environment on DOE leadership systems and will be part of our matrixing activities.

Tiger Teams: There are many applications, particularly those in emerging areas of importance to ASCR and the DOE, that need access to experts in RAPIDS. To that end, we will form "tiger teams" as needed with specific targets and objectives, including software engineering. For example, a team of performance engineering experts (total of 1 FTE) might spend 6–12 months addressing the computational and data challenges of a specific application or a solver. Tiger teams will leverage best practices in standards-based programming models to not only realize high-performance, scalability, and portability across DOE's leadership computing platforms (NERSC's Cori, OLCF's Titan/Summit, ANL's Mira/Theta, and LLNL's Sierra), but also ensure the application developers are well-versed in both best practices and fundamental insights into the ultimate performance bottlenecks within their applications.

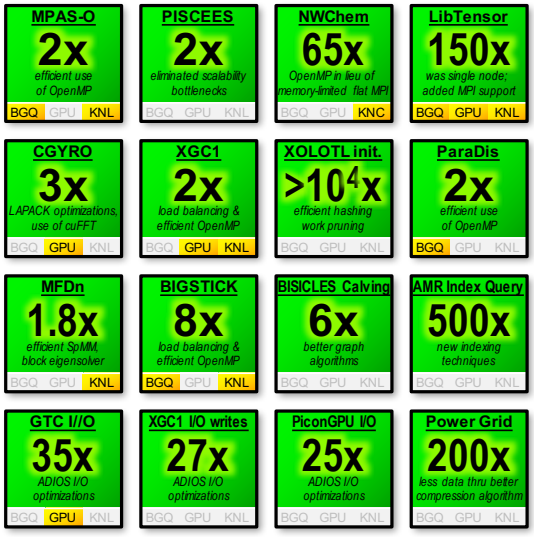


Figure 2: SUPER and SDAV performance accomplishments highlighting application partnership, speedup, and architecture awareness.

The SciDAC3 Data Institute (SDAV), Computer Science Institute (SUPER), and their predecessors, all used this model to great avail, collaborating with SciDAC part-

[1]Given funding constraints, we expect that collaborations requiring longer-term efforts can be realized only if funded as an augmentation to the application project.

nerships as well as the SciDAC3 FastMath Institute to maximize performance, scalability, and I/O throughput on multicore, manycore, and GPU-accelerated DOE systems. Often, SUPER and SDAV attained a 2× speedup at scale, and attained speedups exceeding 10×, 100×, and 10,000× when allowed to implement progressively more aggressive and algorithmic changes (see Figure 2).

Management: We expect to find far more potential application partners than funding (including matrixing) can support. To that end, we will establish a clear set of guidelines for evaluating potential engagements and impose a management structure designed to maximize the effectiveness of matrixed developers and tiger team members. We will begin by surveying the broad computational and data needs of the SciDAC4, BSSD, EFRC, and HPC4mfg applications, as well as any SciDAC4 Math Institute. We will incorporate data from facilities as part of the survey. Based on the survey, members of the relevant RAPIDS thrust areas will be informed of potential users of their technologies and any technological gaps that RAPIDS thrust area development can remedy. This survey will also identify potential candidates for tiger team activities. With appropriate feedback from the thrust areas and tiger team members, the RAPIDS application engagement team will form an initial set of RAPID-application collaborations.

Once we have selected applications, we will devise a plan for needed resources and a timeline. A member of the RAPIDS application engagement team will be assigned to monitor progress on the collaboration in conjunction with the relevant developers and RAPIDS thrust area lead. This will ensure the proper resources and personnel are dedicated on both sides, as well as provide a point of contact for Institute management. Upon the conclusion of each 6–12 month activity, the matrixed team members and application partners, in conjunction with the application engagement team, will write a short report for DOE summarizing the collaboration, how it enhanced the usage of DOE computational resources, any new science enabled, and any longer-term benefits that a potential funding augmentation might enable.

We expect these guidelines to pay dividends, whether the engagement is through matrixing or with a tiger team, as spending some effort in planning for engagement will maximize effective utilization of resources from the Institute as well as the application team. We will build enough flexibility into the process that specific engagements can be extended or customized if the need arises.

2.1.2 Capabilities

In addition to the expertise offered by the technology thrust areas, we will bring strong performance engineering and other capabilities to bear in support of SciDAC and DOE science teams. Performance engineering—traditionally a strong component of application engagements—will benefit the most from the "tiger team" mode of engagement, in which members will focus on using standards-based programming models, performance instrumentation, performance modeling, and best practices for KNL-based and GPU-accelerated systems such as Cori and Summit, respectively, to improve performance and scalability. Additionally, because of increasing heterogeneity in platforms, the platform readiness thrust (§2.3) will track and recommend solutions that are performance portable across architectures to preserve investments in refactoring and optimizing applications. Better software productivity is an emerging concern with regard to the quality of both software and the results generated with them [2–4]. We will promote effective software engineering practices such as verification and testing, lifecycle management and reproducibility through both direct engagements and outreach activities. We will also assist applications in devising software process customized for maximizing their team's productivity [5–7]. Based on application needs, we can facilitate the evaluation of alternative implementation strategies; assist in the design and implementation of resilience strategies; and help address the challenges of coupling codes, at multiple levels [8–10]. These capabilities connect to ASCR research programs (X-Stack, Resilience, and Extreme-Scale Operating Systems and Runtime); and code coupling, in particular, might motivate collaborative interactions with an institute focused on the mathematical side of coupling.

2.1.3 Community Outreach (Coordinator: Bernholdt)

In addition to focused engagements with specific applications, RAPIDS will have a program of outreach to the broader computational science and engineering (CSE) community. We envision a modest program of training events, workshops, and documentation of best practices that will bring the tools, libraries, and experiences of the RAPIDS team to a broader audience. Although our offerings will be widely useful, we expect to make a special effort to reach early career researchers who are in the formative stages of learning both HPC and CSE. The Application Engagement focus area will have responsibility for coordinating these activities, but they will be staffed from throughout the Institute according to expertise and interest. Plans for specific outreach activities will be based on progress in the RAPIDS thrust areas (e.g., the release of significant new capabilities), requests from application teams and the community at large, and the targeted venues. The following are broad examples of some of the various Outreach activities we anticipate organizing.

Training, Tutorials, and Webinars provide key opportunities to disseminate information about RAPIDS technologies and approaches. Potential topics include Roofline, ADIOS, PnetCDF, OpenMP, Graph Analytics Building Blocks, VisIt, Software Best Practices, and many others. Webinars can be offered jointly with the ASCR computing facilities. The recently-launched Exascale Computing Project (ECP) training webinars [11] have been attracting hundreds of registrations per event, indicating both the levels of interest and the potential to reach a sizable audience while avoiding travel costs and more extensive time commitments. Webinars have the further advantage that they are easily recorded and made available for later use. Many larger conferences, such as Supercomputing and SIAM Computational Science and Engineering, include a tutorial track that accommodates tutorials with 3–6 hours of content [12–14]. Other types of training events can also be considered—for example, project meetings or institutional settings—as well as providing content for events organized by others, such as the Argonne Training Program on Extreme-Scale Computing (ATPESC) [15, 16], which is aimed at students and postdocs.

Best Practices Documentation is a strategy for capturing the extensive experience embodied in the RAPIDS team and making it available in a form that is easily digestible and actionable. Potential topics include, but are not limited to, software engineering to facilitate performance portability, design strategies for portable high-performance I/O, data curation and analysis workflows, and strategies for the creation of sustainable software. Best Practices documents help fill a gap: many aspects of the experience of developing CSE software in HPC environments are not captured in the traditional peer-reviewed scientific literature, and yet often have a significant impact on scientific productivity. We plan to partner with the ASCR computing facilities and the IDEAS-ECP project [17] to ensure broad awareness and dissemination of our Best Practices.

Hackathons and Coding Camps are just a couple of the names used for events designed to bring users and developers of technologies together over an extended period of time to work closely together toward goals like learning new technologies, or implementing them in a particular (application) code. Such events often last several days to a week. Earlier in the SciDAC program, the Center for Scalable Application Development and Software (CScADS) [18] offered a series of workshops in the summer that provide some of the inspiration for this aspect of our Outreach program. More recently, the ASCR computing facilities and others in the community have been organizing hackathons on topics such as GPU programming and data-intensive computing. Given the major effort and expense involved in such events, we would expect to organize no more than one per year. We envision that these events might have an educational component to them, with travel, room, and board (partly) covered for some participants.

2.2 Data Understanding (Area Leads: Bethel, Balaprakash)

A key part of the scientific process is gaining understanding from data generated by scientific simulations and experiments, as well as performance data obtained by code instrumentation or other methods. To facilitate understanding of the increasingly large and complex data being produced by DOE science projects, one dimension of our work will center on adapting and applying scalable methods for finding and analyzing

features of importance hidden in scientific data (§2.2.1). Since data understanding tools and pipelines by necessity must run on large-scale DOE HPC platforms, another dimension of our work will focus on the key infrastructure needed to simplify the use of these increasingly complex platforms via collections of tools for data understanding (§2.2.2). Finally, driven by the increasing complexity of scientific data, new analysis approaches using machine learning methods offer the potential to accelerate data-intensive scientific discovery; such methods also have application to HPC performance data analysis and modeling to enable better utilization of large-scale HPC platforms (§2.2.3).

2.2.1 Scalable Scientific Data Analysis (Participants: Bethel, Guo, Morozov, Peterka, Rübel, Steed, Weber)

Modern simulations and experiments generate large volumes of complex data. To resolve the natural phenomena accurately, the data are collected at high spatial and temporal resolution, the measurements are multivariate, and data often are generated as part of large ensembles. The work we propose is motivated by science challenges from several different disciplines: quantitative feature detection and analysis of large-scale structures in cosmological simulations of universe formation; gaining insight into unsteady flow field characteristics in atmospheric and oceanic modeling problems; coping with the challenges of new and complex data types like high-dimensional phase-space data; better understanding time-varying evolution of systems from materials science.

Approach and Preliminary Work: To understand complex data, scientists require robust, scalable methods capable of detecting high-level features that can be analyzed statistically, compared across simulations and experiments, and visualized interactively. Detecting such features is a challenging and computationally expensive task. For example, to evaluate uncertainty in vector fields, it is necessary to understand stochastic flow features [19, 20]. Efforts have been made to visualize and analyze uncertain steady flows [21–23], but feature detection in this setting remains both a theoretical and a practical grand challenge. Stochastic flow features are estimated from a Monte Carlo sampling stochastic particle advection through the flow field, an expensive operation because of the large number of particles and their load imbalance. Similarly, to understand topological features in scalar fields represented by adaptive simulation techniques, such as adaptive mesh refinement (AMR), it is necessary to extend the algorithms to correctly process the complicated topology of the AMR meshes. Simulations in six-dimensional phase space require robust analysis to understand the behavior of regions where particle velocities deviate similarly from a Maxwell–Boltzmann distribution. On the visualization front, human-in-the-loop information visualization and visual analytics techniques have been slow to impact scientific data analysis. Our existing tools [24, 25] focus on either multivariate or temporal analysis, but a combined approach is needed that scales to large data sets.

We will build on techniques developed in ASCR research projects and as part of the SciDAC3 SDAV Institute. We will tackle the data analysis challenges along three directions: by improving existing algorithms to process more diverse types of data; by hardening their underlying infrastructure; and by combining advanced analysis and visualization methods to bring the user into the data analysis loop.

Proposed Work and Impact: We will pursue the following algorithmic improvements. Using decoupled time dependence, adaptive Monte Carlo sampling, and surface density estimation, we will build robust stochastic flow maps for uncertain vector fields [20]. We will further process such maps to extract Lagrangian coherent structures and ridge surfaces to identify regions of uniform flow, features that can be readily understood by the domain scientists. To detect topological features in AMR data, we will build on our recent work on triplet merge trees [26] and distributed merge trees [27] to process irregular topologies at the common boundaries of different AMR levels. The improved merge tree computation allows us to process arbitrary graphs, in a streaming fashion, and to build trees for individual AMR patches that contain just enough information to identify global connected components with minimal communication. We will extend and integrate image analysis and topological techniques, such as contour trees, to automate the processing of atomic force

microscopy image time series data to derive quantitative measurements of growth and dissolution processes, molecular speciation, and forces. We will employ interactive visualization and clustering for segmenting phase-space simulations, where each location in physical space has an associated distribution of particle velocities, using similarity metrics for comparison of such distributions.

We will improve the underlying infrastructure, specifically, the DIY library [28], developed in the SciDAC3 SDAV Institute, by integrating it more tightly with the data decomposition used in the simulations. We will integrate the collective operations that preserve the locality of data with AMR decompositions. We will further add dynamic load balancing to address greater imbalance in simulations [29]. We will relax DIY's bulk synchronous parallel processing (BSP) model to allow overlapping computation and communication phases to improve processing of unpredictable computation patterns in particle advection schemes.

We will combine advanced visualization techniques with statistical data analytics to help the user better understand multivariate, time-varying data. We will begin with variants of the parallel coordinates technique to enable multivariate analysis, adding special visual encodings for time. The data visualizations will include embedded human interaction techniques to manipulate the views, as well as linked visualizations that emphasize statistical aspects of variables and make important associations visually salient to enable effective information foraging.

The proposed effort will improve the resolution, accuracy, performance, and efficiency of scalable scientific data analysis. Topological analysis of simulations with much higher effective resolution will be possible, matching today's finest simulations. Stochastic flow analysis will aid scientists in comprehending the accuracy in ensemble and downsampled simulation data sets. Performance will be augmented with accelerated and automated methods to analyze material interfaces and other features at extreme scale. Efficiency, both machine and human, will be enhanced through phase-space approximations of and improved human interaction with high-dimensional data. Improved methods and infrastructure will enable easier, faster, and more insightful analysis of large-scale, high-dimensional, uncertain, and dynamic data.

2.2.2 Visualization for HPC Platforms and Application Understanding (Participants: Ahrens, Brugger, Grosset, Patchett, Pugmire)

There exists a synergy and interplay among the increasing scale and complexity of DOE HPC platforms, the size and complexity of data produced by simulations that run on those platforms, and the tools and methods for gaining insight from such data. The extraction of understanding from the data generated on HPC platforms is dependent on analysis and visualization tools and techniques that are feature rich and can be efficiently run at scale.

There are currently several gaps in the capabilities of these tools and simulation codes for emerging HPC systems. First, HPC systems are being deployed with deep memory hierarchies to cope with the I/O bandwidth problem. This hierarchy includes main memory, burst buffers, off-node memory, parallel file systems, and archival storage. Middleware layers are being developed to address the movement and placement of data across this hierarchy (c.f. §2.4.2). And at present there are no scalable methods for the seamless integration of analysis and visualization across these layers of memory. Second, as described in §2.2.1, adaptive methods are becoming increasingly important. These methods require complex data and execution models and require special handling to ensure efficient execution. Finally, the increased power and heterogeneity of supercomputers creates a more complex relationship between simulations and hardware. The decrease in situational awareness makes it difficult for scientists to achieve optimal performance.

Approach and Preliminary Work: We will build upon the work done in the SciDAC 3 Institute (SDAV) and other ASCR-funded projects to provide solutions to the aforementioned problems. These tools include VisIt [30] and Paraview [31], which both depend on the Visualization Tool Kit (VTK) [32]. To address the massive hybrid parallelism that exists on nodes, VTK-m [33] has been developed to provide portable

performance for analysis and visualization algorithms. Middleware frameworks like ADIOS [34] have been developed that manage the movement of data across the memory hierarchy, and provide an infrastructure for operating on data using in situ and in in transit processing techniques. Code performance tools, such as TAU [35], HPC Toolkit [36], and Allinea [37], provide views into hardware performance and can be coupled with in situ tools like Libsim [38] and ParaView Catalyst [39] and frameworks, such as SENSEI [40] to obtain access to the simulation state. We will build upon the software from these projects to develop and deploy our solutions, and will provide outreach with science teams for integration into their simulations.

Proposed Work and Impact: First, we will work with science stakeholders to identify sets of common analysis and visualization operations that are frequently required in their scientific simulations and experiments. We will prioritize methods that span multiple scientific disciplines and focus on building blocks that can be used to support more complex operations. These sets of operations, which are sometimes called "services," will be designed to be interoperable with middleware components discussed in §2.4, such as ADIOS. This flexibility will allow execution of these services in a way that is best suited to the data layouts across the memory hierarchy. Key technologies we plan to leverage include VTK-m, which provides portable algorithm performance, and DIY [28] for off-node communication.

Second, we will provide improved support for simulations using adaptive techniques in VisIt. While the increase of computing power has allowed simulations to be performed at increasing spatial and temporal fidelity, movement of this higher-resolution data remains a significant bottleneck, and can slow down simulations. To support adaptive simulation techniques, efficient distributed data models are required. These data models need to be carefully designed and implemented in order to strike a proper balance between memory footprint and execution efficiency. Analysis and visualization algorithms in existing tools need be modified to make use of this data and execution model.

Finally, we will work to provide improved understanding of how the simulation code is interacting with the HPC hardware. Understanding this requires a variety of types of information, including details of the HPC system (e.g., architecture, memory, bandwidth, network); real time resource utilization provided by tools like TAU and Darshan; the real-time state of the simulation. Once information about these three components has been gathered, it needs to be linked together and presented in ways that provide situational awareness of the performance of the HPC system and the state of the simulation code. We will partner with the platform readiness thrust (§2.3) to identify the performance bounds and opportunities of the simulation code on the given HPC system by using the performance modeling tools described in §2.3.5, as appropriate.

The proposed work will enable much more efficient usage of HPC systems in the extraction of knowledge from scientific data. First, embedding key analysis and visualization methods into the layer that manages data movement will greatly reduce the performance bottleneck caused by the I/O imbalance. This will also allow us to more directly leverage the work being done by middleware and runtime systems that will execute these tasks in ways that minimize various metrics, including data movement, time, execution environment, data size, and so on. Second, it will enable scientists to visualize and analyze their data on the newest systems at DOE supercomputing facilities using ParaView and VisIt, which are the primary parallel visualization and analysis tools used throughout the SciDAC community. And finally, it will provide scientists with a variety of performance data associated with the simulation that will help them understand how the HPC hardware is impacting the performance of their simulations.

2.2.3 Machine Learning for Application and Performance Data (Participants: Balaprakash, Buluc, Liao, Patton, Yoo)

Extracting knowledge from data produced by DOE applications has become a significant hurdle for scientific discovery because of the structure, size, heterogeneity, and complexity of these data sets [41, 42]. In addition, achieving scalable performance on leadership-class machines is becoming a daunting task for application developers as a result of heterogeneous node architectures, deep memory hierarchies, power–

energy efficiency demands, resiliency, and I/O bottlenecks [43]. These difficulties are further exacerbated by new data-intensive applications that require algorithms exhibiting irregular data access patterns, sparsity, and poor scalability. To address the above challenges, we will focus on machine learning (ML), in particular deep learning (DL) [44], which is emerging as a critical driver for data-intensive scientific discovery [41–43].

Approach and Preliminary Work: We will support domain-specific ML/DL methods for data analysis in close collaboration with SciDAC application teams. In particular, we will coordinate with AE survey activities (§ 2.1.1) to identify potential SciDAC users and begin a dialog. We will partner with a subset of the SciDAC4 applications on utilizing our ML/DL software.

Adaptation of DL to scientific big data with minimal human intervention is a key challenge facing domain scientists and DL researchers. To that end, we will build upon our (semi-) automatic approach to create deep networks [45, 46] that are scalable on parallel computers. To optimize resource mapping and performance for applications running on leadership-class machines, our goal is to develop predictive ML/DL models using costs of computation, communication, and power–energy at different levels (node, group of nodes, and machine). We will build and expand on AutoMOMML [47]—an ML approach that we developed within SUPER—for modeling the single-node performance of linear algebra kernels and adapt it for load-balancing [48] and I/O performance modeling [49]. To scale ML/DL software, we will develop distributed training by leveraging the rapidly evolving open source software, using our work on communication-hiding algorithms in GraphBLAS that were designed for handling irregular data access patterns, sparsity, and overall scalability.

Proposed Work and Impact: We propose to develop domain-specific adaptations for several ML/DL methods, which are required across SciDAC applications. This includes statistical emulators (such as Gaussian processes [50]), sparse training data handling, (such as transfer learning [51]), active learning [52], and generative adversarial networks (GANs) [53], representation learning (such as autoencoders [54]), and manifold learning [55]. Moreover, we will reach out to application teams, survey key requirements, and prioritize the support and development of ML/DL methods. We will scale ML/DL on leadership-class systems using distributed training approaches. We will also work with the existing DOE facility-vendor partnerships to map data/model parallel and (a)synchronous distributed training frameworks (e.g., [56–60]) and new libraries (e.g., see [61–63]) on leadership-class systems. For automatic DL adaptation for different applications, we will apply our existing deep network creation approach [45, 46] and integrate distributed evolutionary algorithms [64, 65] that can exploit ASCR computing resources. Furthermore, we propose to leverage Google neural architecture search [66] and related reinforcement learning approaches [67] to generate model descriptions of deep neural networks and optimize them using efficient black-box hyperparameter search algorithms [68–70] .

To enable full application modeling and tuning, we will develop data-efficient learning by adapting unsupervised learning approaches, such as generative adversarial networks [53], for data augmentation, and dimension reduction approaches, such as autoencoders, [54] for learning useful domain-specific representations and low-dimensional embedding from the HPC data. For increased modeling capability, we will bring DL into AutoMOMML using open-source software packages [71]. In addition, we will repurpose DL approaches and leverage the distributed training infrastructure to be developed in this project. We will use the offline and online predictive models from AutoMOMML for model-based application tuning on leadership-class machines using case studies in PR, data management, and application engagement. This will include a range of application tuning scenarios such as assigning tasks to cores and accelerators, data movement in deep memories, topology-aware communication, load balancing, and I/O.

We propose to deploy GraphBLAS [63] for ML/DL workloads that implement linear-algebraic building blocks. We will focus on the novel concepts of GraphBLAS, such as masks, semirings, and accumulators, which are flexible enough to eliminate the abstraction penalty of using a software library. We have been

developing communication-avoiding parallel algorithms for key GraphBLAS functions [72, 73]. This high-level algorithmic approach seamlessly captures and efficiently harnesses both model and data parallelism in DL. The GraphBLAS approach is particularly advantageous for sparse input and outputs and will have impact on various SciDAC and BSSD applications that work with irregular data. We propose to develop parallel algorithms and software for GraphBLAS to accelerate a large portion of graph and ML workloads by at least an order of magnitude. We will leverage our Combinatorial BLAS library [74] as a starting point and comply with the GraphBLAS standard [75] that we co-developed.

With the proposed development of robust, high-performance ML/DL approaches, we will address a key challenge in enabling data analysis at extreme scales; doing so will enhance the ability of DOE research to understand extreme-scale data from simulations, experiments, and observations. Integrated ML approaches for predictive modeling and model-informed application tuning can significantly increase the utilization of leadership-class machines for DOE applications, and thereby the rate of scientific discovery. ML-empowered performance modeling is also critical for accelerating applications such as multiscale modeling and in situ analysis and visualization. Our work with GraphBLAS will boost productivity by readily providing scalable performance that is portable across a variety of modern computer architectures.

2.3 Platform Readiness (Area Leads: Vetter, Hovland)

As mentioned in §1, it is critical that applications be prepared to use the major new architectural features, such as GPUs or nonvolatile memory, on current and upcoming DOE platforms. These new features will often require that application software be redesigned in order to exploit each of these innovative capabilities while balancing portability, performance, accuracy, and correctness. To this end, *the **Platform Readiness** thrust has the overarching goal of engaging application teams to prepare their software for these increasingly complex architectures by providing expertise and tools for these new features and programming solutions*. In addition to using traditional performance analysis and engineering capabilities, PR will focus on four architectural trends emerging in contemporary DOE systems: (1) heterogeneous and manycore processing [76], (2) deep memory hierarchies, (3) nonvolatile memory [77, 78], and (4) multilevel interconnection networks. It is important to note that, at this time of rapid change in architectures, programming systems and tools—such as OpenMP, OpenACC, and CUDA—will also be actively changing to support these architectural features. To solve these challenges, the PR thrust must be agile and leverage the experiences and artifacts of ASCR programs, vendors, and the broader community. In fact, our preferred solutions may change based on software maturity and our experiences. More broadly, many of these new features differ across platforms, resulting in few performance-portable programming models. Therefore, the PR Thrust will track and recommend solutions that offer **performance-portable solutions** across architectures to preserve investments in refactoring and optimizing applications. Our engagement strategy will consist of several steps. First, we will use performance tools (e.g., TAU, Roofline, NVPROF) to understand and benchmark the performance of existing applications across multiple existing platforms. Second, based on this analysis, we will prioritize components of the application that may benefit from heterogeneous or manycore computing. Third, we will attempt to introduce high-level approaches such as directives (e.g., OpenACC) or libraries to redesign the application for the target architecture for performance portability, while aligning with the application team's overall objectives. As a last resort, we will help to port applications to DOE systems using low-level programming models, such as CUDA [79].

2.3.1 Portable Programming Solutions for Heterogeneous and Manycore Computing (Participant: Lee)

Heterogeneous and manycore processors are becoming the de facto architectures for HPC systems. These architectures are drastically different from those of the past two decades (primarily X86-based processors), as well as from each other. A critical challenge of these architectures is programmability. Numerous active efforts are working to improve programmability/portability, including directive-based programming

models [80, 81], library APIs [82, 83], and parallel programming languages [84], among others. However, experiences with GPUs [85, 86] and KNL [87] illustrate that it is often difficult to obtain performance near that offered by specialized programming systems. Furthermore, because the architectures are changing so quickly, new programming systems often lack complete toolchains that offer insights into how to optimize performance.

Approach and Preliminary Work: The community has developed several portable solutions with varying levels of maturity and performance: OpenCL [84, 88], OpenACC [80, 85, 89, 90], and OpenMP [81, 91–93]. Library and meta-programming solutions (e.g., cuBLAS [94], Thrust [95], Legion [82], and Kokkos [83]) have also demonstrated promise. Despite the availability of these solutions, programming these systems may require significant architectural expertise; and the current solutions often lack performance portability.

Proposed Work and Impact: We will rely on a combination of software tools from vendors, the community, and this project to deploy and support the best contemporary solutions for applications teams. Consequently, we will continue to be actively engaged in activities, such as the OpenMP Architecture Review Board and the OpenACC Technical Forum, to understand and influence future options. Initially, our application porting work will use the vendor compilers, LLVM [96], and our OpenARC infrastructure [89]. Then, we will transition to the LLVM infrastructure [96] as LLVM support for target architectures matures. Interoperability of programming models will be critical at this time of rapid change; so we will develop tools that perform automatic translation of OpenACC into OpenMP offload directives (and vice versa), which allows interchange or intermixing of both OpenACC and OpenMP to maximize application portability. We will also prototype directive-based support for (1) architecture-specific features such as device-specific memories and (2) deep copy of hierarchical user data structures. Second, we will deploy our integrated MPI+OpenACC programming system [97] to (1) eliminate redundant communication and synchronization and (2) allow efficient load balancing across heterogeneous devices. Finally, we will continue to improve and extend LLVM support for heterogeneous computing, including contributing to Clang and Flang front-ends.

2.3.2 Managing Deep Memory Hierarchies Including Nonvolatile Memory (Participants: Perarnau, Vetter)

Deep memory hierarchies are a recent yet critical trend in HPC system architectures, as seen in recent DOE systems. These deep memory hierarchies require careful data orchestration for high performance. Additionally, platforms are adding massive amounts of nonvolatile memory to increase memory capacity and compensate for plateauing I/O capabilities [77, 98]. We will work with §2.4.1 to recommend strategies for applications. There are currently no portable tools, abstractions, or strategies for the application community to use to analyze and program these deep memory layers. In fact, nearly all of these technologies use their own interfaces, hidden from the application completely, or sometimes using vastly different modes, depending on boot-time parameters (e.g., cache or flat mode on KNL).

Approach and Preliminary Work: First, a number of tools have been developed to characterize and analyze memory access patterns to inform users regarding how to map their data structures into these deep memory hierarchies [99–101]. These tools can provide great insight into an application's memory behaviors, but the optimization and refactoring of these applications is left to the user. Second, compilers use advanced static analysis techniques [102, 103] or compiler directives [89, 90] to deal with deep memories in heterogeneous architectures, but no practical solutions currently exist for contemporary deep memory hierarchies like those in KNL. Finally, libraries and runtime systems can provide support to help users manage data either directly or indirectly; ORNL has developed DRAGON [104] to extend a GPU's managed memory to include massive amounts of nonvolatile memory using paging. As part of the Argo project, Argonne is building various scratchpad methods and low-level allocation APIs for explicit memory management [105]. For scalable nonvolatile memory, ORNL has developed a novel programming system, called Papyrus [106, 107],

that provides abstraction to access scalable, aggregate persistent memory in an extreme-scale HPC system.

Proposed Work and Impact: First, we will focus on integrating the Argo scratchpad system into library codes requiring explicit memory management strategies, so that applications transparently benefit from our performance improvements. We will then harden our implementation and port it to other deep memory architectures. Second, we will focus on developing portable static-analysis mechanisms on common programming models (e.g., OpenMP/ACC), prototype them using source-to-source translators like Rose and OpenARC, and then deploy them in LLVM. Specific optimizations will optimize data placement, focusing initially on the proposed OpenMP memory allocators mechanism. We will also explore the optimization of proposed extensions to support pipelining. Similar optimizations are likely to be applicable to mechanisms being considered within the OpenMP community to support deep copy of complex data structures. We will also develop optimizations that may not require OpenMP extensions, such as data layout changes, including tiling and array of structures (AoS) versus structures of arrays (SoA). Third, we will optimize Papyrus to better exploit RDMA features for more efficient communication. We will develop a Papyrus-based, MPI-IO implementation to allow existing MPI applications to transparently exploit advanced Papyrus features. We will also add additional interface capabilities as we learn, from our applications engagements, how applications are using Papyrus.

2.3.3 Enabling Application Performance on Highly Contended Networks (Participant: Iancu)

At scale, overall application performance is often determined by communication performance. Often, manual optimizations for communication overlap and message aggregation are required to hide network latency and amortize MPI overheads. Moreover, optimal performance execution requires dynamic adjustments to application-level message injection rates and message orderings to mitigate any variability in the network.

Approach and Preliminary Work: We recently explored a combination of message throttling [108] and reordering [109] techniques that greatly improved application performance at scale on InfiniBand and Cray Aries -based systems. For example, we have demonstrated a 75% end-to-end speedup for the GPU version of Nyx/BoxLib's HPGMG [110] gravity solver when running on 512 nodes on the ORNL Titan GPU-based system. We have also demonstrated a 40% speedup in the communication phases of NWChem [111] when running on 10,000 cores on Cori at NERSC; this translated into a 10% end-to-end speedup.

Proposed Work and Impact: As current tools provide little automated support, we will develop program analysis and transformation tools for communication optimizations in production codes. This work will result in a dynamic analysis toolkit built in LLVM that supports multiple programming languages, including C, C++, and Fortran; supports multiple communication paradigms, including 1-sided and 2-sided MPI; performs guided transformations on codes, and provides developers with source level feedback on transformations. We will also provide a runtime for MPI applications able to perform on-the-fly message reordering to adapt to variability in bandwidth availability and transient congestion. We will work with applications to ensure this technology mitigates their communication bottlenecks at scale.

2.3.4 Code Generation and Autotuning (Participants: Hall, Norris)

The growing diversity and complexity of architectural features makes writing optimal implementations challenging. Moreover, manually optimized implementations are typically not performance portable among architectures, requiring application developers to maintain multiple code versions. One solution is to use optimized libraries through high-level APIs when possible. When no library exists, code generation and autotuning can perform aggressive optimizations while preserving performance portability and code readability.

Approach and Preliminary Work: Over the past decade, we have successfully applied autotuning to scientific applications and libraries. For example, we have developed optimization technology that automates the code generation of inspectors and executors for sparse matrix computations within the CHiLL compiler [112–116], and parallelization in the presence of runtime dependences [117]. In SciDAC3, a manually optimized Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG) solver developed by FASTMath and SUPER consisted of more than 2000 lines of code. From a 7-line input code, CHiLL generated code that outperformed the manually tuned code by 3% [116]. In some cases, it is possible to use code generation and autotuning to avoid indirect memory accesses altogether; one such case is our Orio-based generation of optimized CUDA implementations for stencil-based computations, which significantly outperform manually tuned or vendor library implementations [118–120].

Proposed Work and Impact: We plan to apply code generation and autotuning, including the sparse matrix optimization technology, to SciDAC applications and extend our frameworks as required by these applications. For example, a number of additional transformations were added to CHiLL while LOBPCG was being optimized. We also anticipate developing new technologies, such as autotuning for deep memory hierarchies and nonvolatile memory (§2.3.2), parallelization strategies for new architectures, and enhancements to the inspectors we can generate for sparse matrix computations.

2.3.5 Identifying Performance Bounds and Optimization Opportunities (Participant: Williams)

Performance optimization is an iterative process requiring instrumentation, benchmarking, analysis, and prediction as optimizations are employed to eliminate bottlenecks. Application-facing tools must be straightforward to use and readily understandable. Moreover, having the same tool work not only across centers but also across architectures greatly reduces the burden on application developers. In addition to pointing toward effective performance optimizations, the Roofline Model helps identify application resource requirements and performance bottlenecks that can be addressed using our PR technologies.

Approach and Preliminary Work: SUPER developed the Empirical Roofline Toolkit (ERT) to benchmark memory, cache, and floating-point performance of manycore and GPU-accelerated systems and present the results in the context of the Roofline Model [121, 122]. In collaboration with NERSC, we developed a methodology using Intel's VTune/SDE to instrument applications and measure actual DDR and MCDRAM data movement on Cori's KNL nodes [123], which the NESAP project used to great effect [124, 125]. This proof-of-concept demonstration was followed by a more in-depth collaboration that resulted in Intel's integrating the Roofline Model into Intel Advisor [126]. This integration allowed applications to be instrumented, benchmarked, and analyzed in a GUI environment with cross-referencing among the Roofline, source code, and disassembly. We view this model of vendor collaboration and tool integration as highly attractive, as it produces robust, continually supported tools the centers are willing to install and support.

Proposed Work and Impact: We will conduct four Roofline-related development and integration activities. First, although Intel has been highly motivated to add Roofline support to Advisor, it will never support NVIDIA GPUs. To that end, we will collaborate with NVIDIA and OLCF to develop a methodology for Roofline-based analysis of application execution on GPUs to ensure application developers are presented with a common approach to performance analysis across DOE computing platforms. Second, we will expand the Roofline model and ERT to more accurately capture cache behavior on codes with complex locality patterns. Third, as SciDAC's application breadth expands to encompass new fields, including bioinformatics, the performance of integer-intensive kernels may become more important. To ensure programmers are cognizant of both the capabilities of their underlying machine and the properties of their application's execution, we will extend Roofline and ERT support for integer-dominated computations on CPUs and GPUs. Finally, we will extend our Roofline-based instrumentation and analysis to incorporate MPI communication using

benchmarking and performance counters (e.g. CrayPat Cori/Aries).

2.3.6 Program Correctness (Participants: Hovland, Lee, Siegel)

Our experience indicates that many DOE applications have as a goal to transform an underperforming, possibly sequential, component of a computational science workflow into a high-performance implementation with equivalent semantics. The new implementation might be written by hand, generated via a polyhedral or DSL compiler, or implemented in an alternative programming model such as OpenACC and OpenMP 4.x that maps better to contemporary computer architectures. In all cases, an important goal is to preserve semantics while improving performance ("first do no harm").

Approach and Preliminary Work: By combining static analysis and dynamic testing, we will help developers verify that application semantics are preserved as codes undergo refactoring, performance optimization, and porting to new platforms. For static analysis, we will use and extend the CIVL verification framework [127–129]. The CIVL translators convert programs written using MPI, OpenMP, Pthreads, CUDA, or a combination of these into the CIVL-C programming language. The CIVL model checker uses symbolic execution to determine whether race conditions, assertion violations, division by zero, or other safety violations could possibly occur in a CIVL-C program, or to verify that two CIVL-C programs are functionally equivalent. The ASCR-funded ProVESA project improved CIVL performance and added Fortran support.

To complement static analysis with dynamic testing, we have developed a directive-based, interactive program debugging and optimization framework for heterogeneous computing [130]. In this framework, the compiler generates several code variants, and the runtime system executes them to identify incorrect kernels and detect incorrect/missing/redundant memory transfers. Under ProVESA, we used statistical measures of the expected variability in numerical computations to provide thresholds for dynamic testing.

Proposed Work and Impact: For static analysis, we will extend CIVL to meet the needs of SciDAC and other DOE applications. Provisionally, we plan to improve Fortran language coverage and add support for OpenMP tasks and SIMD directives and MPI-3 remote memory access. For dynamic testing, we will enhance the OpenARC interactive framework by adding support for OpenACC2.x and OpenMP4.x and by improving support for handling asynchronous compute regions. Threshold determination in dynamic testing provides a potential point of collaboration with the potential SciDAC Math Institute.

2.4 Scientific Data Management (Area Leads: Klasky, Wu)

As we advance toward the exascale era, we must re-examine traditional interfaces for computation, I/O, and storage. The actual utility of a machine for large-scale simulation is increasingly limited by the growing mismatch between FLOPs and I/O bandwidth. Current approaches, oblivious to the context and intent of science users, largely write arrays of bytes, relying on said users to apply additional codes, scripts, and data analyses to interpret the outputs of their simulations. These approaches can quickly become inflexible and can significantly impact productivity as complexity and scales grow. Without further support from data management infrastructures, it falls to the scientists to resolve all the issues of maintenance of codes and data, as well as develop new approaches for advancing hardware capabilities.

The scientific data management team in this project will deploy and support efficient methods of moving data between applications and storage, and enable the reduction of data through compression, statistical analysis or other approaches. We will partner with the SciDAC and LCF application community to take software researched and developed by our other activities (e.g., ASCR research and ECP) and deploy those products at scale. SciDAC and LCF applications will doubtless require assistance in adopting these tools on the evolving set of hardware and software technologies, and we are well-positioned to provide that aid.

We will focus on three areas, chosen for their impact on applications. First, we will target high performance for *moving data to and from storage.* Data movement involves not just using traditional I/O, such as

MPI-IO (ROMIO), but also higher-level APIs, such as ADIOS, Parallel-NetCDF, and HDF5. These libraries provide an opportunity for data reduction and access pattern transformations in the I/O layer. Achieving high performance involves ensuring those I/O transformations are well tuned to both the application and the hardware/software architecture. Second, we will work on efficient in situ infrastructure to *couple codes* together, which may involve multi-physics application-to-application connections or may involve analysis and visualization applications (§ 2.2). Third, we will work on ways to *manage data across the entire life cycle* of an application's data, to enable a broader concept of a knowledge life cycle. All of our solutions will be co-developed with SciDAC applications through our tiger-team activities.

2.4.1 I/O and Storage (Participants: Forner, Huck, Klasky, Latham, Liao, Podhorszki, Sim)

A major challenge for applications running in the next three years at the LCFs and NERSC will be in extracting and preserving the output from simulations despite the fact that storage and I/O bandwidth has not kept pace with advances in computational throughput. This means that we must make the I/O performance of our tools (ADIOS, HDF5, Parallel-NetCDF and ROMIO) as efficient as possible on new technology that includes deep storage and memory hierarchies including nonvolatile memory (§ 2.3.2) and highly variable shared networks (§ 2.3.3). This problem is already impacting applications; and science teams are being forced to develop ad hoc data management solutions that, while they may be effective, are highly customized and application specific. We will combine many of our tools to monitor (Darshan, TAU), reduce [131–133], move (ROMIO, ADIOS) and create flexible I/O solutions (ADIOS, HDF5, Parallel-NetCDF).

Approach and Preliminary Work: To understand the I/O performance, we have developed the Darshan [134] I/O characterization tool. Darshan's characterization imposes little memory or performance overhead, and it has been enabled on many computer centers (including NERSC and ALCF) by default. The resulting statistics aid in understanding an application's I/O behavior and the corresponding storage performance. Furthermore, we have been engaged in integrating TAU performance data captured during a simulation with ADIOS to provide performance data as part of the routine output of a simulation for finer-grained performance understanding.

The ROMIO [135] MPI-IO implementation, part of most vendors' I/O software stack, provides a low-level I/O interface. Applications can use this interface directly or indirectly through higher-level I/O solutions. The libraries Parallel-NetCDF [136] and HDF5 [137], which provide data structures and interfaces well suited for scientific applications, leverage the MPI-IO low-level details in higher-level abstractions.

ADIOS [34] provides a high-level I/O interface while allowing the underlying implementation to transparently and leverage emerging hardware technologies. During past SciDAC institutes, ADIOS has provided very high I/O performance to applications running on Titan with the Lustre file system and on Mira's BlueGene/Q architecture with General Parallel File System. The impact from this work was highlighted in the FY 2018 Congressional Budget Request [138], which cited the example of ADIOS running at scale on the OLCF to improve the efficiency of simulations to improve fire protection standards across industry.

Proposed Work and Impact: We will work directly with SciDAC applications and emerging areas to address HPC I/O shortfalls. As applications grow in scale, they typically exercise our I/O libraries to overcome these shortfalls. New storage technologies such as burst buffers will help science applications achieve better I/O performance and will lead us to reconsider our I/O transformation strategy. The introduction of solid state, nonvolatile memory, and other persistent memory devices means that we will need to consider latency as well as optimization for bandwidth. We will focus on the burst buffers on Summit, Aurora, and Cori to further improve the I/O performance of applications. Through the institute survey (§2.1.1) we will identify several exemplar SciDAC applications and then, motivated by those applications' access patterns, will adjust or re-implement ROMIO and ADIOS optimizations to further improve performance. We will

collaborate with the researchers in § 2.3.2 to analyze data movement and to use analysis and programming tools for deep memory and nonvolatile memory, where these are beneficial to our efforts.

To understand new architectures and new applications, we will extend Darshan's functionality and coverage. Darshan currently has limited coverage of the Parallel-NetCDF, HDF5, and ADIOS interfaces; but modularity work done under prior funding lays the groundwork for easily extending Darshan coverage of existing APIs and new interfaces of interest. We will further enhance Darshan's statistics and characterizations. We will also partner with the TAU team to write performance data with ADIOS, and partner with our data understanding RAPIDS team to understand and visualize this data. The new storage hierarchy also has impact on Parallel-NetCDF, ADIOS, and HDF5 [139]. We will work with application scientists to debug and enhance performance on new hardware and to experiment with non-traditional storage layouts. Where needed, we will also develop external utilities to access or convert such layouts as standard formats.

New I/O patterns are also appearing as codes dive further into running uncertainty quantification (UQ) and data-driven I/O such as from ML techniques. We will optimize our I/O software for these new and emerging patterns. Furthermore, we will often utilize ML techniques to optimize the I/O for auto-tuning [140] and pre-fetching [141] in conjunction with work in § 2.2.3. Another focus will be to support and harden new data reduction techniques, which are being added to ADIOS and HDF5, from various research groups and to help applications use them for their data reduction needs. Recently added examples are the lossy compression techniques ZFP and SZ that are ready for application adoption. These techniques take advantage of the continuity in data. We will explore using more advanced statistical techniques for data reduction, as well as use advanced numerical techniques, which create a Multigrid-like theory [132], to enable a quasi-optimal reduction technique for all scientific data.

Overall, we will monitor, optimize, and reduce the I/O from SciDAC simulations and read and write output using self-describing methods. We will partner not just with applications but also with any sibling Math Institute, and with the other efforts in RAPIDS, to provide a Data Management backbone that is well-tuned for applications on Summit, Theta, and Cori.

2.4.2 In situ Infrastructure for Coupling (Participants: Parashar, Podhorszki)

The newest generation of DOE computing resources will enable complex simulations in a range of application domains at unprecedented resolution. However, the ability to derive insights from these simulations will be limited by the capacity to address the data management challenges associated with coupled simulation and analytics workflows. These applications will require complex multi-physics / multi-resolution couplings between component simulations, including ensembles of simulations to quantify uncertainty, explore parameter spaces, and integrate complex data processing and analytics workflows.

These couplings are variously referred to as in situ, in transit, or other sorts of in-memory techniques for processing data through I/O workflows. For RAPIDS, we use the term *staging* to focus on the user's view that there are multiple stages for processing between the simulation's write operation and the ultimate consumption of the data. The user should not be concerned about how those stages are linked or executed beyond the absolute minimum necessary; the runtime needs to assist users in determining placement on cores and allocation of different types of memory (e.g., DRAM, nonvolatile memory).

Approach and Preliminary Work: Current ADIOS staging solutions, Flexpath [142,143] and DataSpaces [8, 144], facilitate data staging from compute nodes to auxiliary nodes on high-end systems. We have supported the SciDAC EPSI Institute in its efforts in code coupling as well as in situ analytics [145]. A key goal of the ADIOS framework development is to support in situ workflows at large scale, using data staging as a service to support new data access patterns for significantly unbalanced I/O. In this project, we will focus on creating a reliable staging framework on the next-generation machines, which applications can trust to make use of our in situ processing capability.

Proposed Work and Impact: Leveraging our existing in situ capabilities, we will support applications in their efforts at coupling codes together, or at building in situ workflows from simulation, analysis, and visualization tasks. Although these solutions support applications running on current DOE systems, we must reengineer them to work robustly and more efficiently on emerging systems with new memory technologies, architectures, and interconnects—ensuring that staging can take advantage of storage resources both *on and off node*, leverage accelerators for processing data while in staging, and address concerns related to resilience and energy efficiency. The complexity of in situ workflows is addressed by decomposing the tasks into discrete, simple services that are easy for a user to comprehend by reasoning. We then monitor and capture the provenance and performance information so that it may be stored alongside the analyzed scientific data, maintaining a record of what the costs are for each of the analytics and reduction actions.

We will collaborate with the PR thrust on analyzing and optimizing coupled codes for heterogeneous architectures by using the solutions described in § 2.3.1 as appropriate. We will also collaborate with the Data Understanding thrust for the development of analysis and visualization services described in § 2.2.2. We will also partner with the potential SciDAC Math Institute and the Data Understanding group in RAPIDS to ensure that our infrastructure can be used together with their executables/services.

2.4.3 Knowledge Management (Participants: Mehta, Wu)

We use the term "knowledge management" to cover all aspects of the data life cycle for the data generated by modern scientific applications and workflows. In addition to managing large data volumes data across the storage hierarchy, we need to address the broader issues associated with metadata. In the context of this work, we will focus primarily on the aspects that have been identified as critical to the SciDAC applications. These include provenance capture, including performance metadata; feature extraction; metadata organization; and metadata indexing and querying.

An emerging trend in scientific computing is analyzing data before storing it on persistent media. Coupled codes and workflows require complex data movement among the different computational processes, with many intermediate data elements. It is important to capture the metadata and provenance of these analysis processes, in order to be able to optimize them and make them reproducible.

Approach and Preliminary Work: A number of different strategies have been devised to collect metadata and provenance for data and workflows [146–150]. However, recording information about online workflows is challenging because of high levels of concurrency. We plan to target this key gap in supporting in situ processes. ADIOS is an in situ processing system, so it is a prominent target for incorporating provenance capturing. In the past, research efforts have leveraged performance metadata to improve in situ workflow performance [151], to enable active processing of data [152], and to enable optimizations for non-standard read patterns [153].

Once all of the metadata is collected, it is critical to find data in the collection which can be quickly discovered. Here we plan to use leverage our FastBit indexing technology to organize and index the metadata information [154]. We have demonstrated it to be effective in for scientific data [155,156], and will leverage this technology for indexing all of the campaign data.

Proposed Work and Impact: Advanced data management services that can span whole collections of runs and experiments require managing new and different types of metadata. Gathering such accurate and robust provenance and metadata is therefore an important challenge in scientific computing. The challenge is further exacerbated for experimental setups consisting of multiple, heterogeneous data sources. As part of this proposal, we will apply and tune our distributed campaign metadata management service(s) for recording valuable provenance and performance metadata about data and workflows of SciDAC applications. More specifically, we will enrich and extend the ADIOS API to provide users with a consistent API for creating and managing metadata and provenance information through a single interface. We will capture metadata

about data, as well as source codes that are used in the generation and visualization of data.

Metadata in present day scientific computing comprises multiple file formats, APIs, schemas, and ad hoc coding standards. Based on our experience with several such user communities, we will structure a community-driven activity to develop shared expectations, schemas, and ontologies, which our tools will then implement and support. Specifically, connecting with the ongoing efforts in understanding the performance of the code-coupling infrastructures will allow us to develop clear usage scenarios, as will engaging with other existing research efforts [157–159].

One key challenge in metadata management is the need to handle a large number of active processes simultaneously. This is critical not only in the data collection phase, but also in the data access phase where billions of computing threads might want to access the same information at the same time. Good organization and indexing will be key to supporting these operations efficiently. To that end, we will continue our development efforts on FastBit indexing and scientific data services. Another way to improve data access speed is to leverage faster storage technologies, such as burst buffers [160, 161]. We will expand on our preliminary work through targeted efforts that extend the work undertaken in §2.4.1.

Since many of the quantities that will be queried during a scientific campaign will not be stored, and often calculated on-the-fly, we will extend our indexing techniques to support the ability to index features inside of the data. Here we will extend FastBit for the task. In addition, some compression techniques such as IDEALEM [131, 162], captures distinct statistical properties that could be useful for data understanding. We will implement them in both ADIOS and HDF5 files, which can be queried during a scientific campaign.

3 Project Management Plan

Management. Figure 3 depicts the organization of RAPIDS. Rob Ross (ANL) will serve as the overall lead, with Lenny Oliker (LBNL) acting as deputy. Dr. Ross recently stepped down after serving as interim division director for the Mathematics and Computer Science Division at ANL and has served as the deputy for both the SciDAC2 and SciDAC3 Data Institutes, while Dr. Oliker was the recent director of the SciDAC3 SUPER Performance Institute. Together they bring many years of expertise in successfully managing large multi-institutional SciDAC projects and ASCR computer science and data teams. Their oversight roles will include setting the overall project direction, overseeing project progress and meetings, preparing reports and reviews, connecting to other SciDAC4 institute(s) and application partnerships, and representing RAPIDS in various community activities. We plan to hold coordination and reporting teleconferences with ASCR management every month, or more frequently if preferred by the program management. Annual reports will detail activities and identify institutions and personnel responsible for outcomes.

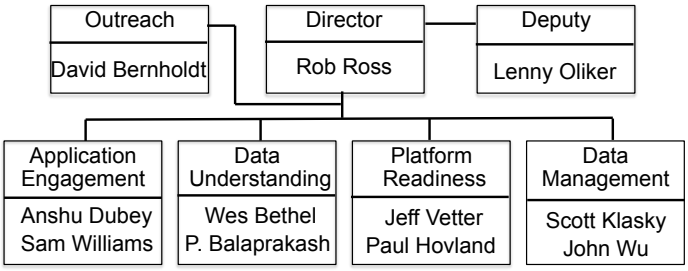


Figure 3: Overview of RAPIDS organizational chart.

Our thrust areas consist of Application Engagement led by Anshu Dubey (ANL) and co-led by Samuel Williams (LBNL), Data Understanding led by Wes Bethel (LBNL) and co-led by Prasanna Balaprakash (ANL), Platform Readiness led by Jeff Vetter (ORNL) and co-led by Paul Hovland (ANL), and Data Management led by Scott Klasky (ORNL) and co-led by John Wu (LBNL). These individuals are leaders in their respective fields with track records of successful research and application engagement projects within the DOE portfolio. Thrust area leads will coordinate the technical activities across our numerous participating laboratories, universities and vendor partners. Regular leadership teleconferences across all areas and insti-

Table 5: RAPIDS deliverables and milestone for Application Engagement (AE), Data Understanding (DU), Platform Readiness (PR), and Scientific Data Management (DM).

Area	Task	Yrs	Milestones and Deliverables
AE	Tiger Tm & Matrix	1–3	Develop engagement methodology and process, engage tiger teams with applications, refine and document methodology based on lessons learned
AE	Tutorials	1–3	Assess tutorial needs, develop material, develop workshop concept, conduct tutorials and workshop, update materials, document outcomes
DU	Scientific Data Analysis	1	Prepare for new alg. (DIY & VisIt); extend analysis & visualization capabilities
		2	Extend infrastructure (collectives, relaxed BSP); demonstrate impact on user data
		3	Harden infrastructural changes, polish user interfaces, engage with applications
DU	High Perf. Comp.	1	Data model, algorithm development, and perf. integration for science partners
		2	Algorithm development and optimization for software stack & HPC platforms
		3	Outreach, deployment, optimization of visualization techniques w/ science partners
DU	Machine Learning	1	Develop ML/DL approaches & GraphBLAS parallelism for apps & data
		2	Implement ML/DL distributed training/inference approaches for scalability
		3	Demonstrate ML/DL approaches on applications data and performance tuning
PR	Hetero-geneity & Deep Memory	1	Evaluate and recommend APIs for deep memory for applications; optimize Papyrus for RDMA; evaluate OpenMP memory hierarchy & affinity prototype
		2	Demonstrate task-based GPU framework on apps; CUDA Scratchpad support; extend Papyrus MPI I/O support; harden OpenMP mem hierarchy & affinity prototype in apps
		3	Demonstrate integrated MPI+OpenACC framework on applications; extend Papyrus Template Container Library for additional applications data structures
PR	Modeling	1	Integrate GPU into Roofline with NVprof & collaborate with OLCF/NVIDIA
		2	Extend Roofline app. execution characterization for deep CPU memory hierarchy
		3	Extend to characterize integer-heavy codes, network capability & actual performance
PR	Correct-ness	1	Enhance CIVL support for Fortran
		2	Prototype support for OpenMP Tasks and SIMD in CIVL
		3	Prototype support for MPI-3 remote memory access in CIVL
PR	Optimize Comm-unication	1	Extend support for Static reordering of MPI operations
		2	Enhance Dynamic analysis to maximize overlap of MPI operations
		3	Enhance Dynamic analysis to maximize overlap for hybrid parallelism (MPI/OpenMP)
PR	Integrate	1–3	Integrate PR tools into applications
DM	I/O	1	Optimize I/O for burst buffers & HPC storage for ADIOS, HDF5, PnetCDF, ROMIO
		2	Extend ADIOS + PnetCDF for Darshan, w/data reduction, non-traditional I/O tuning
		3	Performance data integration from TAU into I/O solutions; ML for I/O optimization
DM	Code Coupling	1	Integrate ADIOS staging into applications using persistent data-staging (Dataspaces)
		2	Integrate Data Understanding Services into ADIOS staging for SciDAC applications
		3	Tune Staging Services with NVRAM for in-situ understanding with applications
3-4 DM	Knowl-edge Mgmt	1	Capture and Index data from simulations and in situ workflows
		2	Deploy querying services for KM, and integrate visualization and analysis capabilities
		3	KM software for data, analysis operations, and in situ and post hoc workflows

tutional PIs, in addition to area meetings, will ensure that Institute contributions are orchestrated effectively across the entire project. An annual "all hands" meeting will bring the team together in person.

Outreach. A key aspect of our project will be outreach activities that collaborate and synergize across the SciDAC portfolio and broader DOE community. David Bernholdt (ORNL) will coordinate these activities. Additionally, we plan to work closely with the future SciDAC4 Math Institute and ensure coordination and avoid duplicative work across the two projects. Samuel Williams (LBNL) has successfully participated on multiple previous SciDAC3 SUPER/FASTMath collaborations, and will act as the point of contact for inter-Institute interaction. We are aware of numerous SciDAC4 Partnership proposals: our team participated in thirty such proposals. The institute solicitation also declares an intention to fund a Math Institute, but at the time of writing final announcements of these projects have not been made. Once these teams are formed, we will immediately commence collaborative discussions, with Dr. Bernholdt leading our contacts with partnerships, and Dr. Williams engaging leadership at the expected Math Institute.

Another important aspect of outreach will be to interface with the ASCR computing facilities, and we have therefore identified liaisons for each of the three centers led by local personnel on our team: Kevin Harms (ANL) at ALCF, Oliver Ruebel (LBNL) at NERSC, and David Bernholdt (ORNL) at OLCF. Similarly, we plan to have Jim Ahrens (LANL) act as the liaison to the Exascale Computing Project (ECP). Dr. Ahrens will leverage his role as an L3 manager within ECP's Software Technology track to ensure successful interaction and technology transfer. We are also aware of DOE's base research funded activities; in fact, many of the technologies we propose to employ were developed or advanced under ASCR research projects. We are tracking other notable SC activities such as BSSD, EFRC, and HPC4Mfg. Numerous members of our team have significant experience with these communities and we will actively look for opportunities to build productive collaborations. To ensure effective communication, we plan regularly scheduled meetings between the liaisons and Institute leadership, with ASCR facilities and external stakeholders.

Flexibility. The vast experience of our leadership team in successfully driving previous large-scale SciDAC projects highlights our proven track record to evaluate and mitigate risk and to adapt our plans in response to SciDAC and SC needs. Our dynamic process of identifying and engaging with application teams and facilities inherently enables flexibility and risk mitigation, allowing teams to adapt to changing priorities and unforeseen technical challenges. As detailed in the proposal, the level of engagement and resources allocated will be critically evaluated and adapted in the context of application needs, potential impact, required resources, and ASCR priorities. Through the coordination activities listed above we will closely track the RAPIDS activities (Table 5) and quickly respond to new developments that affect our team's ability to deliver. Proposed changes to plan will be discussed with ASCR program management, and, if necessary, resources will be reallocated and task schedules revisited in the context of shifting priorities.

The organization and goals of the RAPIDS SciDAC4 Institute are based on the collective experience that our team members have acquired over more than 15 years of SciDAC participation. We have organized a broadly-based project with thrust areas organized to maximize our impact on the SciDAC and SC community. We have identified a key set of advanced computation, information, and data science technologies that address common challenges of strategic importance to the success of SciDAC and DOE science on HPC platforms. Additionally, we have designed a highly collaborative environment that will leverage and group key technologies across thrust areas to deliver significant contributions on current and forthcoming advanced computing platforms at the DOE facilities. While our focus specifically targets the next 3 years of the proposed period, our enabled technologies will continue paying dividends for many years to come.

Appendix 1: Biographical Sketches

As requested in the call, the comprehensive list of conflicts of interest is provided in Table 2.

James Ahrens

Education and Training

University of Washington, Computer Science, Ph.D. 1996
University of Washington, Computer Science, M.S 1992
University of Massachusetts at Amherst, Computer Science, B.S. 1989

Research and Professional Experience

Los Alamos National Laboratory	1997-present
Los Alamos National Laboratory	Summer 1993-1994
Thinking Machines Corporation	Summer 1991
General DataComm Incorporated	Summer 1987-1989

Selected Publications

1. J. Ahrens, "Increasing Scientific Data Insights about Exascale Class Simulations under Power and Storage Constraints", IEEE Computer Graphics and Applications (2015)
2. J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D. Rogers, M. Petersen, "An Image-based Approach to Extreme Scale In Situ Visualization and Analysis", Proceeding of International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), SC14 (2014)
3. J. Ahrens, B. Hendrickson, G. Long, S. Miller, R. Ross, D. Williams, "Data-Intensive Science in the US Department of Energy: Case Studies and Future Challenges", Computing in Science and Engineering (2011)
4. J. Ahrens, K. Heitmann, M. Petersen, J. Woodring, S. Williams, P. Fasel, C. Ahrens, C. Hsu,
5. B. Geveci, "Verifying Scientific Simulations via Comparative and Quantitative Visualization", IEEE Computer Graphics and Applications, 30(6) (2010)
6. sJ. Ahrens, B. Geveci, C. Law, "ParaView: An End User Tool for Large Data Visualization", Visualization Handbook, Academic Press (2005)
7. C. Mitchell, J. Ahrens, J. Wang, "VisIO: Enabling Interactive Visualization of Ultra-Scale, Time Series Data via High-Bandwidth Distributed I/O Systems", IEEE International Parallel and Distributed Processing Symposium (2011)
8. J. Ahrens, K. Heitmann, M. Petersen, J. Woodring, S. Williams, P. Fasel, C. Ahrens, C. Hsu, B. Geveci, "Verification of the Scientific Simulations via Hypothesis-Driven Comparative and Quantitative Visualization", IEEE Computer Graphics and Applications **30** Number 6 (2010)
9. E. Anderson, J. Ahrens, K. Heitmann, S. Habib, C. Silva, "Provenance in Comparative Analysis: A Study in Cosmology", Comp. Sci. Eng. **10**, 30 (2008)
10. J. Ahrens, B. Geveci, C. Law, "ParaView: An End User Tool for Large Data Visualization", Visualization Handbook, Academic Press (2005)
11. J. Ahrens, K. Brislawn, K. Martin, B. Geveci, C. Law, M. Papka, "Large Scale Data Visualization Using Parallel Data Streaming", IEEE Comp. Graphics and Applications (2001)

Synergistic Activities

1. Founder and design lead of ParaView, a widely used, open-source visualization tool designed to handle extremely large datasets. Developed parallel VTK, a parallel visualization software infrastructure. ParaView is built upon parallel VTK. From 1997-present.
2. General Chair for IEEE Visualization 2017, SciVis Papers co-chair for IEEE Visualization 2015 and 2016, SciVis Executive Committee Member starting 2016.
3. From 2015 to the present James led the Data Analytics and Visualization Area as a control account manager for the Exascale Computing Project. From 2011 to 2015 he co-led the Data Visualization and Analytics Exascale Planning effort. These activities included helping to define an exascale strategy

and writing a national white paper that set the scope, schedule and budget of an exascale visualization and data analysis activity.
4. Led DOE's Big Data and Privacy briefing for John Podesta's White House 90-day study on this topic, March 2014.
5. Briefed the JASON Advisory Group for a report to examine the technical challenges and economic impacts associated with an Exascale Initiative. I spoke on the relationship between Data Intensive Computing and Exascale Computing, 2011.
6. Co-led the creation of a white paper for the White House's Office of Science and Technology Policy on "Data Intensive Science in the Department of Energy". J. Ahrens, B. Hendrickson, G. Long, S. Miller, R. Ross, D. Williams, Available as Technical Report LA-UR-10-07088, 2010.
7. Each summer, led and organized a Los Alamos data science summer school with professors from many universities visiting and presenting to laboratory community about data science topics. Helped mentor and guide students from these institutions to develop state of the art data science solutions to laboratory problems of interest, 2013-present.
8. Provided computer science community leadership through positions as: DOE's SciDAC-3 SDAV Institute Executive and Visualization Lead, Founding Executive Committee Member of the IEEE Large Data Analysis and Visualization Symposium from 2011-present, Computer Science Liaison to Utah University as part of ASC Alliance Program in 2014, Program Chair for the Eurographics Symposium on Parallel Graphics and Visualization in 2012, Guest Editor, IEEE Computer Graphics and Applications in 2010.

Collaborators and Co-Editors

Collaborators: S. Ahern (CEI Inc), E. Anderson (Utah), R. Armstrong (UC Davis), D. Bader (ORNL), W. Bethel (LBNL), R. Bujack (LANL), E. Brugger (LLNL), H. Childs (UO), J. Freire (NYU), B. Geveci (Kitware), P. Grosset (LANL), S. Habib (ANL), C. Hansen (Utah), B. Hamann (UC Davis), K. Heitmann (ANL), C. Hsu (ORNL), S. Jourdain (Kitware), M. Larsen (LLNL), L. Lins (Utah), Z. Lukic (U. Illinois), K. Martin (Kitware), K.L. Ma (UC Davis). P. McCormick (Los Alamos), K. Myers (LANL), P. O'Leary (Kitware), B. O'Shea (Michigan State), V. Pascucci (Utah), J. Patchett (LANL), M. Petersen (LANL), P. Ricker (NCSA), G. Roth (NVidia), O. Ruebel (LBNL), E. Santos (Utah), C. Sewell (LANL), C. Silva (NYU), V. Springel (Max-Planck Institute), J. Stadel (U. Zurich), H. Trac (Princeton), C. Ware (UNH), G. Weber (LBNL), J. Wendelberger (Los Alamos), M. White (UC Berkeley), D. Williams (LLNL), S. Williams (UC Davis), J. Woodring (LANL), Co-editors: Kurt Debattista (Warwick, UK)

Graduate Advisors

Linda Shapiro, Steven Tanimoto - University of Washington

Advisees

Boonthanome Nouanesengsy (Ohio State), Carson Brownlee (Utah), Sohail Shafii (UC Davis), Sean Williams (UC Davis), Christopher Mitchell (Central Florida), Jonathan Woodring (Ohio State), Emanuele Santos (Utah), Brett Wilson (UC Davis), RunZhen Huang (UC Davis), Kim Edlund (UNM), Garrett Aldrich (UC Davis), Divya Banesh (UC Davis), Jesus Pulido (UC Davis), Max Zeyen (U. Kaiserslautern)

Prasanna Balaprakash

Computer Scientist
Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439-4844
Phone: 630-252-1109
Fax: 630-252-5986
pbalapra@anl.gov
http://www.mcs.anl.gov/~pbalapra

Fellow
Computation Institute,
University of Chicago
NAISE, Northwestern University

Education and Training

- Université Libre de Bruxelles, Belgium, Computer Science, Ph.D., 2009
- Université Libre de Bruxelles, Belgium, Computer Science, Master of Advanced Studies., 2005
- Otto-von-Guericke-Universität Magdeburg, Germany, Computer Science, M.S., 2004
- Periyar University, India, Computer Science, Bachelor of Engineering, 2002

Research and Professional Experience

- 2017 – Present: *Computer Scientist*, MCS Division/LCF, Argonne National Laboratory
- 2017 – Present: *Fellow*, NAISE, Northwestern University
- 2014 – Present: *Computation Institute*, University of Chicago
- 2014–2017: *Assistant Computer Scientist*, MCS Division/LCF, Argonne National Laboratory
- 2010–2013: *Postdoctoral Appointee*, MCS Division, Argonne National Laboratory
- 2009–2010: *Chief Technology Officer*, Mentis SA, Brussels, Belgium.

Honors and Awards

- Best Paper Award, 21st High Performance Computing Symposia (HPC), The Society for Modeling & Simulation International, San Diego, April 2013
- F.N.R.S. Charge de Recherches Postdoc Fellowship, Le Fonds de la Recherche Scientifique, Belgium, October 2010 to September 2013 (Declined)
- F.N.R.S. Aspirant Graduate Fellowship, Le Fonds de la Recherche Scientifique, Belgium, October 2006 to September 2008
- Marie Curie Fellowship, European Commission, November 2004 to September 2006
- International Student Scholarship, Otto-von-Guericke-Universität Magdeburg, Germany, April 2003 to March 2004

Selected Publications Relevant to this Proposal

1. P. Balaprakash, A. Tiwari, S.M. Wild, L. Carrington, and P.D. Hovland. AutoMOMML: Automatic multiple objectives modeling with machine learning. In Intl. Supercomputing Conference (ISC), 2016.
2. O. Subasi, S. Di, L. Bautista-Gomez, P. Balaprakash, O. Unsal, J. Labarta, A. Cristal, and F. Cappello. Spatial support vector regression to detect silent errors in the exascale era. In 16^{th} IEEE/ACM Intl. Symp. on Cluster, Cloud and Grid Computing (CGGRID), 2016.
3. M. Berry, T.E. Potok, P. Balaprakash, H. Hoffmann, R. Vatsavai, and Prabhat. Machine learning and understanding for intelligent extreme scale scientific computing and discovery. DOE ASCR Workshop Report, 2015.
4. F. Isaila, P. Balaprakash, S.M. Wild, D. Kimpe, R. Latham, R. Ross, and P.D. Hovland. Collective I/O tuning using analytical and machine learning models. In 2015 IEEE Intl. Confön Cluster Computing (CLUSTER). IEEE, 2015.

5. A. Mametjanov, P. Balaprakash, C. Choudary, P. D. Hovland, S. M. Wild, and G. Sabin. Autotuning FPGA design parameters for performance and power. In 2015 IEEE 23rd Annual Intl. Symp. on Field-Programmable Custom Computing Machines (FCCM), 2015.
6. T. Nelson, A. Rivera, P. Balaprakash, M. Hall, P. D. Hovland, E. Jessup, and B. Norris. Generating efficient tensor contractions for GPUs. In 2015 44th Intl. Conference on Parallel Processing (ICPP), pp. 969-978. IEEE, 2015.
7. P. Balaprakash, K. Rupp, A. Mametjanov, R.B. Gramacy, P.D. Hovland, and S.M. Wild. Empirical performance modeling of GPU kernels using active learning. In Proc. of the International Conference on Parallel Computing (ParCo2013), Sep. 2013.
8. P. Balaprakash, A. Tiwari, and S.M. Wild. Multi-objective optimization of HPC kernels for performance, power, and energy. In Proc. of the 4th Intl. Wkshp. on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS13), Nov. 2013.
9. P. Balaprakash, R.B. Gramacy, and S.M. Wild. Active-learning-based surrogate models for empirical performance tuning. In Proc. of IEEE Cluster, September 2013.
10. P. Balaprakash, S.M. Wild, and P. Hovland, Can Search Algorithms Save Large-scale Automatic Performance Tuning? In Proc. of the International Conference on Computational Science, ICCS 2011, Procedia Computer Science, Vol. 4, pp. 2136-2145, 2011.

Synergistic Activities

- *Co-lead* for self-aware runtime and operating systems, and member of workshop organizing committee, DOE ASCR Machine learning workshop, January 2015, Rockville, MD.
- *Lead developer* for **AutoMOMML**: An end-to-end, machine-learning-based framework to build predictive models for objectives such as performance, power, and energy.
- *Contributor* to **irace Package**: An iterated racing procedure to automatically configure optimization algorithms by finding the most appropriate settings given a set of instances of an optimization problem.
- *Member* of computational readiness review committee, 2016 INCITE Program; PC member, first IEEE Workshop on Emerging Parallel and Distributed Runtime Systems and Middleware (IPDRM2016) May 2016, Chicago, IL; PC member, PMBS15, PMBS16, PMBS17.
- *Vice president*, Argonne Postdoctoral Association, January — December 2012.

Collaborators and Co-editors: Y. Alexeev (Argonne), L.A. Bautista Gomez (BSC, Spain), M. Birattari (ULB, Belgium), M.-S. Bouguerra (Yahoo), C. Catlett (Argonne), F. Cappello (Argonne), L. Carrington (San Diego Supercomputer Center), A. Chien (Univ. Chicago and Argonne), C. Choudary (RNet), L. Curfman McInnes (Argonne), M. Dorigo (ULB, Belgium), I. Foster (Argonne), R. Gramacy (Univ. Chicago), A. Guha (Univ. Chicago), R. Gupta (Argonne), M. Hall (Univ. Utah) P. Hovland (Argonne), F. Isaila (Univ. Madrid), R. Jacob (Argonne), A. Kannan (IBM), R. Latham (Argonne), O.A.V. Lilienfeld (VUB, Belgium), S. Leyffer (Argonne), A. Mametjanov (Argonne), K. Maheshwari (Argonne), S. Mickelson (Argonne), A. Moawad (Argonne), J. Moré (Argonne), V. Morozov (Argonne), T. Munson (Argonne), S. Narayanan (Argonne), B. Norris (Univ. Oregon), R. Ross (Argonne), A. Rousseau (Argonne), K. Rupp (Vienna Univ.), G. Sabin (Rnet), J. Sarich (Argonne), T. Stützle (ULB, Belgium), A. Tiwari (San Diego Supercomputer Center), A. Vazquez-Mayagoitia (Argonne), S. Wild (Argonne)

Graduate and Postdoctoral Advisors and Advisees: Marco Dorigo (ULB, Belgium), Stefan M. Wild (Argonne), Sandeep Madireddy (Argonne), Yi Ming Yu (New Jersey Institute of Technology), Amal Fethi (ENS Paris-Saclay), Prateek Agarwal (IIT, Chicago)

David E. Bernholdt — Oak Ridge National Laboratory

Research Interests

David Bernholdt is a Distinguished R&D Staff Member at Oak Ridge National Laboratory. He is Group Leader for the Computer Science Research Group in the Computer Science and Mathematics Division; his group is also part of the National Center for Computational Sciences, where he leads Programming Environment and Tools for the Oak Ridge Leadership Computing Facility (OLCF). His research interests are in programming environments for high-performance scientific computing, broadly interpreted. This includes programming languages and programming models, resilience, frameworks, and the "engineering of scientific software". Although his research is now in computer science, David's formal training and early research was in computational chemistry, and he continues to motivate his work through close interactions with scientific applications in a variety of domains, including fusion, nuclear engineering, climate modeling, and chemistry. David's group spans a broad range, including program translation (i.e. compilers, source code analysis and transformation, and runtime systems), tools (i.e. debuggers, performance, and other tools), system software (i.e. operating systems, scheduling, file systems, and resilience), and work on computer science issues arising in scientific applications.

Education and Training

Undergraduate	University of Illinois	Chemistry	BS	1986
Graduate	University of Florida	Chemistry	PhD	1993
Postdoctoral	Pacific Northwest Nat'l Lab	Computational Chemistry	1993–1995	

Research and Professional Experience

2012–present	Distinguished Research Staff Member and Group Leader, Computer Science and Mathematics Division and National Center for Computational Sciences, Oak Ridge National Laboratory
2005–2011	Senior Research Staff Member, Computer Science and Mathematics Division, Oak Ridge National Laboratory
2000–2004	Research Staff Member, Computer Science and Mathematics Division, Oak Ridge National Laboratory
1995–2000	Sr. Research Scientist and Alex G. Nason Fellow, Northeast Parallel Architectures Center, Syracuse University.

Synergistic Activities

- ASCR lead for three current (SciDAC-3) and three proposed (SciDAC-4) application projects
- Lead for Programming Env. and Tools for the Oak Ridge Leadership Computing Facility (OLCF) and for the OLCF-4 project (a.k.a. CORAL or Summit)
- Lead for Outreach and ORNL PI for Interoperable Design of Extreme-Scale Application Software (IDEAS) project and IDEAS-ECP within the Exascale Computing Project
- Lead PI or ORNL PI for two Exascale Computing Project Software Technologies projects

Recent and Related Publications

[1] D. E. Bernholdt, W. R. Elwasif, C. Kartsaklis, S. Lee, T. Mintz, *Programmer-Guided Reliability for Extreme-Scale Applications* in *Cluster Computing (CLUSTER), 2015 IEEE International Conference on*, IEEE, (1st International Workshop on Fault Tolerant Systems).

[2] B. Kocoloski, J. Lange, H. Abbasi, D. E. Bernholdt, T. R. Jones, J. Dayal, N. Evans, M. Lang, J. Lofstead, K. Pedretti, P. G. Bridges, *System-Level Support for Composition of Applications* in *Proceedings of the 5th International Workshop on Runtime and Operating Systems for Supercomputers (ROSS 2015)*. http://dx.doi.org/10.1145/2768405.2768412.

[3] R. Brightwell, R. Oldfield, A. B. Maccabe, D. E. Bernholdt, E. Brewer, P. Bridges, P. Dinda, J. Dongarra, C. Iancu, M. Lang, J. Lange, D. Lowenthal, F. Mueller, K. Schwan, T. Sterling, P. Teller, *Hobbes: Composition and Virtualization as the Foundations of an Extreme-scale OS/R* in *Proceedings of the 3rd International Workshop on Runtime and Operating Systems for Supercomputers*, ACM, New York, NY, USA, of *ROSS '13*, pp. 2:1–2:8. http://doi.acm.org/10.1145/2481425.2481427.

[4] W. R. Elwasif, D. E. Bernholdt, S. S. Foley, A. G. Shet, R. Bramley, *Multi-Level Concurrency in a Framework for Integrated Loosely Coupled Plasma Simulations* in *Computer Systems and Applications (AICCSA), 2011. 9th IEEE/ACS International Conference on*. http://dx.doi.org/10.1109/AICCSA.2011.6126601.

[5] A. G. Shet, W. Elwasif, S. S. Foley, B. H. Park, D. E. Bernholdt, R. Bramley. Strategies for Fault Tolerance in Multicomponent Applications. *Procedia Computer Science* **2011**, *4*, 2287–2296, Proceedings of the International Conference on Computational Science, ICCS 2011.

[6] S. S. Foley, W. Elwasif, D. E. Bernholdt, A. G. Shet, R. Bramley, *Many-Task Applications in the Integrated Plasma Simulator* in *Many-Task Computing on Grids and Supercomputers (MTAGS), 2010 IEEE Workshop on*. http://dx.doi.org/10.1109/MTAGS.2010.5699425.

[7] W. Elwasif, D. E. Bernholdt, A. G. Shet, S. S. Foley, R. Bramley, D. B. Batchelor, L. A. Berry, *The Design and Implementation of the SWIM Integrated Plasma Simulator* in *Parallel, Distributed and Network-Based Processing (PDP), 2010 18th Euromicro International Conference on*, pp. 419–427. http://doi.ieeecomputersociety.org/10.1109/PDP.2010.63.

[8] R. F. Barrett, S. R. Alam, V. de Almeida, D. E. Bernholdt, W. R. Elwasif, J. A. Kuehn, S. W. Poole, A. G. Shet, *Exploring HPCS Languages in Scientific Computing* in *SciDAC 2008, 14-17 July 2008, Washington, USA*, Institute of Physics, Vol. 125 of *Journal of Physics: Conference Series*, p. 012034, invited paper. http://www.iop.org/EJ/article/1742-6596/125/1/012033/jpconf8_125_012034.pdf.

[9] A. G. Shet, W. R. Elwasif, R. J. Harrison, D. E. Bernholdt, *Programmability of the HPCS Languages: A Case Study with a Quantum Chemistry Kernel* in *2008 IEEE International Parallel and Distributed Processing Symposium*.

[10] B. A. Allan, R. Armstrong, D. E. Bernholdt, F. Bertrand, K. Chiu, T. L. Dahlgren, K. Damevski, W. R. Elwasif, T. G. W. Epperly, M. Govindaraju, D. S. Katz, J. A. Kohl, M. Krishnan, G. Kumfert, J. W. Larson, S. Lefantzi, M. J. Lewis, A. D. Malony, L. C. McInnes, J. Nieplocha, B. Norris, S. G. Parker, J. Ray, S. Shende, T. L. Windus, S. Zhou. A Component Architecture for High-Performance Scientific Computing. *Intl. J. High-Perf. Computing Appl.* **2006**, *20*, 163–202.

Collaborators, Advisors and Advisees

Excluding collaborators at institutions participating in this proposal.

Gabrielle Allen[14], Randall Allen[20], Dorian Arnold[40], Rodney J. Bartlett[8], Robert Blackmore[11], Paul T. Bonoli[17], Dylan Brennan[30], Eric Brewer[4], Patrick Bridges[40], Ron Brightwell[33], Henri Calandra[37], Jeff Candy[10], Jeffrey Carver[1], Barbara Chapman[35], Venkatesh Chopella[12], Anthony Curtis[34], Jay Dayal[16], Peter Dinda[24], Greg Eisenhauer[9], Donald Estep[6], Noah Evans[23], Kurt Ferreira[33], Samantha S. Foley[38], Ada Gavrilovska[9], Richard Graham[19], Ryan Grant[33], Martin Greenwald[17], Karl D. Hammond[21], Robert J. Harrison[35], Michael Heroux[33], Chris Holland[39], Lin Hu[18], Douglas L. Hudson[31], Maxime Hugues[37], Laxmikant Kale[14], Chuck Kessel[27], Omar Knio[7], Brian Kocoloski[28], Rick Kurtz[26], Jack Lange[28], James Laros[33], Gerald Lofstead[33], David Lowenthal[2], Bob Lucas[41], Andrew Lumsdaine[26], Dimitrios Maroudas[18],

Reed Maxwell[5], Henik Mikre[3], Frank Mueller[22], Habib Najm[33], Boyana Norris[25], Kathryn O'Brien[11], Ron Oldfield[33], Zachary W. Parchman[36], Giovanni Pastore[13], Kevin Pedretti[33], Khachik Sargsyan[33], Vivek Sarkar[32], Tim Scheibe[26], Stephen L. Scott[36], Aniruddha G. Shet[16], Brian E. Smith[11], Phil Snyder[10], Carl Sovinec[43], Thomas Sterling[15], Patricia Teller[42], Michael Wolfe[29]. **Institutional Affiliations:** [1]Alabama, [2]Arizona, [3]BLEN Corp, [4]Berkeley, [5]Colo. School Mines, [6]Colorado St, [7]Duke, [8]Florida, [9]GaTech, [10]Gen. Atomics, [11]IBM, [12]IIIT-Hyderabad, [13]INL, [14]Illinois, [15]Indiana, [16]Intel, [17]MIT, [18]Massachusetts, [19]Mellanox, [20]Mentor Graphics, [21]Missouri, [22]NCSU, [23]New Mexico Consortium, [24]Northwestern, [25]Oregon, [26]PNNL, [27]PPPL, [28]Pitt, [29]Portland Group, [30]Princeton, [31]Retired, [32]Rice, [33]SNL, [34]Stonybrook, [35]Stonybrook/BNL, [36]Tennessee Tech., [37]Total E&P, [38]U Wisc.-La Crosse, [39]UCSD, [40]UNM, [41]USC, [42]UTEP, [43]Wisconsin.

Graduate and Post Doctoral Advisors

Graduate: Rodney J. Bartlett, University of Florida
Postgraduate: Robert J. Harrison, Stony Brook U/Brookhaven National Laboratory

Thesis Advisor and Postgraduate-Scholar Sponsor

Venkatesh Chopella, Indian Inst. of Information Technology-Kerala; Aniruddha Shet, Intel; Samantha Foley, U. Wisconsin-La Crosse; Henok Mikre, BLEN Corp.; Jay Billings, ORNL; Tiffany Mintz, ORNL; Sophie Blondel, Tennessee; Ozgur Cekmer, ORNL

Curriculum Vitae

E. Wes Bethel

July, 2017

Senior Computer Scientist, Data Analytics and Visualization Group Leader
Computer and Data Sciences Department
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

http://vis.lbl.gov/~wes
Voice: (510) 486-7353
Fax: (510) 486-5812
E-mail: ewbethel@lbl.gov

BIOGRAPHICAL SKETCH

Bethel, a Senior Scientist at Lawrence Berkeley National Laboratory, and his team conduct basic and applied research and development on techniques and tools for enabling scientific knowledge discovery in some of the world's largest collections of scientific data generated by computational models, simulations, and experiments. Since taking over as Group Leader in 2001, he has grown the program from two persons with an annual budget of about $500K to a vibrant program that has earned an internationally acclaimed reputation for excellence, and that consists of about 12 staff and an annual budget of over $4M. He conceived and was the Coordinating Principal Investigator for the SciDAC Visualization and Analytics Center for Enabling Technology (VACET) which made production-quality, petascale-capable visualization a reality at DOE supercomputing facilities, and has produced software tools in use by a worldwide scientific community in disciplines ranging from high energy physics to climate modeling. His research interests include software architecture, high performance computing, scientific and information visualization, computer graphics, computer vision, image analysis, and machine learning. He is an ACM Distinguished Scientist, and a Senior Fellow at the Berkeley Institute for Data Science (UC Berkeley).

EDUCATION AND TRAINING

University of California, Davis	Computer Science	Ph.D.	2010
University of Tulsa	Computer Science	M.S.	1986
University of Tulsa	Information Systems	B.S.	1983

RESEARCH AND PROFESSIONAL EXPERIENCE

2017–present	Old Dominion University, Norfolk, VA – Adjunct Associate Professor, Computer Science
2011–present	Lawrence Berkeley National Laboratory, Berkeley, CA – Group Leader, Senior Computer Scientist
2001–2011	Lawrence Berkeley National Laboratory, Berkeley, CA – Group Leader, Computer Scientist
1990–2001	Lawrence Berkeley National Laboratory, Berkeley, CA – Computer Scientist
1997–2010	R3vis Corporation, Novato, CA – Founding Technical Director, Chief Technology Officer, Software Architect
2000–2001	University of California, Berkeley, Institute for Transportation Studies – Principal Development Engineer
1988–1989	Bethel Software – Principal
1987–1988	Island Graphics Corporation, San Rafael, CA – Software Engineer
1987	Amoco Research Center, Tulsa, OK – Consultant
1986-1987	Geoscan Inc., Tulsa, OK – Senior Graphics Engineer
1984-1986	University of Tulsa – Graduate Research Assistant

Select Recent Publications

1. Utkarsh Ayachit, Andrew Bauer, Earl P. N. Duque, Greg Eisenhauer, Nicola Ferrier, Junmin Gu, Kenneth Jansen, Burlen Loring, Zarija Lukić, Suresh Menon, Dmitriy Morozov, Patrick O'Leary, Michel Rasquin, Christopher P. Stone, Venkat Vishwanath, Gunther H. Weber, Brad Whitlock, Matthew Wolf, K. John Wu, and E. Wes Bethel. Performance Analysis, Design Considerations, and Applications of Extreme-scale *In Situ* Infrastructures. In *ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC16)*, Salt Lake City, UT, USA, November 2016.

2. E. Wes Bethel, Martin Greenwald, Kerstin Kleese van Dam, Manish Parashar, Stefan M. Wild, and H. Steven Wiley. Management, Analysis, and Visualization of Experimental and Observational Data – The Convergence of Data and Computing. In *Proceedings of the 2016 IEEE 12th International Conference on eScience*, Baltimore, MD, USA, October 2016.
3. Andrew C. Bauer, Hasan Abbasi, James Ahrens, Hank Childs, Berk Geveci, Scott Klasky, Kenneth Moreland, Patrick O'Leary, Venkatram Vishwanath, Brad Whitlock, and E. Wes Bethel. *In Situ* Methods, Infrastructures, and Applications on High Performance Computing Platforms, a State-of-the-art (STAR) Report. *Computer Graphics Forum, Proceedings of Eurovis 2016*, 35(3), June 2016.
4. O. Rübel, B. Loring, J. L. Vay, D. P. Grote, R. Lehe, S. Bulanov, H. Vincenti, and E. W. Bethel. WarpIV: In Situ Visualization and Analysis of Ion Accelerator Simulations. *IEEE Computer Graphics and Applications*, 36(3):22–35, May 2016.
5. E. Wes Bethel, Hank Childs, and Charles Hansen, editors. *High Performance Visualization—Enabling Extreme-Scale Scientific Insight*. Chapman & Hall, CRC Computational Science. CRC Press/Francis–Taylor Group, Boca Raton, FL, USA, November 2012. `http://www.crcpress.com/product/isbn/9781439875728`.

Select Recent Grants

1. U. S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. "Towards Exascale: High Performance Visualization and Analytics." FY2015–2017. E. Wes Bethel (PI).
2. U. S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. "SciDAC Scalable Data Analysis and Visualization Institute." FY2012–2017. A. Shoshani (PI), E. Wes Bethel (co-PI), along with co-PIs from 12 other institutions.
3. U. S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. "Scalable Analysis Methods and *In Situ* Infrastructure for Extreme Scale Knowledge Discovery." FY2015–FY2017. E. Wes Bethel (PI).

Graduate Advisors/Committee

- PhD Dissertation Committee: Dr. Kenneth I. Joy (chair), Dr. John D. Owens, Dr. Horst D. Simon.
- MS Thesis Committee: Dr. Samuel P. Uselton (chair), Dr. Dale A. Shoenefeld, Dr. Thomas W. Cairns.
- Students supervised/committee member: Kenes Bekatayev (PhD, UC Davis); Christine de Bianchi (Emory); Dean Brederson (Utah); Ian Bowman (MS, UCD); David Camp (PhD, UC Davis); Jerry Chi-Li Chen (MS, SFSU); Luke Gosink (PhD, UC Davis); Chengcheng Hu (PhD, UC Davis); T. J. Jankun-Kelly (PhD, UC Davis); Hoa Nguyen (PhD, Utah); Nameeta Shah (PhD, U.C. Davis);

Synergistic Activities

- As a Senior Fellow in the Berkeley Institute for Data Science, he serves as a bridge between the academic data science research community and DOE's large-scale HPC-focused R&D efforts in the data space.
- As organizer and general chair for the SC workshop *In Situ* Infrastructures for Enabling Extreme-scale Analysis and Visualization, he fostered growth of a broad *in situ* community in the HPC space.
- As Data Analytics and Visualization sub-plexus co-lead for the DOE Office of Advanced Scientific Computing Research, he was instrumental in shaping the R&D agenda for visualization and analytics for the entire DOE-funded analytics and visualization community that will lead to algorithms and software infrastructure usable by the scientific community on future exascale-class platforms.
- As Coordinating Principal Investigator on DOE's largest-ever open-science visualization program, he was the visionary and leader for a five-institution program that made production-quality, petascale-capable visual data exploration and analysis a reality, and showed positive impact on computational science research projects.
- *De facto* chair of the DOE Computer Graphics Forum (DOECGF), an *ad hoc* organization of computer graphics and visualization professionals who receive DOE funding dating back to the 1970s. See `www.doecgf.org`.

Other Certifications: Instrument-rated private pilot with complex and high performance aircraft endorsements, FAA certificate No. 2791855; Taekwondo 4th Dan, Kukkiwon certificate No. 5905313.

Eric Brugger

Education and Training

University of California, Berkeley Computer Science B.A. 1985

Research and Professional Experience

Project leader and software developer for VisIt	Lawrence Livermore National Laboratory	2000 -
Project leader and software developer for MeshTV	Lawrence Livermore National Laboratory	1993 - 2000
Software developer for MeshTV	Lawrence Livermore National Laboratory	1989 - 1993
Simulation code developer	Lawrence Livermore National Laboratory	1986 - 1990

Selected Publications

1. M. Larsen, E. Brugger, H. Childs, J. Eliot, K. Griffin, and C. Harrison. Strawman: A batch in situ visualization and analysis infrastructure for multi-physics simulation codes. In *Proceedings of the First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (ISAV), held in conjunction with SC15*, Austin, TX, November 2015, pages 30-35.
2. H. Childs, et al. VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data. In *Proceedings of SciDAC 2011*, Denver, CO, July 2011.
3. E. W. Bethel, J. Rosendale, D. Southard, K. Gaither, H. Childs, E. Brugger, S. Ahern. Visualization Viewpoints Visualization at Supercomputing Centers: The Tale of Little Big Iron and the Three Skinny Guys. In *Proceeding of the IEEE Computer Graphics and Applications*, pp. 90–95, January/February, 2011
4. B. Paul, S. Ahern, E. W. Bethel, E. Brugger, R. Cook, J. Daniel, K. Lewis, J. Owen, D. Southard. Chromium Renderserver: Scalable and Open Remote Rendering Infrastructure. In *Proceeding of the IEEE Transactions on Visualization and Computer Graphics"*, vol. 14, no. 3, pp. 627–639, May/June, 2008
5. H. Childs, E. S. Brugger, K. S. Bonnell, J. S. Meredith, M. Miller, B. J. Whitlock, and N. Max. A Contract-Based System for Large Data Visualization. In *Proceedings of IEEE Visualization (Vis05)*, pages 190–198, Minneapolis, MN, Oct. 2005.

Software Systems Developed

- VisIt software system: Co-designer of VisIt. Made the key architectural decisions to utilize a distributed, component-based architecture, and basing the visualization and analysis capabilities on VTK. Implemented the initial viewer component. Put in place the management, development and SQA processes to enable VisIt to support a large, distributed user base. Received an R&D 100 award in 2005 for VisIt. Fellow award recipients were Sean Ahern, Kathleen Bonnell, Hank Childs, Linnea Cook, Jeremy Meredith, Mark Miller, and Brad Whitlock.
- MeshTV software system: Implemented the majority of 3D visualization capabilities, implemented the initial graphical user interface, designed the parallel version and contributed to its development. This was the first truly scalable, parallel, visualization tool. Enabled MeshTV to visualize data directly from a remotely running simulation using a socket based library, which was demonstrated at Supercomputing 94.

Synergistic Activities

- Founding project leader and developer for VisIt (2000-), a widely-used large-scale visualization and analysis software package.

Recent Collaborators

E. Wes Bethel, Lawrence Berkeley Lab
Patricia Crossno, Sandia Lab
Berk Geveci, Kitware
Jeremy Meredith, Oak Ridge Lab
Ken Moreland, Sandia Lab
John Patchett, Los Alamos Lab
Dave Pugmire, Oak Ridge Lab
Allen Sanderson, University of Utah
Gunther Weber, Lawrence Berkeley Lab
Brad Whitlock, Intelligent Light

Aydın Buluç
Computational Research Division,
Lawrence Berkeley National Laboratory,
Berkeley, CA 94720
Phone: 510-486-5197;
Email: abuluc@lbl.gov

Education and Training

- UC Center for Executive Education: Emerging Leaders Development Program, 2015–2016
- Lawrence Berkeley National Laboratory: Luis W. Alvarez Postdoctoral Fellow, 2010–2012
- University of California-Santa Barbara, Computer Science: Ph.D., 2010; M.S., 2009
- Sabancı University, Computer Science and Engineering (minor: Mathematics): B.S., 2005

Research and Professional Experience

- *Staff Scientist*, Lawrence Berkeley National Laboratory, Computational Research Division, 2016 – present
- *Adjunct Assistant Professor*, University of California-Berkeley, 2017 – present
- *Research Scientist (Career)*, Lawrence Berkeley National Laboratory, Computational Research Division, 2014 – 2016
- *Visiting Scientist*, Simons Institute for the Theory of Computing, Fall 2013.
- *Research Scientist (Career-Track)*, Lawrence Berkeley National Laboratory, Computational Research Division, 2012 – 2014

Selected Honors and Awards

- *Excellence in Early Career Research* by IEEE Tech. Committee on Scalable Computing, 2015
- *DOE Early Career Award* by Office of Science, Department of Energy, 2013

Selected Publications Relevant to this Proposal

1. Y. You, A. Buluç, and J. Demmel. Scaling deep learning on GPU and Knights Landing clusters. In *Intl. Conference for High Performance Computing, Networking, Storage and Analysis (SC'17)*, ACM, 2017. To appear.
2. A. Azad, A. Buluç. A work-efficient parallel sparse matrix-sparse vector multiplication algorithm. *IEEE Intl. Parallel & Distributed Processing Symposium (IPDPS)* 2017.
3. A. Buluç, T. Mattson, S. McMillan, J. Moreira, and C. Yang. Design of the GraphBLAS API for C. *IEEE Workshop on Graph Algorithm Building Blocks, IPDPSW*, 2017.
4. A. Azad, A. Buluç, and A. Pothen. Computing maximum cardinality matchings in parallel on bipartite graphs via tree-grafting. *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 28(1):44–59, 2017.
5. A. Azad, G. Ballard, A. Buluç, J. Demmel, L. Grigori, O. Schwartz, S. Toledo, and S. Williams. Exploiting multiple levels of parallelism in sparse matrix-matrix multiplication. *SIAM Journal on Scientific Computing (SISC)*, 38(6):C624–C651, 2016.
6. A. Azad and A. Buluç. Distributed-memory algorithms for maximum cardinality matching in bipartite graphs. *IEEE Intl. Parallel & Distributed Processing Symposium (IPDPS)*, 2016.
7. A. Buluç, H. Meyerhenke, I. Safro, P. Sanders, and C. Schulz Recent advances in graph partitioning *Algorithm Engineering - Selected Results and Surveys*, Vol. 9220. Lecture Notes in Computer Science, 2016
8. E. Georganas, A. Buluç, J. Chapman, L. Oliker, D. Rokhsar, and K. Yelick. Parallel de bruijn graph construction and traversal for de novo genome assembly *Intl. Conference for High Performance Com-*

puting, Networking, Storage and Analysis (SC), 2014.
9. A. Buluç, E. Duriakova, A. Fox, J. Gilbert, S. Kamil, A. Lugowski, L. Oliker, and S. Williams High-productivity and high-performance analysis of filtered semantic graphs *IEEE Intl. Parallel & Distributed Processing Symposium (IPDPS)*, 2013
10. A. Buluç and J. Gilbert The Combinatorial BLAS: Design, implementation, and applications *The Intl. Journal of High Performance Computing Appl. (IJHPCA)*, 25(4):496 – 509, 2011

Selected Synergistic Activities

- **Significant Software**:
  - The Combinatorial BLAS is a parallel library for running large-scale graph algorithms on distributed-memory architectures.
  - HipMer is an extreme-scale de novo genome assembler for large complex genomes that scales over O(10,000) processors.
  - GraphBLAS provides standards for graph algorithmic primitives in the language of linear algebra. Mathematical description and C API Specification is released.
- **Editorship**: Founding associate editor, ACM Transactions on Parallel Computing, 2013-present. Guest Editor, Parallel Computing, issue on 'Graph Analysis for Scientific Discovery' (2015).
- **Program Committee** SC (applications vice chair: 2017, member: 2013-14, 2016), IPDPS (member: 2013-17), GABB (co-chair: 2017, member: 2015-), SIAM CSC (2016), ICS (2015)
- Exascale Requirements Review for Biological and Environmental Research, 2016
- **Steering Committee**, The GraphBLAS Forum, IEEE GABB Workshop

Collaborators within the Past 48 Months

Md. Afibuzzaman (Michigan State), H. Metin Aktulga (Michigan State), Krste Asanovic (UC Berkeley), David Bader (Georgia Tech), Grey Ballard (West Forest), Scott Beamer (LBNL), Jarrod Chapman, James Demmel (UC Berkeley), Erika Duriakova (University College Dublin), Armando Fox (UC Berkeley), Evangelos Georganas (UC Berkeley), John Gilbert (UCSB), Joseph Gonzalez (UC Berkeley), Laura Grigori (INRIA Saclay-Ile de France), Steven Hofmeyr (LBNL), Mathias Jacquelin (LBNL), Stefanie Jegelka (MIT), Shoaib Kamil (Adobe), Jeremy Kepner (MIT-LL), Penporn Koanantakool (UC Berkeley), Nikos Kyrpides (JGI), Andrew Lumsdaine (PNNL), Adam Lugowski, Kamesh Madduri (Penn State), Pieter Maris (Iowa State), Tim Mattson (Intel), Scott McMillan (CMU SEI), Henning Meyerhenke (KIT), Jose Moreira (IBM), Esmond G. Ng (LBNL), Sang-Yun Oh (UCSB), Leonid Oliker (LBNL), John Owens (UC Davis), Alex Pothen (Purdue), Daniel Rokhsar (JGI/UC Berkeley), Ilya Safro (Clemson University), Peter Sanders (KIT), Christian Schulz (KIT), Oded Schwartz (Hebrew University of Jerusalem), Meiyue Shao (LBNL), Veronika Strnadova-Neeley (UCSB), Sivan Toledo (Tel-Aviv University), James P. Vary (Iowa State), Samuel Williams (LBNL), Carl Yang (UC Davis), Chao Yang (LBNL), Yang You (UC Berkeley), Katherine Yelick (LBNL/UC Berkeley).

Advisors

- John R. Gilbert (UC Santa Barbara), PhD advisor
- Juan C. Meza (now at UC Merced), Postdoctoral advisor at LBNL

Graduate/Postdoctoral Advisees

- Ariful Azad (LBNL), Postdoctoral advisee at LBNL
- Sang-Yun Oh (UC Santa Barbara), Postdoctoral advisee at LBNL
- Harsha Simhadri (Microsoft Research India), Postdoctoral advisee at LBNL
- Adam Lugowski (Turn.com), Summer intern advisee at LBNL
- Veronika Strnadova (UC Santa Barbara), Graduate Research Assistant at LBNL
- Carl Yang (UC Davis), Graduate Research Assistant at LBNL

Alok Choudhary

Professor, Electrical Engineering and Computer Science
Director, Center for Ultra-Scale Computing and Information Security
Northwestern University
2145 Sheridan Rd, Evanston
IL-60208-3118
TEL: (847) 467-4129, FAX: (847) 467-4144
EMAIL: choudhar@eecs.northwestern.edu

EDUCATION AND TRAINING:

Birla Institute of Technology and Science	Electrical Engineering	B.E.(hon),1982
University of Massachusetts	Electrical and Computer Engineering	M.S., 1986
University of Illinois	Electrical and Computer Engineering	Ph.D., 1989

RESEARCH AND PROFESSIONAL EXPERIENCE:

Northwestern University, Evanston, Illinois

- Professor, Electrical Engineering and Computer Science Department, Sep. 2000 - present.
- Founder, Chairman, Chief Scientist, 4Cinsights (a startup). 2011-present (CEO 2011 - 2013).
- Director of Center for Ultra-Scale Computing and Information Security, 2000 - present.
- Professor, Technology Industry Management and Marketing, Kellogg School of Management, 2000 - present.
- Chair, Electrical Engineering and Computer Science Department, 2007 - 2011.
- Associate Professor, Electrical and Computer Engineering, Sept 1996 - Aug. 2000.

Syracuse University, Syracuse, NY

- Associate Professor, Electrical and Computer Engineering Department, Aug. 1993 - Aug. 1996.
- Assistant Professor, Electrical and Computer Engineering Department, Aug. 1989 - Aug. 1993.

Intel Corporation – Summer 1995, 1996.
IBM Research Center – Summer 1987, 1988, 1991.
Univ. of Illinois, Urbana-Champaign, Research Assistant, 1986 - 1989.
University of Massachusetts, Amherst, MA, Graduate Assistant, 1984 - 1986.
Tata Consultancy Services, N. Delhi, India, 1982 - 1984.

PUBLICATIONS

1. Z. Yuan, W. Hendrix, S. Son, C. Federrath, A. Agrawal, W. Liao, and A. Choudhary. Parallel Implementation of Lossy Data Compression for Temporal Data Sets. In the 23rd International Conference on High Performance Computing, December 2016.
2. Q. Kang, W. Liao, A. Agrawal, and A. Choudhary. A Filtering-based Clustering Algorithm for Improving Spatio-temporal Kriging Interpolation Accuracy. In 25th ACM International Conference on Information and Knowledge Management, October 2016.
3. E. Rangel, N. Li, S. Habib, T. Peterka, A. Agrawal, W. Liao, and A. Choudhary. Parallel DTFE Surface Density Field Reconstruction (best paper award). In the IEEE International Conference on Cluster Computing, September 2016.
4. D. Palsetia, W. Hendrix, S. Lee, A. Agrawal, W. Liao, and A. Choudhary. Parallel Community Detection Algorithm Using a Data Partitioning Strategy with Pairwise Subdomain Duplication. In the 31st International Supercomputing Conference, June 2016.
5. R. Liu, A. Kumar, Z. Chen, A. Agrawal, V. Sundararaghavan, and A. Choudhary. A Predictive Machine Learning Approach for Microstructure Optimization and Materials Design. Nature Scientific Reports, 5:11551, Macmillan Publishers Limited SN, June 2015.
6. Z. Chen, S. W. Son, W. Hendrix, A. Agrawal, W. Liao, and A. Choudhary, “NUMARCK: Machine

Learning Algorithm for Resiliency and Checkpointing," in Proceedings of 26th International Conference on High Performance Computing, Networking, Storage and Analysis, November 2014, pp. 733–744.
7. H. Xu, R. Liu, A. Choudhary, and W. Chen. A Machine Learning-Based Design Representation Method for Designing Heterogeneous Microstructures, (Design Automation (DAC) Best Paper). In the ASME International Design Engineering Technical Conferences, August 2014.
8. M. Patwary, D. Palsetia, A. Agrawal, W. Liao, F. Manne, and A. Choudhary. Scalable Parallel OPTICS Data Clustering Using Graph Algorithmic Techniques. In the International Conference for High Performance Computing, Networking, Storage and Analysis, November 2013.
9. A Choudhary, William Hendrix, Kathy Lee, Diana Palsetia, and Wei-Keng Liao. "Social media evolution of the Egyptian revolution." Communications of the ACM 55, no. 5: 74-80, 2012.
10. J. Li, W. Liao, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, and M. Zingale. Parallel netCDF: A Scientific High-Performance I/O Interface. In the Proceedings of the Supercomputing Conference, November 2003.

SYNERGISTIC ACTIVITIES:

- At the invitation of the White House attended the Big Data Event in Washington DC (03/2012).
- Member, National Academy of Science Committee on the Potential Impact of High-End Computing on Selected Areas of Sciences. Produced a report in High-End Computing Priorities for the US Govt.
- Served on the program committee of more than 55 conferences including: Architectures and Systems for Big Data Workshop, Supercomputing, International Conference on High Performance Computing & Communications, and International Conference on Data Mining.
- Award recipient: Excellence in Research, Teaching and Service, McCormick School of Engineering, 2006, IBM Faculty Development Award, 1994, 2005, Intel Research Council Award, 1993-1996, 2003-2005, IEEE Engineering Foundation award, National Science Foundation Young Investigator Award, 1993
- Fellow: IEEE, ACM, AAAS

E. IDENTIFICATION OF POTENTIAL CONFLICTS OF INTESEST OR BIAS IN SELECTION OF REVIEWERS

- **Collaborators and Co-editors**
  Peter Beckman (ANL), Ruben Buaba (NC A&T State U.), Auroop Ganguly (Northeastern U.), Bill Gropp (UIUC), Salman Habib (ANL), Abdollah Homaifar (NC A&T State U.), Vassiliki Kalogera (Northwestern U.), Seong Jo Kim (Penn State U.), Scott Klasky (ORNL), Quincey Koziol (HDF Group), Rakesh Krishnaiyer (Intel), Vipin Kumar (UMN), Rob Latham (ANL), Bertram Ludascher (UC Davis), Frederik Manne (U. of Bergen), Gokhan Memik (Northwestern U.), Bhagirath Narahari (GWU), Ron Oldfield (SNL), Gregory Olson (Northwestern U.), Frederic Rasio (Northwestern U.), Robert Ross (ANL), Nagiza Samatova (NCSU), Frederick Semazzi (NCSU), Arie Shoshani (LBNL), Tim Tautges (ANL), Goce Trajcevski (Northwestern U.), Eric Van Wyk (UMN), Peter Voorhees (Northwestern U.), John Wu (LBNL)
- **Graduate and Postdoctoral Advisees**
  Ankit Agrawal (Northwestern U.), Rajesh Bordawekar (IBM), Zhengzhang Chen (NEC), Avery Ching (Facebook), Steve Chiu (Idaho State U.), Kenin Coloma (Ask.com), Abhishek Das (Intel), Al'ona Furmanchuk (Northwestern U.), Kui Gao (Dassault Systemes Simulia Corp), Sanjay Goil (Intel), Dianwei Han (Mount Sinai Medicine School), William Hendrix (U. South Florida), Mahmut Kandemir (Penn State U.), Prabhat Kumar (4C Insights), Jianwei Li (Bloomberg), Wei-keng Liao (Northwestern U.), Ruoqian Liu (Uber), Ying Liu (Chinese Academy of Sciences), Sanchit Misra (Intel India), Sachin More (EMC), Nithin Nakka (UIUC), Ramanathan Narayanan (Goldman Sachs), Arifa Nisar (Salesforce), Jaechun No (Sejong U.), Berkin Ozisikyilmaz (4C Insights), Diana Palsetia (4C Insights), Bharath Pattabiraman (Google), Mostofa Patwary (Intel), Jayaprakash Pisharath (Intel), Ioan Raicu (IIT), Saba Sehrish (FNAL), Seung Woo Son (UMass Lowell), Rajeev Thakur (ANL),

Yves Xie (Baidu), Joe Zambreno (Iowa State U.), Kunpeng Zhang (UIC)

- **Graduate Advisors**
  Janak Patel (UIUC)

Bronis R. de Supinski

Center for Applied Scientific Computing, L-557
Lawrence Livermore National Laboratory
Livermore, CA 94551

Phone: (925) 422-1062
Fax: (925) 423-6961
E-mail: bronis@llnl.gov

Research Interests

- Innovative massively parallel systems
- OpenMP
- Parallel performance evaluation
- Cache coherence and consistency semantics
- Parallel performance and correctness tools
- MPI
- Parallel programming
- Computer architecture and memory systems
- Parallel and distributed simulation
- Compilers

Education

Ph.D. Computer Science, University of Virginia, May 1998
M.S. Computer Science, University of Virginia, January 1995
B.A. Mathematics, University of Chicago, December 1987

Professional Experience

5/15-present	Distinguished Member of Technical Staff, LLNL, Livermore CA
9/12-present	Chief Technology Officer, Livermore Computing (LC), LLNL, Livermore CA
2/12-present	Professor of Exascale Computing, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom
9/10-present	Adjunct Associate Professor, Texas A&M University, College Station, TX
9/05-8/12	ASC Application Development Environment Performance Team Leader, LLNL, Livermore CA
4/08-1/09	Computer Science Group Leader, CASC, LLNL, Livermore CA
4/05-3/08	Data Analysis Group Leader, CASC, LLNL, Livermore CA
5/03-3/05	Advanced Software Technologies Group Leader, CASC, LLNL, Livermore CA
7/98–4/03	Computer Scientist, Center for Applied Scientific Computing (CASC), Lawrence Livermore National Laboratory (LLNL), Livermore CA
2/91–8/92	Computer Coordinator, Dept. of Cell Biology and Anatomy, Mt. Sinai Medical Center, New York, NY
3/90–11/90	Systems Coordinator, Medical Center Development, University of Chicago, Chicago, IL
11/88–3/90	Assistant Director for Systems, Medical Alumni Association, University of Chicago, Chicago, IL

Honors and Organizations

- Chair, OpenMP Language Committee, 2009-present
- Co-Editor in Chief, Editorial Board, International Journal of High Performance Computing Applications (IJHPCA)
- Associate Editor, IEEE Micro, 2016-present
- Associate Editor, IEEE Transactions on Parallel and Distributed Systems (TPDS), 2011-2015
- Technical Papers Co-Chair, SC14
- Program Co-Chair, 25th International Conference on Supercomputing (ICS 2011)
- Program Co-Chair, 18th International Conference on Parallel Architectures and Compilation Techniques (PACT 2009)
- Workshop Co-Chair, 8th and 9th Workshop on High Performance Power-Aware Computing (HPPAC 2012, 2013)
- Chair, International Workshop on OpenMP Steering Committee, 2005-2009 (Member, 2004-present)
- President, IBM Scientific Computing User Group (ScicomP), 2000-2004
- Institute for Electrical and Electronics Engineers (IEEE) and IEEE Computer Society
- Association of Computing Machinery (ACM)
- ACM Distinguished Speaker (2007 – 2014)
- Chair, OpenMP Tools Committee, 2002-2009
- Best Paper Award, ICPP 2014
- Best Paper Award, ICS 2013
- Best Student Paper Award, SC12
- Two Best Paper Awards, EuroMPI 2011
- Member, R&D 100 Prize Winning Team (for *STAT: The Stack Trace Analysis Tool*), 2011
- Best Paper Award, Software Track, IPDPS 2007
- Member, Gordon Bell Prize Winning Team, 2006
- Best Paper Award, PADTAD – IV, 2006
- Best Paper Award, IWOMP 2006
- Member, Gordon Bell Prize Winning Team, 2005

Selected Publications and Presentations

Journal Papers

Bronevetsky, Greg, John Gyllenhaal and Bronis R. de Supinski, "CLOMP: Accurately Characterizing OpenMP Application Overheads," *International Journal of Parallel Programming,* Vol. 37, No. 3, June 2009, pp. 250-265. (LLNL-JRNL-408738*).

Bronevetsky, Greg and Bronis R. de Supinski, "Complete Formal Specification of the OpenMP Memory Model," *International Journal of Parallel Programming,* Vol. 35, No. 4, August 2007, pp. 335-392. (UCRL-JRNL-226950*).

Peer Reviewed Conference Papers

Scogland, Thomas R.W., Barry Rountree, Wu-chun Feng and Bronis R. de Supinski, "CoreTSAR: Adaptive Worksharing for Heterogeneous Systems," *The International Supercomputing Conference (ISC'14)*, Leipzig, Germany, June 22-26, 2014. (LLNL-CONF-662205*)

Olivier, Stephen, Bronis R. de Supinski, Martin Schulz and Jan F. Prins, "Characterizing and Mitigating Work Time Inflation in Task Parallel Programs," *SC2012*, Salt Lake City, Utah, November 11–16, 2012. (Best Student Paper Award) (LLNL-CONF-555492*).

Scogland, Thomas R. W., Barry Rountree, Wu-chun Feng and Bronis R. de Supinski, "Heterogeneous Task Scheduling for Accelerated OpenMP," *Twenty Sixth International Parallel and Distributed Processing Symposium (IPDPS 2012)*, Shanghai, China, May 21–25, 2012. (LLNL-CONF-522311*).

Li, Dong, Bronis R. de Supinski, Martin Schulz, Kirk Cameron and Dimitrios S. Nikolopoulos, "Hybrid MPI/OpenMP Power-Aware Computing," *Twenty Fourth International Parallel and Distributed Processing Symposium (IPDPS 2010)*, Atlanta, Georgia,
April 19–23, 2010. (LLNL-CONF-422990*).

Chapman, Barbara M., Lei Huang, Haoqiang Jin, Gabriele Jost and Bronis R. de Supinski, "Toward Enhancing OpenMP's Work-Sharing Directives," *Euro-Par 2006*, Dresden, Germany, August 29 – September 1, 2006. (UCRL-CONF-221478*).

Peer Reviewed Workshop Papers

Ciesko, Jan, Sergi Mateo, Xavier Teruel, Xavier Martorell, Eduard Ayguade, Jesus Labarta, Alejandro Duran, Bronis R. de Supinski, Stephen Olivier, Kelvin Li and Alexandre E. Eichenberger, " Task-Parallel Reductions in OpenMP," *Eleventh International Workshop on OpenMP (IWOMP 2015)*, Aachen, Germany, September 30 – October 2, 2015. (LLNL-CONF-677684*).

Scogland, Thomas R.W., John Gyllenhaal, Jeff Keasler, Rich Hornung and Bronis R. de Supinski, "Enabling Region Merging Optimizations in OpenMP," *Eleventh International Workshop on OpenMP (IWOMP 2015)*, Aachen, Germany, September 30 – October 2, 2015. (LLNL-CONF-670944*).

Scogland, Thomas R.W., Jeff Keasler, John Gyllenhaal, Rich Hornung, Bronis R. de Supinski and Hal Finkel, "Supporting Indirect Data Mapping in OpenMP," *Eleventh International Workshop on OpenMP (IWOMP 2015)*, Aachen, Germany, September 30 – October 2, 2015. (LLNL-CONF-671602*).

Wong, Michael, Eduard Ayguade, Justin Gottschlich, Victor Luchangco, Bronis R. de Supinski and Barna L. Bihari, "Towards Transactional Memory for OpenMP," *Tenth International Workshop on OpenMP (IWOMP 2014)*, Salvador, Brazil, September 28–30, 2014. (LLNL-CONF-677916*).

Liao, Chunhua, Yonghong Yan, Bronis R. de Supinski, Daniel J. Quinlan and Barbara Chapman, "Early Experiences With The OpenMP Accelerator Model," *Ninth International Workshop on OpenMP (IWOMP 2013)*, Canberra, Australia, September 16–18, 2013. (LLNL-CONF-636479*).

Beyer, James C., Eric J. Stotzer, Alistair Hart and Bronis R. de Supinski, "OpenMP for Accelerators," *Seventh International Workshop on OpenMP (IWOMP 2011)*, Chicago, IL, June 13–15, 2011. (LLNL-CONF-474253*).

Duran, Alejandro, Roger Ferrer, Michael Klemm, Bronis R. de Supinski and Eduard Ayguade, "A Proposal for User-Defined Reductions in OpenMP," *Sixth International Workshop on OpenMP (IWOMP 2010)*, Tsukuba, Japan, June 14–16, 2010. (LLNL-CONF-426239*).

Liao, Chunhua, Daniel J. Quinlan, Thomas Panas and Bronis R. de Supinski, "A ROSE-based OpenMP 3.0 Research Compiler Supporting Multiple Runtime Libraries," *Sixth International Workshop on OpenMP (IWOMP 2010)*, Tsukuba, Japan, June 14–16, 2010. (LLNL-CONF-422873*).

Anshu Dubey

Education and Training

Ph.D., Computer Science, 1993, Old Dominion University, Norfolk, VA

M.S., Electrical Engineering, 1990, Auburn University, Auburn, AL

B.Tech, Electrical Engineering, 1985, Indian Institute of Technology, New Delhi, India

Research and Professional Experience

2015– present	Computer Scientist, MCS Division Argonne National Laboratory, Argonne, IL
2015–present	Senior Fellow, Computation Institute, University of Chicago, Chicago IL
2013– 2015	Computer Systems Engineer, CRD, Lawrence Berkeley National Laboratory, Berkeley CA
2003–2013	Associate Director, ASC/Flash Center, University of Chicago, Chicago IL
1998–2003	Research Scientist ASC/Flash Center, Astronomy & Astrophysics, University of Chicago, Chicago IL
1985-1987	Senior Research Assistant, Comp Science & Engineering, Indian Institute of Technology, New Delhi, India

Related Publications

A. Dubey, H. Fujita, D.T. Graves, A. Chien, and D. Tiwari. Granularity and the cost of error recovery in resilient amr scientific applications. 2016. SC16.

A. Dubey, K. Antypas, E. Coon, and K.M. Riley. Software process for multicomponent multiphysics codes. In J. Carver, N.C. Hong, and G. Thiruvathukal, editors, *Software Engineering for Science*, chapter 1. Taylor and Francis, 2016.

A. Dubey, K. Weide, D. Lee, J. Bachan, C. Daley, S. Olofin, N. Taylor, P.M. Rich, and L.B. Reid. Ongoing verification of a multiphysics community code: FLASH. *Software: Practice and Experience*, 45(2), 2015.

Anshu Dubey, Ann Almgren, John Bell, Martin Berzins, Steve Brandt, Greg Bryan, Phillip Colella, Daniel Graves, Michael Lijewski, Frank Lffler, Brian OShea, Erik Schnetter, Brian Van Straalen, and Klaus Weide. A survey of high level frameworks in block-structured adaptive mesh refinement packages. *Journal of Parallel and Distributed Computing*, 74(12):3217–3227, 2014.

A. Dubey, K. Antypas, A.C. Calder, C. Daley, B. Fryxell, J.B. Gallagher, D.Q. Lamb, D. Lee, K. Olson, L.B. Reid, P. Rich, P.M. Ricker, K.M. Riley, R. Rosner, A. Siegel, N.T. Taylor, F.X. Timmes, N. Vladimirova, K. Weide, and J. ZuHone. Evolution of FLASH, a multiphysics scientific simulation code for high performance computing. *International Journal of High Performance Computing Applications*, 28(2):225–237, 2013.

A. Dubey, A.C. Calder, C. Daley, R.T. Fisher, C. Graziani, G.C. Jordan, D.Q. Lamb, L.B. Reid, D. M. Townsley, and K. Weide. Pragmatic optimizations for better scientific utilization of large supercomputers. *International Journal of High Performance Computing Applications*, 27(3):360–373, 2013.

A. Dubey, C. Daley, J. ZuHone, P. M. Ricker, K. Weide, and C. Graziani. Imposing a Lagrangian particle framework on an Eulerian hydrodynamics infrastructure in FLASH. *ApJ Supplement*, 201:27, aug 2012.

A. Dubey, K. Antypas, and C. Daley. Parallel algorithms for moving Lagrangian data on block structured Eulerian meshes. *Parallel Computing*, 37(2):101 – 113, 2011.

A. Dubey, K. Antypas, M.K. Ganapathy, L.B. Reid, K. Riley, D. Sheeler, A. Siegel, and K. Weide. Extensible component based architecture for FLASH, a massively parallel, multiphysics simulation code. *Parallel Computing*, 35:512–522, 2009.

E.-J. Rijkhorst, T. Plewa, A. Dubey, and G. Mellema. Hybrid characteristics: 3D radiative transfer for parallel adaptive mesh refinement hydrodynamics. *Astronomy and Astrophysics*, 452:907–920, June 2006.

Synergistic Activities

Associate Editor: IEEE TPDS, JPDC, IJHPCA

Applications area chair SC17, and co-chair SC14 technical program committee

Member: SC16 tech papers committee, SC15 posters committee, SC13 tutorials committee

Co-lead: Software Engineering and Community Codes Tracks at Argonne Training Program for Extreme Scale Computing, 2013-2016

Organizer, FLASH Tutorials: 2004 (onsite), 2006(at AAS meeting), 2007(at ICTP, Trieste), 2009 (onsite), 2012 (at RAL, UK)

Berk Geveci

Education and Training

Lehigh University	Mechanical Engineering and Mechanics	Ph.D.	1999
Lehigh University	Mechanical Engineering and Mechanics	M.S.	1996
Bogazici University (Turkey)	Mechanical Engineering	B.S.	1994

Research and Professional Experience

Senior Director of Scientific Computing	Kitware Inc.	2013–
Director of Scientific Computing	Kitware Inc.	2009–2013
Project Lead	Kitware Inc.	2005–2009
R&D Engineer	Kitware Inc.	2000–2005
Post-Doctoral Fellowship	University of Pennsylvania	1999–2000

Selected Publications

1. "VTK-m: Accelerating the Visualization Toolkit for Massively Threaded Architectures," K. Moreland, C. Sewell, W. Usher, L. Lo, J. Meredith, D. Pugmire, J. Kress, H. Schroots, K. Ma, H. Childs, M. Larsen, C. Chen, R. Maynard, and B. Geveci, *IEEE Computer Graphics and Applications,* vol. 36, no. 3, pp. 48-58 2016.
2. "Flying edges: A high-performance scalable isocontouring algorithm," W. Schroeder, R. Maynard, and B. Geveci, 2015, pp. 33-40.
3. "A Classification of Scientific Visualization Algorithms for Massive Threading," Kenneth Moreland, Berk Geveci, Kwan-Liu Ma, and Robert Maynard. In *Proceedings of Ultrascale Visualization Workshop*, November 2013. DOI 10.1145/2535571.2535591.
4. "Research Challenges for Visualization Software," Hank Childs, Berk Geveci, Will Schroeder, Jeremy Meredith, Kenneth Moreland, Christopher Sewell, Torsten Kuhlen, and E. Wes Bethel. *IEEE Computer*, Volume 46, Number 4, May 2013. DOI 10.1109/MC.2013.179.
5. "Dax Toolkit: A Proposed Framework for Data Analysis and Visualization at Extreme Scale," Kenneth Moreland, Utkarsh Ayachit, Berk Geveci, and Kwan-Liu Ma. *Proceedings of the IEEE Symposium on Large-Scale Data Analysis and Visualization*, October 2011.
6. "The ParaView Coprocessing Library: A Scalable, General Purpose In Situ Visualization Library," Nathan Fabian, Kenneth Moreland, David Thompson, Andrew C. Bauer, Pat Marion, Berk Geveci, Michel Rasquin, and Kenneth E. Jansen. *Proceedings of the IEEE Symposium on Large-Scale Data Analysis and Visualization*, October 2011.
7. "Large-Scale Data Visualization Using Parallel Data Streaming," James Ahrens, Kristi Brislawn K., Ken Martin, Berk Geveci, Charles Law, Mike Papka, *IEEE Computer Graphics & Applications*, July 2001

Synergistic Activities

- Dr. Geveci regularly publishes and teaches courses at conferences including IEEE Visualization and Supercomputing conferences.
- VTK Architecture Review Board member

Recent Collaborators

James Ahrens, LANL
E. Wes Bethel, LBNL
Peer-Timo Bremer, LLNL
Eric Brugger, LLNL
Surendra Byna, LBNL
Hank Childs, University of Oregon
Charles Doutriaux, LLNL
Attila Gyulassy, University of Utah
Salman Habib, ANL
Charles Hansen, University of Utah
Cyrus Harrison, LLNL
Katrin Heitmann, ANL
Mark Hereld, ANL
Joe Insley, ANL
Kenneth Jansen, UC Boulder
Chris Johnson, University of Utah
Ken Joy, UC Davis
Wesley Kendall, Ambition, LLC
Scott Klasky, ORNL
Rob Latham, ANL
Burlen Loring, LBNL
Kwan-Liu Ma, UC Davis
Jeremy Meredith, ORNL
Michael Papka, ANL
Manish Parashar, Rutgers University
Valerio Pascucci, University of Utah
Tom Peterka, ANL
Norbert Podhorszki, ORNL
Michel Rasquin, ANL
Robert Ross, ANL
Oliver Rübel, LBNL
Allen Sanderson, University of Utah
Christopher Sewell, Intuitive Surgical
Han-Wei Shen, Ohio State
Arie Shoshani, LBNL
Venkat Vishwanath, ANL
Gunther Weber, LBNL
Matt Wolf, ORNL
Kesheng (John) Wu, LBNL

Graduate Advisor and Advisees

J.D.A. Walker, Lehigh University (deceased)

Biographical Sketch: Neil Fortner

Neil Fortner
1800 S. Oak Street, Suite 203
Champaign, IL 61820
nfortne2@hdfgroup.org

Education and Training

University of Maryland, College Park, MD BS, Aerospace Engineering 2004

Research and Professional Experience

2008 – present Software Engineer, The HDF Group.

- Extensive programming on the core HDF5 library, fixing many bugs and adding many features
- Implemented new software packages including a threading engine (AXE) and a plugin to allow remote access to HDF5 (h5netvol)
- Managing a project funded by a contract with a National Lab to add features to HDF5
- Wrote documentation for various features
- Gave HDF5 tutorials for new users

2004 – 2005 Research Assistant, The University of Illinois at Urbana-Champaign

- Studied the problem of finding nearly optimal paths for Unmanned Air Vehicles
- Adapted algorithms for the traveling salesperson problem for this purpose, including Lin-Kernighan

2003 – 2004 Engineering Intern, Naval Surface Warfare Center, Carderock Division, Bethesda, MD

- Designed and coded a MATLAB user interface to facilitate the use of empirical mode decomposition (EMD), a time-frequency method for signal processing
- Discovered an effective way to use such a method to detect damage in some structures
- Coded an EMD routine in C for integration into an existing analysis program

Publications

- Venkatesan, V., Chaarawi, M., Koziol, Q., Fortner, N., & Gabriel, E. (2016). A framework for collective I/O style optimisations at staging I/O nodes. International Journal of Big Data Intelligence, 3(2).

Synergistic Activities

- Regular development work on the HDF5 library and related software, with a focus on HPC

Collaborators and Co-editors

- Quincey Koziol, LBNL
- Mohamad Chaarawi, Intel
- Johann Lombardi, Intel
- Vishwanath Venkatesan, Intel
- Venkat Vishwanath, ANL

Graduate and Postdoctoral Advisors and Advisees

- Natasha Neogi, National Institute of Aerospace

Pascal Grosset

Los Alamos National Laboratory
Computer, Computational, and Statistical Sciences Division
PO Box 1663, Mail Stop B287
Los Alamos, NM 87544
pascalgrosset@lanl.gov

Education and Training

2016	PhD	Computer Graphics and Visualization	University of Utah
2004	MSc	Computer Graphics	University of Teeside, England
2003	BSc	Computer Science	University of Mauritius, Mauritius

Research and Professional Experience

Los Alamos National Laboratory	Staff Scientist	Apr 2017 – present
Los Alamos National Laboratory	Postdoc Research Associate	Sep 2016 – Mar 2017
SCI Institute, University of Utah	Postdoc Fellow	May 2016 – Aug 2016
SCI Institute, University of Utah	Research Assistant	Aug 2009 – May 2016
University of Mauritius	Lecturer	Jan 2005 – July 2009

Selected Publications

- A.V. Pascal Grosset, A. Knoll, C.D. Hansen Dynamically Scheduled Region-Based Image compositing, In Proceedings of the 16th Eurographics Symposium on Parallel Graphics and Visualization (PGV '16). Eurographics Association, Aire-la-Ville, Switzerland, Switzerland.
- A. V. P. Grosset, M. Prasad, C. Christensen, A. Knoll and C. Hansen, "TOD-Tree: Task-Overlapped Direct Send Tree Image Compositing for Hybrid MPI Parallelism and GPUs," in IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 6, pp. 1677-1690, June 1 2017
- A.V. Pascal Grosset, M. Prasad, C. Christensen, A. Knoll, C.D. Hansen TOD-Tree: Task-Overlapped Direct send Tree Image Compositing for Hybrid MPI Parallelism, In Proceedings of the 15th Eurographics Symposium on Parallel Graphics and Visualization (PGV '15). Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 67-76.
- M. Schott, T. Martin, A.V.P. Grosset, S.T. Smith, C.D. Hansen. Ambient Occlusion Effects for Combined Volumes and Tubular Geometry , In IEEE Transactions on Visualization and Computer Graphics (TVCG), Vol. 19, No. 6
- A.V. Pascal Grosset, M. Schott, G-P Bonneau, C.D. Hansen Evaluation of Depth of Field for Depth Perception in DVR, Visualization Symposium (PacificVis), 2013 IEEE Pacific, Sydney, NSW, 2013, pp. 81-88. doi:10.1109/PacificVis.2013.6596131
- M. Schott, T. Martin, A.V.P. Grosset, C. Brownlee, T. Hollt, B. P. Brown, S.T. Smith, C. D. Hansen. Combined Surface and Volumetric Occlusion Shading, Visualization Symposium (PacificVis), 2012 IEEE Pacific, Songdo, 2012, pp. 169-176.
- M. Schott, A.V.P. Grosset, T. Martin, V. Pegoraro, S.T. Smith, C.D. Hansen. Depth of Field Effects for Interactive Direct Volume Rendering, In Proceedings of Eurographics/IEEE Symposium on Visualization 2011, Vol. 30, No. 3, Edited by H. Hauser, H. Pfister, and J. J. van Wijk, 2011
- A.V. Pascal Grosset, Peihong Zhu, Shusen Liu, Suresh Venkatasubramanian, Mary Hall. Evaluating Graph Coloring on GPUs, ACM SIGPLAN Principles and Practice of Parallel Programming (PPoPP), Feb. 2011, San Antonio, Texas, USA PPoPP'11, February 12-16, 2011.

Synergistic Activities

Reviewer IEEE Eurovis, IEEE Cluster

Collaborators (last 48 months) and Co-Editors (last 24 months)

Hansen, Charles (University of Utah), Knoll, Aaron (University of Utah), Sanderson, Allen (University of Utah), Junghans, Christoph (LANL), Bergen, Ben (LANL), Daniel, David (LANL), Woodring, Jon (LANL), Heitmann, Katrin (ANL), Habib, Salman (ANL), Almgren, Ann (LBL), Lukic, Zarija (LBL),

Graduate and Postdoctoral Advisors and Advisees:

Charles Hansen (University of Utah)

Hanqi Guo
9700 S. Cass Ave.
Argonne IL 60439

Phone 630-252-7225
Email hguo@mcs.anl.gov
Web http://mcs.anl.gov/~hguo

Education

- **Doctor of Philosophy** Computer Science. Peking University, Beijing, China. July, 2014. Dissertation: "Scalable Visual Analysis of Pathlines in Large-Scale Flow Field Data."
- **Bachelor of Science** Mathematics and Applied Mathematics. Beijing University of Posts and Telecommunications, Beijing, China. July, 2009.

Appointment

- **Argonne National Laboratory** Argonne IL, since August 2017, Assistant Computer Scientist, Mathematics and Computer Science Division.
- **Argonne National Laboratory** Argonne IL, August 2014–August 2017. Postdoctoral Appointee, Mathematics and Computer Science Division.

Related Publications

- **Hanqi Guo**, Wenbin He, Tom Peterka, Han-Wei Shen, Scott M. Collis, and Jonathan J. Helmus, "Finite-Time Lyapunov Exponents and Lagrangian Coherent Structures in Uncertain Unsteady Flows." *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE PacificVis '16)*, 22(6):1672–1682, 2016.
- Jiang Zhang, **Hanqi Guo**, Xiaoru Yuan, Fan Hong, and Tom Peterka, "Dynamic Load Balancing Based on Constrained K-D Tree Decomposition for Parallel Particle Tracing." *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE SciVis '17)*, 2017. (Accepted)
- **Hanqi Guo**, Jiang Zhang, Richen Liu, Lu Liu, Xiaoru Yuan, Jian Huang, Xiangfei Meng, and Jingshan Pan, "Advection-based Sparse Data Management for Visualizing Unsteady Flow." *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE SciVis '14)*, 20(12):2555–2564, 2014.
- **Hanqi Guo**, Xiaoru Yuan, Jian Huang, and Xiaomin Zhu, "Coupled Ensemble Flow Line Advection and Analysis." *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE SciVis '13)*, 19(12):2733–2742, 2013.

Other Significant Publications

- **Hanqi Guo**, Tom Peterka, and Andreas Glatz, "In Situ Magnetic Flux Vortex Visualization in Time-Dependent Ginzburg-Landau Superconductor Simulations." In *Proceedings of IEEE Pacific Visualization Symposium (PacificVis '17)*, pages 71–80, Seoul, Korea, April, 18–21, 2017.
- **Hanqi Guo**, Carolyn L. Phillips, Tom Peterka, Dmitry Karpeyev, and Andreas Glatz, "Extracting, Tracking, and Visualizing Vortices in 3D Complex-Valued Superconductor Simulation Data." *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE SciVis '15)*, 22(1):827–836, 2016.

Synergistic Activities

- **Conference Committees:** IEEE VIS 2015-2017 Program Committees; IEEE VIS 2015 and 2017 Organizing Committee; IEEE PacificVis 2016-2017 Program Committee; Eurographics Symposium on Parallel Graphics and Visualization (EGPGV) 2017 Program Committee
- **Selected Reviews:** IEEE VIS, EuroVis, IEEE PacificVis, IEEE Transactions on Visualization and Computer Graphics, Computer Graphics Forum, The Visual Computer, Journal of Visualization, Information Visualization, National Science Foundation

Mary W. Hall
School of Computing, University of Utah, 50 S. Central Campus Dr., Salt Lake City, UT 84112
ph: (801) 585-1089 fax: (801) 581-5843 mhall@cs.utah.edu http://www.cs.utah.edu/~mhall

Professional Preparation

Rice University	Computer Science/Mathematical Sciences	B.S.	1985
Rice University	Computer Science	M.S.	1989
Rice University	Computer Science	Ph.D.	1991

Appointments

July 2011 to present: *Professor, School of Computing, University of Utah*
September 2008 to June 2011: *Associate Professor, School of Computing, University of Utah*
March 2003 to August 2008: *Research Associate Professor, USC Dept. of Computer Science, and Project Leader, USC Information Sciences Institute*.
June 1996 to February 2003: *Research Assistant Professor, USC Dept. of Computer Science, and Project Leader, USC Information Sciences Institute*.
January 1995 to May 1996: *Visiting Assistant Professor and Senior Research Fellow, Dept. of Computer Science, California Institute of Technology.*
June 1992 to December 1994: *Research Scientist, Center for Integrated Systems, Stanford University*.
October 1990 to June 1992: *Research Scientist, Center for Research on Parallel Computation, Rice University.*

Selected Publications

1. ``Compiler Transformation to Generate Hybrid Sparse Computations," Huihui Zhang, Anand Venkat and Mary Hall, in Proceedings of the Sixth Workshop on Irregular Applications: Architectures and Algorithms (IA^3 2016), held in conjunction with SC16, November 2016.
2. "Optimizing LOBPCG: Sparse Matrix Loop and Data Transformations in Action," Khalid Ahmad, Anand Venkat and Mary Hall, LNCS, 2017, Volume 10136, Languages and Compilers for Parallel Computing, Springer Verlag, Pages 221-231.
3. "Automating Wavefront Parallelization for Sparse Matrix Codes,"A. Venkat, M. Mohamadi, J. Park, R. Barik, H. Rong, M. Strout, M. Hall, SC'16, Nov. 2016. ***Best Paper Finalist.***
4. "Synchronization Tradeoffs in GPU Implementations of Graph Algorithms," R. Kaleem, A. Venkat, S. Pai, M. Hall, K. Pingali, Proceedings of the IEEE International Parallel and Distributed Processing Symposium (IPDPS), May 2016.
5. "Generating Efficient Tensor Contractions for GPUs," T. Nelson, A. Rivera, P. Balaprakash, M. Hall, P.D. Hovland, E. Jessup, B. Norris, Proceedings of the IEEE International Conference on Parallel Processing (ICPP),Sept. 2015.
6. "Loop and Data Transformations for Sparse Matrix Code," A. Venkat, M. Hall, M. Strout, Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI), June 2015.
7. "Compiler-Directed Transformation for Higher-Order Stencils", Protonu Basu, Samuel Williams, Brian Van Straalen, Mary Hall, Leonid Oliker, Phillip Colella, International Parallel and Distributed Processing Symposium (IPDPS), 2015.
8. "Non-affine Extensions to Polyhedral Code Generation," Anand Venkat, Manu Shantharam, Michelle Strout and Mary Hall. IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2014).
9. "Compiler generation and autotuning of communication-avoiding operators for geometric multigrid", 20th International Conference on High Performance Computing (HiPC), 2013.
10. "Speeding up Nek5000 with Autotuning and Specialization," J. Shin, M.W. Hall, J. Chame, C. Chen, P.F. Fischer, P.D. Hovland, *Proceedings of the International Conference on Supercomputing, June 2010.*

Synergistic Activities

1. **Mentoring programs:** Univ. of Utah Presidential Commission on the Status of Women (member); Univ. of Utah Women in Engineering Advisory Council (chair; School of Computing Diversity Committee (chair); Organizer SC16 Students@SC Mentoring Program; Organizer ACM SIGPLAN Programming

Languages Mentoring Workshop 2016; Invited panelist and selection committee for CRA-W Programming Languages Summer School, 2007; PhD Scholarship Committee for Grace Hopper Celebration of Women in Computing Conference (GHC), 2006, 2007, 2011, 2012, 2013; PhD Forum committee member for GHC, Oct., 2004; Invited panelist at the Committee on the Status of Women in Computing Research (CRA-W) workshop, July, 2003; Directed undergraduate in Caltech CRPC *Summer Program in Computing for Undergraduate Women,* 1995; Mentor for PhD student in Stanford American Women in Science Mentoring Program, 1992-93; Directed undergraduate in Rice CRPC *Spend the Summer with a Scientist Program*, 1991.

2. **Conference and Workshop Organization:** Served on more than 50 program committees in compilers, parallel computing, embedded computing and architecture, including general chair for PLDI 2011, program chair for PPOPP 2010, CGO 2009 and PLDI 2005, program co-chair for PACT 2003, and Software Area chair for SC 2004. Also, Technical Papers Co-Chair for SC14, Silver Anniversary Chair for SC13, Exhibits Chair for SC12, Awards Chair for SC09, Workshop Co-Chair for SC07 and SC08. Created the Software Tools for Multi-Core Systems (STMCS) Workshop and organized it in 2006 and 2007. Organized the HPCSW Autotuning of Applications and Libraries workshop, 2008.
3. **Service to field:** Computing Research Association Board of Directors, 2015-present; ACM History Committee (2005-2015, Chair 2009-2014); Member of the ACM Committee on the Health of Conferences, a task force identifying best practices for encouraging innovation in conference submissions; Associate Editor, ACM Computing Surveys

Biographical Sketch

Mary Hall joined University of Utah in September 2008 as an associate professor. Her current research focus is on auto-tuning compiler technology, targeting conventional multi-core and many-core GPU architectures as sockets in high-end systems. She works with application and library developers to understand how to accelerate the human process of performance tuning, and with architects to understand how to maximize the performance potential of novel systems. Under DOE funding, she serves as the lead PI on the X-TUNE X-Stack project, and the Performance area lead on DOE SciDAC SUPER project. She was previously autotuning lead on the SciDAC PERI Institute (precursor to SUPER), and lead PI on TUEN, an ASCR-funded collaboration with Argonne and USC/ISI to apply compiler-based auto-tuning to accelerate DOE applications on jaguar. She was a principal investigator in a SciDAC-e project with USC's EFRC. Hall was an active participant in the 2008-2009 DARPA Exascale Software Study, and organized the discussion and writing of enabling technologies for tools.

Collaborators and Other Affiliations

Graduate advisors: Ken Kennedy, Keith Cooper, Linda Torczon, Rice University.

Postdoctoral advisors: John Hennessy, Monica Lam, Stanford University.

Advisees: Sungdo Moon, Hewlett-Packard, Santa Clara, CA; Byoungro So, Intel Research, Santa Clara, CA; Jaewook Shin, Argonne National Laboratories, Argonne, IL; Heidi Ziegler, Boeing, El Segundo, CA; Melina Demertzi, Yoon-Ju Lee, Nelson, Intel, Phoenix, AZ; Muhammad Murtaza, Norwegian Institute of Science and Technology; Protonu Basu, Lawrence Berkeley National Laboratory; Saurav Muralidharan, Nvidia Research; Anand Venkat, Intel Research

Collaborators: Jacqueline Chame, Pedro Diniz, Robert Lucas, Yolanda Gil, Ewa Deelman, Kristina Lerman (USC/ISI); Priya Vashishta, Rajiv Kalia, Aiichiro Nakano, Murali Annavaram (USC); Craig Steele (Exogii Systems); Rob Fowler (RENCI); Michelle Strout (University of Arizona); Catherine Olschanowsky (Boise State); Michael Garland, Duane Merrill, Albert Sidelnk (Nvidia)

Kevin Harms
9700 South Cass Ave
Lemont, IL 60439
harms@alcf.anl.gov

Education

Purdue University - West Lafayette, IN
Bachelor of Science - Computer Science, May 1998, GPA: 3.55

Illinois Institute of Technology - Chicago, IL
Master of Science - Computer Science, August 2006, GPA: 4.00

Work Experience

Argonne National Laboratory - Lemont, IL (Leadership Computing Facility)

Lead, I/O Libraries & Benchmarks, July 2016 – Present
Performance engineer providing architectural input for design of future I/O systems
Contributing developer on MCS I/O team

Senior Software Developer, July 2007 – July 2016
Software developer working on the production ALCF file systems including PVFS and GPFS.
Contributing developer on MCS I/O team

Publications

"Modular HPC I/O characterization with Darshan" - Shane Snyder, Philip Carns, Kevin Harms, Robert Ross, Glenn K Lockwood, Nicholas J Wright, 2016/11/13, Proceedings of the 5th Workshop on Extreme-Scale Programming Tools

"Impact Of Data Placement on Resilience in Large-Scale Object Storage Systems" - Philip Carns, Kevin Harms, John Jenkins, Misbah Mubarak, Robert Ross, Christopher Carothers, 2016/5/6, 32ND INTERNATIONAL CONFERENCE ON MASSIVE STORAGE SYSTEMS AND TECHNOLOGY (MSST 2016)

"A Multiplatform Study of I/O Behavior on Petascale Supercomputers" - Huong Luu, Marianne Winslett, William Gropp, Robert Ross, Philip Carns, Kevin Harms, Mr Prabhat, Suren Byna, Yushu Yao, 2015/6/15, Proceedings of the 24th International Symposium on High-Performance Parallel and Distributed Computing on 33-44, ACM

"A Case for Epidemic Fault Detection and Group Membership in HPC Storage Systems" - Shane Snyder, Philip Carns, Jonathan Jenkins, Kevin Harms, Robert Ross, Misbah Mubarak, Christopher Carothers, 2014/11/16, High Performance Computing Systems. Performance Modeling, Benchmarking, and Simulation on 237-248, Springer International Publishing

"Scalable parallel I/O on a Blue Gene/Q supercomputer using compression, topology-aware data aggregation, and subfiling" - Huy Bui, Hal Finkel, Venkatram Vishwanath, Salman Habib, Katrin Heitmann, Jason Leigh, Michael Papka, Kevin Harms, 2014/2/12, Parallel, Distributed and Network-Based Processing (PDP), 2014 22nd Euromicro International Conference on 107-111, IEEE

"Toward a unified object storage foundation for scalable storage systems" - C Karakoyunlu, D Kimpe, P Carns, K Harms, R Ross, L Ward, Cluster Computing (CLUSTER), 2013 IEEE International Conference on, 1-8

"Production I/O Characterization on the Cray XE6" - P Carns, Y Yao, K Harms, R Latham, R Ross, K Antypas, Proceedings of the Cray User Group meeting 2013

"A Case for Optimistic Coordination in HPC Storage Systems" - Philip Carns, Kevin Harms, Dries Kimpe, Robert Ross, Justin Wozniak, Lee Ward, Matthew Curry, Ruth Klundt, Geoff Danielson, Cengiz Karakoyunlu, John Chandy, Bradley Settlemeyer, William Gropp, 2012/11/10, High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion

"Understanding and improving computational science storage access through continuous characterization" - Philip Carns, Kevin Harms, William Allcock, Charles Bacon, Samuel Lang, Robert Latham, Robert Ross, 2011/10/1, ACM Transactions on Storage (TOS)

"I/O performance challenges at leadership scale"- Samuel Lang, Philip Carns, Robert Latham, Robert Ross, Kevin Harms, William Allcock, 2009/11/14, Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis

Paul D. Hovland
Mathematics and Computer Science Division
Argonne National Laboratory, Argonne, IL 60439-4844

Education

1997	Ph.D.	Computer Science	University of Illinois at Urbana-Champaign, IL, U.S.A.
1993	M.S.	Computer Science	Michigan State University, East Lansing, MI, U.S.A.
1991	B.S.	Computer Engineering	Michigan State University, East Lansing, MI, U.S.A.

Research and Professional Experience

- Deputy Division Director, Argonne National Laboratory, MCS Division, 2015 – present
- Senior Computer Scientist, Argonne National Laboratory, MCS Division, 2015 – present
- Senior Fellow, Computation Institute, University of Chicago, 2008 – present
- Computer Scientist, Argonne National Laboratory, MCS Division, 2003 – 2015
- Assistant Computer Scientist, Argonne National Laboratory, MCS Division, 1998 – 2003
- Fellow, Computation Institute, University of Chicago, 2001 – 2008
- Enrico Fermi Scholar, Argonne National Laboratory, MCS Division, 1997 – 1998

Selected Publications

- J. Hückelheim, P. Hovland, M. M. Strout, and J.-D. Müller. *Reverse-mode algorithmic differentiation of an OpenMP-parallel compressible flow solver*, The International Journal of High Performance Computing Applications, to appear, http://dx.doi.org/10.1177/1094342017712060.

- F. Isaila, P. Balaprakash, S. M. Wild, D. Kimpe, R. Latham, R. Ross, and P. Hovland. *Collective I/O tuning using analytical and machine learning models*, 2015 IEEE International Conference on Cluster Computing, 2015.

- T. Nelson, A. Rivera, P. Balaprakash, M. Hall, P. D. Hovland, E. Jessup, and B. Norris. *Generating efficient tensor contractions for GPUs*, International Conference on Parallel Processing, 2015.

- P. Balaprakash, L. A. B. Gomez, M.-S. Bouguerra, S. M. Wild, F. Cappello, and P. D. Hovland. *Analysis of the tradeoffs between energy and run time for multilevel checkpointing*, Proceedings of the 5th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS14), New Orleans, 2014.

- P. Balaprakash, D. Buntinas, A. Chan, A. Guha, R. Gupta, S. H. K. Narayanan, A. A. Chien, P. D. Hovland, and B. Norris. *Exascale workload characterization and architecture implications*. 21st High Performance Computing Synposium, San Diego, 2013.

- P. Balaprakash, K. Rupp, A. Mametjanov, R. B. Gramacy, P. D. Hovland, and S. M. Wild. *Empirical performance modeling of GPU kernels using active learning*, Proceedings of the International Conference on Parallel Computing - ParCo2013, Munich, 2013.

- P. Balaprakash, S. M. Wild, and P. D. Hovland. *Can Search Algorithms Save Large-Scale Automatic Performance Tuning?*, Proceedings of the Iternational Workshop on Automatic Performance Tuning, Singapore, June 2011.

- J. Shin, M. W. Hall, J. Chame, C. Chen, P. F. Fischer, P. D. Hovland. *Speeding up Nek5000 with Autotuning and Specialization*, Proceedings of the 24th International Conference on Supercomputing (ICS'10), June 1-4, 2010, Tsukuba, Japan.

- S. Krishna Narayanan, B. Norris, and P. Hovland. *Generating performance bounds from source code*. 39th International Conference on Parallel Processing Workshops (ICPPW), pages 197–206, Sept 2010.

- I. Safro, P. Hovland, J. Shin, and M. Strout. *Improving Random Walk Performance*, In the Proceedings of the International Conference on Scientic Computing (CSC), July 2009.

Collaborators (48 months) and Co-Editors (24 months)

Guillaume Aupy (INRIA), Torsten Bosse (Friedrich Schiller University Jena), Mohamed Slim Bouguerra (ENSIMAG), Darius Buntinas (Jump Trading), Laura Carrington (UCSD), Bronis de Supinski (Lawrence Livermore National Laboratory), Anthony Chan (MathWorks), Andrew Chien (University of Chicago), Chekuri S. Choudary (RNET), Jack Dongarra (University of Tennessee), Anshu Dubey (LBNL), Robert Fowler (RENCI), Leonardo A. Bautista Gomez (Barcelona Supercomputing Center), Robert Gramacy (Virginia Tech), Apala Guha (IIT Delhi), Mary Hall (Utah), Jeffrey Hollingsworth (University of Maryland), Jan Hueckelheim (Imperial College London), Elizabeth Jessup (Colorado), Dries Kimpe (KCG Holdings), PMBS 2014 Robert Lucas (USC ISI), Shirley Moore (University of Texas El Paso), Jens-Dominik Mueller (Queen Mary), Mahesh Narayanamurthi (Virginia Tech), Thomas Nelson (Colorado), Boyana Norris (University of Oregon), Daniel Quinlan (LLNL), Axel Rivera (Intel), Yves Robert (INRIA), Amit Roy (drchrono), Karl Rupp (TU Vienna), Gerald Sabin (RNET), Markus Schordan (LLNL), Stephen Siegel (Delaware), Michelle Strout (Arizona), Ananta Tiwari (UCSD), Jean Utke (Allstate), Andrea Walter (Paderborn), Sam Williams (LBNL), Patrick Worley (ORNL),

Graduate and postdoctoral advisors

Lionel Ni (M.S.), Hong Kong University of Science and Technology
Michael T. Heath (Ph.D.), University of Illinois at Urbana-Champaign
Christian Bischof (postdoc), TU Darmstadt

Predoctoral-scholars supervised and current affiliations

Priyadarshini Malusare, Flextrade Systems

Postdoctoral-scholars supervised and current affiliations

Prasanna Balaprakash (Argonne), Sanjukta Bhowmick (University of Nebraska - Omaha), Jong Gyun Kim (Argonne), Azamat Mametjanov (Argonne), Uwe Naumann (RWTH Aachen), Boyana Norris (University of Oregon), Ilya Safro (Clemson), Jaewook Shin (NVIDIA), Michelle Mills Strout (University of Arizona), Jean Utke (Allstate)

Kevin A. Huck

Education and Training

Spanish Ministry of Science and Innovation *Juan de la Cierva* Postdoctoral Fellowship, Barcelona Supercomputing Center, 2009-2011
Ph.D., Computer and Information Science, University of Oregon, 2009
M.S., Computer and Information Science, University of Oregon, 2004
B.S., Computer Science, University of Cincinnati, 1995

Research and Professional Experience

2012 – present, Research Associate, Neuroinformatics Center, University of Oregon
2011 – 2012, Computer Scientist, ParaTools, Inc., Eugene, OR
2009 – 2011, Senior Researcher, Barcelona Supercomputing Center, Spain
2002 – 2009, Graduate Research Assistant, Neuroinformatics Center, University of Oregon
2001 – 2002, Senior Software Engineer, Southwest Financial Services, Ltd., Cincinnati OH
1997 – 2001, Senior Systems Engineer, Triple-I Systems, Inc., Cincinnati OH
1992 – 1997, Software Engineer, International TechneGroup, Inc., Milford, OH

Related Publications

1. M. A. S. Bari, N. Chaimov, A. M. Malik, K. A. Huck, B. Chapman, A. D. Malony, and O. Sarood, "ARCS: Adaptive Runtime Configuration Selection for Power-Constrained OpenMP Applications," in *IEEE Cluster*, 2016.
2. K. A. Huck, K. Potter, D. W. Jacobsen, H. Childs, and A. D. Malony, "Linking Performance Data into Scientific Visualization Tools," in *Proceedings of the First Workshop on Visual Performance Analysis*, pp. 50–57, IEEE Press, 2014.
3. K. Huck, A. Porterfield, N. Chaimov, H. Kaiser, A. Malony, T. Sterling, and R. Fowler, "An Autonomic Performance Environment for Exascale," *Supercomputing Frontiers and Innovations*, vol. 2, no. 3, 2015.
4. A. Sarje, S. Song, D. Jacobsen, K. Huck, J. Hollingsworth, A. Malony, S. Williams, and L. Oliker, "Parallel performance optimizations on unstructured mesh-based simulations," *Procedia Computer Science*, vol. 51, no. 0, pp. 2016 – 2025, 2015. International Conference On Computational Science, (ICCS) 2015 Computational Science at the Gates of Nature.
5. A. D. Malony and K. A. Huck, "General Hybrid Parallel Profiling," in *Parallel, Distributed and Network-Based Processing (PDP), 2014 22nd Euromicro International Conference on*, pp. 204–212, IEEE, 2014.
6. K. A. Huck, A. D. Malony, S. Shende, and D. W. Jacobsen, "Integrated Measurement for Cross-Platform OpenMP Performance Analysis," in *IWOMP 2014: Using and Improving OpenMP for Devices, Tasks, and More*, pp. 146–160, Springer International Publishing, 2014.
7. K. Huck and J. Labarta, "Detailed Load Balance Analysis of Large Scale Parallel Applications," in *39th International Conference on Parallel Processing (ICPP 2010)*, pp. 535–544, 2010.
8. K. A. Huck, A. D. Malony, S. Shende, and A. Morris, "Knowledge Support and Automation for Performance Analysis with PerfExplorer 2.0," *Scientific Programming, special issue on Large-Scale Programming Tools and Environments*, vol. 16, no. 2-3, pp. 123–134, 2008.

9. K. A. Huck, O. Hernandez, V. Bui, S. Chandrasekaran, B. Chapman, A. D. Malony, L. C. McInnes, and B. Norris, "Capturing Performance Knowledge for Automated Analysis," in *SC '08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, (Piscataway, NJ, USA), pp. 1–10, IEEE Press, 2008.
10. K. A. Huck and A. D. Malony, "PerfExplorer: A Performance Data Mining Framework For Large-Scale Parallel Computing," in *Proceedings of the 2005 ACM/IEEE Conference on Supercomputing*, SC '05, (Washington, DC, USA), pp. 41–, IEEE Computer Society, 2005.

Synergistic Activities

Tutorial Program Committee, SC 2016-2018, Salt Lake City, UT.
Technical Program Committee, ICCS 2016, San Diego, CA.
Web Chair, VECPAR 2014, Eugene, OR.
Technical Program Committee, ICS 2014, Munich, Germany.
Web Chair, ICS 2013, Eugene, OR.

Collaborators and Co-editors

Hasan Abbasi (Amazon), Abhinav Bhatele (LLNL), Ron Brightwell (SNL), Laura Carrington (UCSD), Nicholas Chaimov (ParaTools), Barbara Chapman (SUNY Stonybrook), Hank Childs (UO), Anthony Danalis (UTK), Bronis de Supinski (LLNL), Greg S. Eisenhauer (GA Tech), Stephane Ethier (PPPL), Robert Fowler (RENCI), Edgar Gabriel (UH), Todd Gamblin (LLNL), Mary Hall (UU), Mike Heroux (SNL), Jeff Hollingsworth (UM), Paul Hovland (ANL), Doug Jacobsen (Intel), Hartmut Kaiser (LSU), Alice Koniges (LBL), Andrew Lumsdaine (PNNL), Abid Muslim Malik (BNL), Heike McCraw (UTK), Kathryn Mohror (LLNL), Shirley Moore (ORNL), Boyana Norris (UO), Leonid Oliker (LBNL), David Ozog (Intel), Allan Porterfield (LogicBlox), Philip Roth (ORNL), Thomas Sterling (IU), Jeffrey Vetter (ORNL), Stefan Wild (ANL), Samuel Williams (LBL), Matthew Wolf (GA Tech), Chad Wood (UO), Patrick Worley (ORNL).

Graduate and Postdoctoral Advisors and Advisees

Doctoral Advisor Allen D. Malony, University of Oregon, Eugene, OR.
Post-Doctoral Advisor Jesus Labarta, Barcelona Supercomputing Center, Barcelona, Spain.
Ph.D. Students and Postdocs None.

Costin Catalin Iancu
Senior Staff Scientist
Computational Research Division
Lawrence Berkeley National Laboratory

Email: cciancu@lbl.gov
Phone: (510) 495-2122

A. EDUCATION

- ***1996-2001 PhD Computer Science***, University of California, Santa Barbara
- ***1994-1995 Advanced Degree in Computer Science***, Politehnica University of Bucharest
- ***1989-1994 BS/MS Computer Science***, Politehnica University of Bucharest

B. EXPERIENCE

- ***2016-, Senior Staff Scientist, CRD,*** Lawrence Berkeley National Laboratory
- ***2001-2016, Staff Scientist***, CRD, Lawrence Berkeley National Laboratory
- ***1998-2000, Systems Software Engineer***, Green Hills Software
- ***1994-1995, Software Design Engineer***, Computer Aided Finance (Bucharest,Romania)

C. SELECTED PUBLICATIONS

i. Most relevant to proposed activity:

Maximizing Communication Overlap with Dynamic Program Analysis . Emmanuelle Saillard , Koushik Sen, Wim Lavrijsen, Costin Iancu. Submitted 2017.

*Application Level Reordering of Remote Direct Memory Access Operations.*Wim Lavrijsen and Costin Iancu. Proceedings IEEE International Parallel & Distributed Processing Symposium (IPDPS 2017).

SReplay: Deterministic Sub-Group Replay for One-Sided Communication. Xuehai Qian, Koushik Sen, Paul Hargrove, Costin Iancu, *International Conference on Supercomputing (ICS), 2016,* February 5, 2016

Barrier Elision for Production Parallel Programs. Milind Chabbi, John Mellor-Crummey, Koushik Sen, Wim Lavrijsen, Wibe de Jong, Costin Iancu, PPoPP 2015

Scaling Data Race Detection for Partitioned Global Address Space Programs. Chang-Seo Park, Koushik Sen, Costin Iancu**,** International Conference on Supercomputing (ICS) 2013.

ii. Other:

Floating-point precision tuning using blame analysis . Cuong Nguyen, Cindy Rubio-Gonzalez, Benjamin Mehne, Koushik Sen, Costin Iancu … *38th International Conference on Software Engineering (ICSE 2016),* January 20, 2016.

Precimonious: Tuning Assistant for Floating-Point Precision. Cindy Rubio-Gonzalez, Cuong Nguyen, Hong Diep Nguyen, James Demmel,…, Supercomputing 2013.

*Characterizing the Performance of Parallel Applications on Multi-Socket Virtual Machines. (***Best Paper Award***)* Khaled Ibrahim, Steven Hofmeyr, Costin Iancu. Proceedings of the 11th IEEE/ACM International Symposium in Cluster, Cloud and Grid Computing (CCGRID 2011).

Scientific Application Performance on Candidate Petascale Platforms (**Best Paper Award**). Leonid Oliker, Andrew Canning, Jonathan Carter, Costin Iancu et al. Proceedings of the 21st IEEE International Parallel & Distributed Processing Symposium (IPDPS 2007).

Hybrid PGAS Runtime Support for Multicore Nodes. Filip Blagojevic, Paul Hargrove, Costin Iancu, Katherine Yelick. Proceedings of the 4th Partitioned Global Address Space Languages Conference (PGAS 2010).

D. SYNERGISTIC ACTIVITIES

- PI Throughput Oriented Runtime Design (THOR), DoD
- Co-PI Berkeley Institute for Compilers, Languages and Runtime Research (Binoculars), DoD
- Main architect of the Berkeley UPC compiler deployed in production at multiple DOE sites. BUPC was also shipped as the default installation by Cray on their XT4/5 series. 46,000 downloads since project inception.

RECENT COLLABORATORS (LBNL,UCB, MIT excluded)

Brad Chamberlain(Cray), Stephane Ethier(Princeton), Mary Hall(U of Utah), Barbara Chapman (U Houston), Richard Graham (ORNL), Scott Klasky (ORNL), D. K. Panda (Ohio State), John Owens (UC Davis), Vivek Sarkar (Rice), John Mellor Crummey (Rice), Mattan Erez (UT Austin), Craig Rasmussen (LANL), Boyana Norris (ANL), Ron Brightwell (Sandia), Kurt Ferreira (Sandia), Jim Laros (Sandia), Jay Lofstead (Sandia), Ron Oldfield (Sandia), Kevin Pedretti (Sandia), Mike Lang (LANL), Pat McCormick (LANL) Arthur B. Maccabe (ORNL), David Bernholdt (ORNL), Christian Engelmann (ORNL), Karsten Schwan (Georgia Tech), Thomas Sterling (Indiana U), Andrew Lumsdaine (Indiana U), Frank Mueller (NC State), Peter Dinda (Northwestern), Eric Brewer (UC Berkeley), Patrick Bridges (UNM), Jack Lange (U Pittsburg)

Graduate Advisors: Urs Holzle (Google), Anurag Acharya (Google)

Postdoctoral students: Seung-Jai Min (LBL), Cindy Gonzales (UC Davis), Xuehai Qian (UC Berkeley), Emmanuelle Saillard (UC Berkeley)

Curriculum Vitae
Scott A. Klasky
Contact Information

Education and Training

Ph.D. in Physics (1994), University of Texas, Austin, TX.
B.S. in Physics (1989), Drexel University.

Research and Professional Experience

2014 – Present, Distinguished Scientist, ORNL
2014 – Present, Adjunct Faculty, Georgia Technical University, Atlanta GA.
2012 – Present, Group Leader: Scientific Data, Computer Science and Mathematics, ORNL,
2009 – Present, Adjunct Professor, Dept. of Electric and Computer Science, University of Tennessee Knoxville, Knoxville TN.
2007 – Present, Adjunct Professor, Dept. of Information Technology, North Carolina State University, Raleigh, NC.
2005 – Present, Visiting Professor, Dept. of Electrical and Computer Engineering, Rutgers University, Piscataway NJ.
2005 – 2011, Senior Research Scientist, and End-to-End Task Lead, Oak Ridge National Laboratory, Oak Ridge, TN.
1999 – 2005, Senior Research Scientist, Princeton Plasma Physics Laboratory, Princeton NJ.
1995 – 1999, Senior Research Scientist, Syracuse University (Northeast Parallel Architecture Center, Syracuse NJ.

Five Publications Most Relevant to This Proposal (hindex=33)

1. J. F. Lofstead, S. Klasky, K. Schwan, N. Podhorszki, C. Jin, Flexible io and integration for scientific codes through the adaptable io system (adios) in Proceedings of the 6th international workshop on Challenges of large applications in distributed environments, ACM, pp. 15–24. (229 citations, 1/24/2017)
2. H. Abbasi, M. Wolf, G. Eisenhauer, S. Klasky, K. Schwan, F. Zheng. Datastager: scalable data staging services for petascale applications. *Cluster Computing* 2010, *13*, 277–290. (193 citations, 1/24/2017)
3. J. Lofstead, F. Zheng, S. Klasky, K. Schwan, Adaptable, metadata rich IO methods for portable high performance IO in Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on, IEEE, pp. 1–10. (125 citations, 1/24/2017)
4. F. Zheng, H. Abbasi, C. Docan, J. Lofstead, Q. Liu, S. Klasky, M. Parashar, N. Podhorszki, K. Schwan, M. Wolf, PreDatA–preparatory data analytics on peta-scale machines in Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on, IEEE, pp. 1–12. (104 citations, 1/24/2017)
5. J. Lofstead, F. Zheng, Q. Liu, S. Klasky, R. Oldfield, T. Kordenbrock, K. Schwan, M. Wolf, Managing variability in the IO performance of petascale storage systems in Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, IEEE Computer Society, pp. 1–12. (111 citations, 1/24/17)

Research Interests and Expertise

Dr. Klasky has over 25 years of experience in designing and creating middleware, visualization, mathematical algorithms, and numerical simulations for high performance- data intensive computing. Dr. Klasky leads many large efforts, and has funding of over $6M/year, and serves as the technical lead and coordinator on all of these projects. Scott is one of the world's leaders in Research into Data Intensive

Science, and is the lead of the ADIOS project, used by many of the largest data producers in the world. Dr. Klasky is starting to lead a new effort which combines the use of Service Oriented Arthictures for High End Computing with advanced data streaming techniques which can process voluminous data from remote resources and efficiently use all of the devices.

The core mission around my work is to perform research and development in core solutions for data intensive computing. I lead a large team of over 75 scientists to research and build data handling infrastructure for moving, reducing, analyzing, visualizing, and understanding massive amounts of data produced by current and next generation facilities, both computational and experimental. My approach is rooted in deep engagements with real science end-users. As such, my team is a strong partner with leading application teams to co-design new algorithms, middleware, and end-to-end systems. My overarching goal is to create infrastructure that enables new capabilities for scientists to make Timely (Near-Real-Time) and Quality-based decisions.

Synergistic Activities

1. Head of the ADIOS Project, (https://www.olcf.ornl.gov/center-projects/adios/
2. Co-PI on the SDAV SciDAC data institute.
3. Co-PI on the CODAR ECP project.
4. PI of the ECP ADIOS project.
5. Co-PI on the ECP WDM project.

Collaborators (*past 5 years including name and current institution*)-ORNL are conflicts

PhD Advisor: R. Matzner (U. Texas at Austin)
Postdoctoral Advisees: H. Abbasi (Amazon), J. Choi (ORNL), C. Gu (Linked In), Q. Liu (NJIT), J. Logan (UTK), N. Podhorszki (ORNL), Y. Tian (Amazon), R. Wang (ORNL), Lipeng Wan(ORNL)

Abbasi, Hasan, Amazon
Adams, Mark, LBNL
Agrawal, Ankit, NWU
Ahrens, James, LANL
Aiken, Alexander, Stanford
Altintas, Ilkay, UCSD
Bell, John, LBNL
Bennett, Janine, SNL
Bethel, Wes, LBNL
Brugger, Eric, LLNL
Chen, Jackie, Sandia
Childs, Hank, U. Oregon
Choudhary, Alok, NWU
Ethier, Stephane, PPPL
Feibush, Elliott, PPPL
Fox, Geoffrey, U Indiana
Gavrilovska, Ada, GT
Geveci, Berk, Kitware
Grout, Ray, NREL
Hanrahan, Patrick, Stanford
Hansen, Charles, U. Utah
Johnson, Chris, U. Utah
Kolla, Hemanth, SNL
Kothe, Doug, ORNL
Kurc, Tahsin, SBU SUNY
Latham, Rob, ANL
Lee, Wei-Li, PPPL
Liao, Wei-keng, NWU
Liu, Qing, Oak Ridge
National Laboratory
Lofstead, Jay, Sandia
National Laboratories
Logan, Jeremy, UTK
Ludaescher, Bertram, UCD
Ma, Xiaosong, Qatar
Research Foundation
Ma, Kwan-Liu, U.C. Davis
Maccabe, Arthur, Oak Ridge
National Laboratory
Matzner, Richard, U Texas
Austin
McCormick, Patrick, LANL
Moreland, Kenneth, SNL
Moser, Robert, U. T. Austin
Oldfield, Ron, SNL
Papka, Michael, ANL
Parker, Scott, U. of Colorado
Pascucci, Valerio, U. of Utah
Peterka, Tom, ANL
Quinlan, Dan, LLNL
Ross, Rob, ANL
Saltz, Joel, SBU
Samatova, Nagiza, NCSU
Sanderson, Allen, Utah
Shalf, John, LBNL
Shen, Han-Wei, OSU
Shephard, Mark, RPI
Shoshani, Arie, LBNL
Siegel, Andrew, ANL
Taufer, Michela, U. Delaware
Tchoua, Roselyne, Chicago
Vishwanath, Venkat, ANL
Widener, Patrick, SNL
Wolf, Matthew, GT Tech
Wu, Kesheng, LBNL
Yu, Weikuan, U. Flo

Robert J. Latham
9700 South Cass Avenue
Argonne, IL 60439-4844
robl@mcs.anl.gov

Education and Training

1999–2000	Lehigh University (Bethlehem, PA) Masters of Science, Computer Engineering
1995–1999	Lehigh University (Bethlehem, PA) Bachelor of Science, Computer Engineering

Research and Professional Experience

2013-present:	**Mathematics and Computer Science Division** **Argonne National Laboratory** Principle Software Development Specialist
2005-2013:	**Mathematics and Computer Science Division** **Argonne National Laboratory** Software Development Specialist
2002–2005:	**Mathematics and Computer Science Division** **Argonne National Laboratory** Software Developer
2000–2002:	**Paralogic, Inc.** Systems Developer

Current Activities

- **Parallel netCDF I/O Interface** (`http://www.mcs.anl.gov/parallel-netcdf`): Technical Lead for project; worked closely with Northwestern University collaborators to harden and productize research ideas.
- **ROMIO MPI-IO Implementation** (`http://www.mcs.anl.gov/romio`): Technical Lead for project; worked with community to incorporate assorted improvements and modifications; provide feedback, suggestions, and code changes to IBM for improvements to MPI-IO library (based on ROMIO) on BlueGene.

Publications Most Relevant to Research

1. Rob Latham, Matthieu Dorier, and Rob Ross, "Get out of the way! Applying compression to internal data structures", Parallel Data Storage and data Intensive Scalable Computing Systems (PDSW-DISCS), 2016.
2. Rob Latham, Chris Daley, Wei-keng Liao, Kui Gao, Robert Ross, Anshu Dubey, Alok Choudhary," A case study for scientific I/O: improving the FLASH astrophysics code". Computational Science & Discovery 5 (1), 015001
3. Philip Carns, Kevin Harms, William Allcock, Charles Bacon, Samuel Lang, Robert Latham, and Robert Ross. "Understanding and Improving Computational Science Storage Access through Continuous Characterization". Trans. Storage, 7(3), 2011.

4. Florin Isaila, Javier Garcia Blas, Jesus Carretero, Robert Latham, Robert Ross, "Design and Evaluation of Multiple-Level Data Staging for Blue Gene Systems". IEEE Transactions on Parallel and Distributed Systems,22 (6), 946-959

5. Samuel Lang, Philip Carns, Robert Latham, Robert Ross, Kevin Harms, and William Allcock. "I/O Performance Challenges at Leadership Scale," in Proceedings of SC2009: High Performance Networking and Computing, November 2009.

6. W. Kendall, M. Glatter, J. Huang, T. Peterka, R. Latham, and R. Ross, "Terascale Data Organization for Discovering Multivariate Climatic Trends," in Proceedings of Supercomputing, November, 2009.

7. Javier Garca Blas, Florin Isaila, Jess Carretero, Robert Latham, and Robert Ross "Multiple-level MPI file Write-Back and Prefetching for Blue Gene Systems," in Proc. of the 16th European PVM/MPI User's Group Meeting (Euro PVM/MPI 2009), September 2009

8. Robert Latham, Robert Ross, and Rajeev Thakur, "Implementing MPI-IO Atomic Mode and Shared File Pointers Using MPI One-Sided Communication," in the International Journal of High Performance Computing Applications, 21(2):132-143, 2007.

9. Rob Latham, Rob Ross, and Rajeev Thakur. "The Impact of File Systems on MPI-IO Scalability". Lecture Notes in Computer Science, 3241:87-96 (EuroPVM/MPI 2004), September 2004.

10. J. Li, W. Liao, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, and M. Zingale, "Parallel netCDF: A High-Performance Scientific I/O Interface," in Proceedings of SC2003, Phoenix, AZ, November 2003.

Synergistic Activities

- **Parallel I/O in Practice**, Regularly present full-day tutorial at annual SC conference (2005-2017)
- **Argonne Training Program for Exascale Systems (ATPESC)**, Data and I/O day, 2013-2017
- **Interfaces and Abstractions for Scientific Data Storage (IASDS)**, Organizer, 2010 and 2011
- **Argonne Pacesetter award** for exceptional work in production deployment of ALCF BG/P storage system (2009)

Biographical Sketch

Wei-keng Liao
Research Professor
Department of Electrical Engineering and Computer Science
Northwestern University
2145 Sheridan Rd, Rm L-455
Evanston, IL 60208
Phone: (847) 491-2916
Email: w-liao2@northwestern.edu

A. EDUCATION AND TRAINING

College/University	Location	Major	Degree	Year
National Chung-Hsing University	Taichung, Taiwan	Applied Mathematics	BS	1988
Syracuse University	Syracuse, NY	Computer & Information Science	PhD	1999

B. RESEARCH AND PROFESSIONAL EXPERIENCE

2012-present	Research Professor, Department of Electrical Engineering and Computer Science, Northwestern University
2008-2012	Research Associate Professor, Department of Electrical Engineering and Computer Science, Northwestern University
1999-2008	Research Assistant Professor, Department of Electrical Engineering and Computer Science, Northwestern University

C. PUBLICATIONS

1. S. Son, S. Sehrish, W. Liao, R. Oldfield, and A. Choudhary. Reducing I/O Variability Using Dynamic I/O Path Characterization in Petascale Storage Systems. In the Journal of Supercomputing, vol. 73, issue 5, pp. 2069-2097, May 2017.
2. Z. Yuan, W. Hendrix, S. Son, C. Federrath, A. Agrawal, W. Liao, and A. Choudhary. Parallel Implementation of Lossy Data Compression for Temporal Data Sets. In the 23rd International Conference on High Performance Computing, December 2016.
3. D. Palsetia, W. Hendrix, S. Lee, A. Agrawal, W. Liao, and A. Choudhary. Parallel Community Detection Algorithm Using a Data Partitioning Strategy with Pairwise Subdomain Duplication. In the 31st International Supercomputing Conference, June 2016.
4. Z. Chen, S. W. Son, W. Hendrix, A. Agrawal, W. Liao, and A. Choudhary, "NUMARCK: Machine Learning Algorithm for Resiliency and Checkpointing," in Proceedings of 26th International Conference on High Performance Computing, Networking, Storage and Analysis, November 2014, pp. 733–744.
5. Q. Kang, W. Liao, A. Agrawal, and A. Choudhary. A Filtering-based Clustering Algorithm for Improving Spatio-temporal Kriging Interpolation Accuracy. In 25th ACM International Conference on Information and Knowledge Management, October 2016.
6. E. Rangel, N. Li, S. Habib, T. Peterka, A. Agrawal, W. Liao, and A. Choudhary. Parallel DTFE Surface Density Field Reconstruction (best paper award). In the IEEE International Conference on Cluster Computing, September 2016.
7. W. Liao and R. Thakur. "MPI-IO." Chapter 13 in *High Performance Parallel I/O*, edited by Prabhat and Quincey Koziol, pp. 155-165. 1st Edition. Boca Raton: Chapman and Hall/CRC, 2014.
8. M. Patwary, D. Palsetia, A. Agrawal, W. Liao, F. Manne, and A. Choudhary. Scalable Parallel OPTICS Data Clustering Using Graph Algorithmic Techniques. In the International Conference for High Performance Computing, Networking, Storage and Analysis, November 2013.

9. W. Liao. Design and Evaluation of MPI File Domain Partitioning Methods under Extent-Based File Locking Protocol, The *IEEE Transactions on Parallel and Distributed Systems*, **22**, no. 2, pp. 260-272, Feb. 2011.
10. J. Li, W. Liao, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, and M. Zingale. Parallel netCDF: A Scientific High-Performance I/O Interface. In the Proceedings of the Supercomputing Conference, November 2003.

D. SYNERGISTIC ACTIVITIES

- Technical program committee of the 29th International Conference on Scientific and Statistical Database Management (SSDBM), Chicago, June 2017
- Technical program committee of the 18th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing (CCGrid), Madrid, Spain, May 2017.
- Technical papers committee of the International Conference for High Performance Computing, Networking, Storage and Analysis, Salt Lake City, Utah, 2016.
- Technical program committee of the 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Cartagena, Colombia, 2016.
- Technical program committee of the 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS), Chicago, 2016.

E. IDENTIFICATION OF POTENTIAL CONFLICTS OF INTESEST OR BIAS IN SELECTION OF REVIEWERS

- **Collaborators and Co-editors**
  Ankit Agrawal (Northwestern U.), Ruben Buaba (NC A&T State U.), Zhengzhang Chen (NEC), Alok Choudhary (Northwestern U.), Auroop Ganguly (Northeastern U.), Salman Habib (U. Chicago), Dianwei Han (Mount Sinai Medicine School), William Hendrix (Northwestern U.), Abdollah Homaifar (NC A&T State U.), Vassiliki Kalogera (Northwestern U.), Mahmut Kandemir (Penn State U.), Seong Jo Kim (Penn State U.), Quincey Koziol (HDF group), Rob Latham (ANL), Jianwei Liao (Southwest University of China), Ruoqian Liu (Uber), Frederik Manne (U. of Bergen), Ron Oldfield (SNL), Bharath Pattabiraman (Google), Diana Palsetia (4C Insights), Mostofa Patwary (Intel), Frederic Rasio (Northwestern U.), Robert Ross (ANL), Nagiza Samatova (ORNL), Saba Sehrish (FNAL), Frederick Semazzi (NCSU), Seung Woo Son (UMass Lowell), Tim Tautges (ANL), Rajeev Thakur (ANL), John Wu (LBNL), Kunpeng Zhang (UIC)
- **Graduate and Postdoctoral Advisees**
  None
- **Graduate Advisors**
  Alok Choudhary (Northwestern U.), Promad Varshney (Syracuse U.), Donald Weiner (Syracuse U.)

Seyong Lee — ORNL

Education and Training

Graduate	Purdue University, West Lafayette, Indiana	Electrical and Computer Engineering	PhD	2011
Graduate	Purdue University, West Lafayette, Indiana	Electrical and Computer Engineering	MS	2004
Undergraduate	Seoul National University, Seoul, South Korea	Electrical Engineering	BS	1999

Research and Professional Experience

5/2011–present	Computer Scientist, Computer Science and Mathematics Division, Oak Ridge National Laboratory – Developed an Open Accelerator Research Compiler (OpenARC), which supports various heterogeneous architectures, ranging from NVIDIA/AMD GPUs to Intel Xeon Phis and Altera FPGAs. – Developed a programming system (Papyrus) that provides a portable and scalable programming interface to access and manage parallel data structures on the distributed NVM storage.
10/2006–4/2011	Research Assistant, Electrical and Computer Engineering, Purdue University – OpenMP to GPU: Automatic translation and adaptation of OpenMP shared-memory programs onto GPUs – iShare: Internet-sharing middleware and collaboration – MaRCO: MapReduce with Communication Overlap
9/2009–12/2009	Software Engineer (Intern), NEEScomm, Purdue University – Developed a HUBzero-based cloud computing system for NEES (Network for Earthquake Engineering Simulation).

Synergistic Activities

- IEEE-CS TCHPC Award for Excellence for Early Career Researchers in High Performance Computing (https://www.computer.org/web/pressroom/TCHPC-Award-2016).
- Member of 1) OpenACC Technical Forum (www.openacc.org), 2) Science Council, Computer Science and Mathematics Division, Oak Ridge National Laboratory, and 3) Award Committee for 2017 IEEE CS TCHPC Award for Excellence for Early Career Researchers in High Performance Computing, 2017
- The paper on the high-level GPU programming ("OpenMP to GPGPU: A Compiler Framework for Automatic Translation and Optimization") was selected as the most cited paper among all papers published in the Symposium on Principles and Practices of Parallel Programming (PPOPP) between 2009 and 2014 (cited 489 times as of June 2017).
- Program Committee Member : ASPLOS(2018), IPDPS(2017), Euro-Par (2017), ICPADS(2013, 2014, 2015, 2016, and 2017), PPoPP (2014), CCGrid (2015, 2016, and 2017), ICPP (2013), CANDAR (2016), PLC (2015), WRAp (2015 and 2017), WACCPD (2014, 2015, 2016, and 2017), AsHES (2016 and 2017), LHAM (2016 and 2017)
- External Reviewer (Journals, Conferences, Workshops, and Research Proposals)

Related and Recent Publications

[1] J. Kim, S. Lee, and J. S. Vetter. PapyrusKV: A High-Performance Parallel Key-Value Store for Distributed NVM Architectures. In *The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, SC '17, 2017.

[2] J. Kim, K. Sajjapongse, S. Lee, and J. S. Vetter. Design and Implementation of Papyrus: Parallel Aggregate Persistent Storage. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, IPDPS '17, 2017.

[3] J. E. Denny, S. Lee, and J. S. Vetter. Language-Based Optimizations for Persistence on Nonvolatile Main Memory Systems. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, IPDPS '17, 2017.

[4] S. Lee, J. Kim, and J. S. Vetter. OpenACC to FPGA: A Framework for Directive-based High-Performance Reconfigurable Computing. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2016.

[5] J. Kim, S. Lee, and J. S. Vetter. IMPACC: A Tightly Integrated MPI+OpenACC Framework Exploring Shared Memory Parallelism. In *Proceedings of the ACM Symposium on High-Performance and Distributed Computing (HPDC)*. 2016.

[6] J. E. Denny, S. Lee, and J. S. Vetter. NVL-C: Static Analysis Techniques for Efficient, Correct Programming of Non-Volatile Main Memory Systems. In *Proceedings of the ACM Symposium on High-Performance and Distributed Computing (HPDC)*. 2016.

[7] S. Lee, J. S. Meredith, and J. S. Vetter. COMPASS: A Framework for Automated Performance Modeling and Prediction. In *ACM International Conference on Supercomputing (ICS)*, 2015.

[8] S. Lee and J. S. Vetter. OpenARC: Open Accelerator Research Compiler for Directive-Based, Efficient Heterogeneous Computing. In *Proceedings of the ACM Symposium on High-Performance and Distributed Computing (HPDC)*. Short Paper, 2014.

[9] S. Lee, D. Li, and J. S. Vetter. Interactive Program Debugging and Optimization for Directive-Based, Efficient GPU Computing. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2014.

[10] S. Lee and R. Eigenmann. OpenMPC: Extended OpenMP programming and tuning for GPUs. In *the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*. **Best Student Paper**, IEEE press, 2010.

Collaborators and Co-Editors

Vikram S. Adve (UIUC), Saurabh Bagchi (Purdue), Ayon Basumallik (MathWorks), Volker Blum (Duke), Ryan Braby (NICS), Sunita Chandrasekaran (U. of Delaware), Barbara Chapman (StonyBrook), Tom Conte (GT), Anthony Danalis (UTK), Chirag Dave (Qualcomm), Kei Davis (LANL), Joel E. Denny (ORNL), Jack Dongarra (UTK), Rudolf Eigenmann (Purdue), James Ferguson (NICS), Hal Finkel (ANL), Richard Glassbrook (GT), John Gounley (Duke), Gregory Herschlag (Duke), Kevin A. Huck (U. of Oregon), Wen-Mei W. Hwu (UIUC), Jungwon Kim (ORNL), Sang-Ik Lee (Intel), Dong Li (UC Merced), Bruce Loftis (NICS), Allen D. Malony (U. of Oregon), Gabriel Marin (Google), Jeremy Meredith (Google), Samuel Midkiff (Purdue), Nick Moss (LANL), Vijay Pai (Purdue), Zhelong Pan (VMware), Amanda Randles (Duke), Xiaojuan Ren (VMware), James Rogers (ORNL), Philip Roth (ORNL), Karsten Schwan (GT), Sameer S. Shende (U. of Oregon), Mithuna Thottethodi (Purdue), Jeffrey Vetter (ORNL), T. N. VijayKumar (Purdue), Richard W. Vuduc (GT), Michael Wolfe (PGI), Tian Xinmin (Intel), Sudhakar Yalamanchili (GT)

Graduate and Postdoctoral Advisors and Advisees

Graduate Advisor: Rudolf Eigenmann, Purdue University

Contact Information

Dr. Chunhua "Leo" Liao
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory, L-495
Livermore, CA 94551-0808
Phone: (925) 422-8190 , Fax: (925) 424-2831
Email: liao6@llnl.gov

Education and Training

1999	B.E.	Computer Science,	Sichuan University, Chengdu, China
2002	M.E.	Computer Science,	Sichuan University, Chengdu, China
2007	Ph.D	Computer Science,	University of Houston, Texas, USA
2010	Postdoc	Computer Science,	Lawrence Livermore National Laboratory, California, USA

Research and Professional Experience

04/2010 – Present	Computer Scientist, Lawrence Livermore National Laboratory
09/2007 – 04/2010	Postdoctoral Research Fellow, Lawrence Livermore National Laboratory
01/2004 – 08/2007	Research Assistant, University of Houston

Related Publications

1. Chunhua Liao, Pei-Hung Lin, Joshua Asplund, Markus Schordan and Ian Karlin, *DataRaceBench: A Benchmark Suite for Systematic Evaluation of Data Race Detection Tools*, The International Conference for High Performance Computing, Networking, Storage and Analysis, Denver, CO, Nov. 12-17, 2017 (Accepted, Best Paper Nominee).
2. Pei-Hung Lin, Qing Yi, Daniel Quinlan, Chunhua Liao and Yongqing Yan, *Automatically Optimizing Stencil Computations on Many-core NUMA Architectures*, The 29th International Workshop on Languages and Compilers for Parallel Computing (LCPC 2016) September 28-30, 2016 Rochester NY, USA
3. Yue Zhao, Guoyang Chen, Chunhua Liao and Xipeng Shen, *Towards Ontology-Based Program Analysis*, The European Conference on Object-Oriented Programming (ECOOP), Sun 17 - Fri 22 July 2016 Rome, Italy
4. Yonghong Yan, Pei-Hung Lin, Chunhua Liao, Bronis R. de Supinski, and Daniel J. Quinlan, *Supporting multiple accelerators in high-level programming models*, In Proceedings of the Sixth International Workshop on Programming Models and Applications for Multicores and Manycores (PMAM '15), Feb 7-8, 2015, San Francisco, USA
5. Jacob Lidman, Sally McKee, Daniel Quinlan and Chunhua Liao, *An Automated Performance-Aware Approach to Reliability Transformations*, Euro-Par 2014, 25-29 August, Porto, Portugal
6. Hongyi Ma , Steve Diersen, Liqiang Wang, Chunhua Liao, Daniel J. Quinlan and Zijiang Yang, *Symbolic Analysis of Concurrency Errors in OpenMP Programs*, 42nd International Conference on , vol., no., pp.510,516, 1-4 Oct. 2013
7. Chunhua Liao, Yonghong Yan, Bronis R. de Supinski, Daniel J. Quinlan and Barbara Chapman, *Early Experiences With The OpenMP Accelerator Model*, 9th International Workshop on OpenMP (IWOMP), Canberra, Australia, Sept. 16-18, 2013.
8. Chunhua Liao, Daniel J. Quinlan, Thomas Panas and Bronis R. de Supinski, *A ROSE-based OpenMP 3.0 Research Compiler Supporting Multiple Runtime Libraries*, International Workshop on OpenMP (IWOMP) 2010, Tsukuba, Japan. June 14-26, 2010

9. Chunhua Liao, Daniel J. Quinlan, Jeremiah J. Willcock and Thomas Panas, *Semantic-Aware Automatic Parallelization of Modern Applications Using High-Level Abstractions*, International Journal of Parallel Programming, Volume 38, Issue 5-6 , pp 361-378, October 2010
10. Chunhua Liao, Daniel J. Quinlan, Richard Vuduc and Thomas Panas, *Effective Source-to-Source Outlining to Support Whole Program Empirical Optimization*, The 22nd International Workshop on Languages and Compilers for Parallel Computing, Newark, Delaware, USA. October 8-10, 2009

Collaborators and Co-editors Barbara Chapman (University of Houston), Bronis R. de Supinski (LLNL), Craig Rasmussen (University of Oregon), Jacqueline Chame (University of Southern California, Information Sciences Institute), Jeff Hollingsworth(University of Maryland), John Mellor-Crummey (Rice Univ.), Liqiang Wang (University of Central Florida), Mary Hall (University of Utah), P. (Saday) Sadayappan (Ohio State University), Pedro Diniz (University of Southern California-ISI), Peter Pirkelbauer(University of Alabama at Birmingham), Qing Yi (University of Colorado at Colorado Springs), Robert Lucas (USC-ISI), Sally A. McKee (Chalmers University of Technology), Stephen Guzik (Colorado State University), Vivek Sarkar (Rice Univ.), Xipeng Shen (North Carolina State University), Yonghong Yan (Oakland University)

Graduate and Postdoctoral Advisors and Advisees

Postdoctoral Advisor:	Dan Quinlan, Lawrence Livermore National Laboratory
Ph.D. Advisor:	Barbara Chapman, University of Houston
M.E. Advisor:	Hongwei Zhang, Sichuan University, China

Advisee(s): Post-doc: Pei-Hung Lin (Lawrence Livermore National Laboratory), Students Interns: Kewen Meng (Univ. of Oregon), Jordan Riley(Colorado State Univ.), Weilin Wang (College of William & Mary), Weidong Wang, Hongyi Ma (University of Wyoming), Cong Hou (Google), Akshatha Bhat (Mathworks), Tristan Vanderbruggen (University of Delaware), Sara Royuela Alcazar (Technical University of Catalonia), Wooil Kim (University of Illinois at Urbana-Champaign)

Allen D. Malony

Education and Training

1980	B.S.	Mathematics/Computer Science	University of California, Los Angeles
1982	M.S.	Computer Science	University of California, Los Angeles
1991	Ph.D.	Computer Science	University of Illinois, Urbana-Champaign

Research and Professional Experience

2016 – 2017	Visiting Professor, University of Versailles Saint-Quentin-en-Yvelines, France
2003 – present	Professor, Computer and Information Science, University of Oregon
2004 – present	CEO, ParaTools, Inc.
1999 – 2000	Visiting Associate Professor, Technical University Vienna, Austria
1996 – 2003	Associate Professor, Computer and Information Science, University of Oregon
1991 – 1996	Assistant Professor, Computer and Information Science, University of Oregon
1991 – 1992	Visiting Assistant Professor, Utrech University, The Netherlands

Awards

Fulbright-Tocqueville Distinguished Chair, University of Versailles St.-Quentin-en-Yvelines, 2017
IBM Faculty Award, 2016
NVIDIA Professor Partnership, 2009
Research Innovation Award, University of Oregon, 2006
Humboldt Research Award, Alexander von Humboldt Foundation, Germany, 2002
Fulbright Research Scholar, University of Vienna, Austria, 1999
National Young Investigator, National Science Foundation, 1994
Fulbright Research Scholar, Utrecht University, The Netherlands, 1991

Related Publications

1. R. Lim, B. Norris, A. Malony, "Autotuning GPU Kernels via Static and Predictive Analysis," *46th International Conference on Parallel Processing (ICPP)*, August 2017.
2. D. Ellsworth, T. Patki, M. Schulz, B. Rountree, A. Malony, "A Unified Platform for Exploring Power Management Strategies," *4th Workshop on Energy Efficient Super Computing (E2SC 2016)*, SC'16, 8 pages, November 2016.
3. C. Wood, S. Sane, D. Ellsworth, A. Gimenez, K. Huck, T. Gamblin, A. Malony, "A Scalable Observation System for Introspection and In Situ Analytics," *5th Workshop on Extreme-Scale Programming Tools (ESPT 2016)*, SC'16, 8 pages, November 2016.
4. J.-B. Besnard, J. Adam, S. Shende, M. Perache, P. Carribault, J. Jaeger, A. Malony, "Introducing Task-Containers as an Alternative to Runtime-Stacking," *23rd European MPI Users' Group (Euro MPI 2016)*, pp. 5:1-5:13, September 2016.
5. M. Bari, N. Chaimov, A. Malik, K. Huck, B. Chapman, A. Malony, "ARCS: Adaptive Runtime Configuration Selection for Power-Constrained OpenMP Applications," *IEEE Cluster*, pp. 461-470, September 2016.
6. A. Malony, A. Monil, C. Rasmusen, K. Huck, J. Byrnes, D. Toomey "Towards Scaling Parallel Seismic Raytracing," *19th IEEE International Conferences on Computational Science and Engineering (CSE 2016)*, August 2016.
7. N. Chaimov, A. Malony, S. Canon, C. Iancu, K. Ibrahim, J. Srinivasan, "Scaling Spark on HPC Systems," *25th International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC 2016)*, pp. 132-143, June 2016.

8. D. Ozog, A. Kamil, Y. Zheng, P. Hargrove, J. Hammon, A. Malony, W. de Jong, K. Yelick, "A Hartree-Fock Application using UPC++ and the New DArray Library," *IEEE International Parallel and Distributed Processing Symposium (IPDPS 2016)*, May 2016.
9. K. Huck, A. Porterfield, N. Chaimov, A. Malony, R. Fowler, H. Kaiser, T. Sterling, "An Autonomic Performance Environment for Exascale," *Journal on Supercomputing Frontiers and Innovations*, 2015.
10. D. Ellsworth, A. Malony, M. Schultz, B. Rountree, "Dynamic Power Sharing for Higher Job Throughput," *International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, November 2015.

Synergistic Activities

Director, Oregon Advanced Computing Institute for Science and Society (OACISS), University of Oregon, June 2017 – present. OACISS is a newly-established institute with the mission to advance research partnerships across the university that will benefit from computational and data science opportunities.

Director, NeuroInformatics Center (NIC), University of Oregon, 2002 – present. The NIC is developing advanced integrated neuroimaging tools that combine EEG and MRI methods for next-generation brain analysis.

Director, TAU Performance System project, 1992 – present. The University of Oregon is home to the TAU project, a 18-year old research effort focusing on problems in performance analysis and optimization of parallel applications on large-scale HPC systems. The project distributes the TAU Performance System®, an open source performance analysis tool suite.

General Chair, 8th International Conferences on Partitioned Global Address Space Programming Models (PGAS 2014), Eugene, OR, October 2014.

General Chair, International Meeting on High Performance Computing for Computational Science (VECPAR), Eugene, OR, July 2014.

Collaborators and Co-editors

P. Beckman (ANL), J. Besnard (ParaTools, SAS), R. Brightwell (SNL), L. Carrington (UCSD), B. Chapman (SUNY Stonybrook), H. Childs (UO), B. de Supinski (LLNL), G. Eisenhauer (GA Tech), S. Ethier (PPPL), R. Fowler (RENCI), E. Gabriel (UH), T. Gamblin (LLNL), M. Hall (Utah), J. Hammond (Intel), J. Hollingsworth (UM), P. Hovland (ANL), C. Iancu (LBL), W. Jalby (Versailles), H. Kaiser (LSU), K. Karavanic (PSU), J. Linford (ParaTools, Inc.), A. Lumsdaine (PNNL), J. Mellor-Crummey (Rice), B. Miller (Wisconsin), K. Mohror (LLNL), S. Moore (ORNL), B. Norris (UO), L. Oliker (LBNL), D. Panda (OSU), R. Ross (ANL), P. Roth (ORNL), D. Rouson (Sourcery, Inc.), M. Schulz (LLNL), T. Sterling (IU), J. Vetter (ORNL), A. Wissink (U.S. Army), M. Wolf (ORNL), K. Yelick (LBNL).

Ph.D. Students and Postdocs

Ph.D. thesis advisor:
Nicholas Chaimov, ParaTools, Inc.; Daniel Ellsworth, Colorado College; Geoffrey Hulette, Sandia National Laboratory; Kevin Huck, University of Oregon; Kai Li, Electrical Geodesics, Inc.; David Ozog, Intel Federal; Adnan Salman, An-Najah University; Sameer Shende, University of Oregon
Ph.D. students advised:
Annaliese Johnson, Jacob Lambert, Robert Lim, Erin McCarthy, Alaul Monil, Chad Wood
Total number of graduate students advised: 25

Graduate Advisor

Daniel Reed, University of Iowa

Kshitij V. Mehta

Computer Scientist
Scientific Data Group
Computer Science and Mathematics Division
Oak Ridge National Lab
One Bethel Valley Road
P.O. Box 2008, MS-6057
Oak Ridge, TN 37831-6057
Phone: (865) 574-1739
Email: mehtakv@ornl.gov

[Education]

PhD in Computer Science (2013), University of Houston, TX
MS in Computer Science (2009), University of Houston, TX
BS in Computer Engineering (2005), University of Mumbai, India

[Professional Experience]

2017 – current, Computer Scientist, ORNL
Computer Scientist conducting research into data-intensive scientific computing and high performance I/O

2014 – 2016, Visiting Scientist, Oak Ridge National Lab
Visiting scientist as participant of a CRADA between ORNL and TOTAL E&P. Working on research and acceleration of seismic applications using the OpenACC and OpenMP programming models, and large-scale experiments on the Titan supercomputer.

2013 – 2016, HPC Research Scientist, TOTAL E&P R&T, Houston, TX
Research scientist working on high performance computing in seismic imaging. Research in accelerator technologies for seismic applications.

2012 – 2012, Intern, Los Alamos National Lab (LANL)
Research and development of a plugin for HDF5 using the new Virtual Object Layer for use with PLFS (Parallel Log-Structured File System).

2008 – 2013, Research Assistant, Parallel Software Technologies Lab, University of Houston
Research and development of parallel I/O interfaces for shared-memory programming models such as OpenMP. Presented work to OpenMP committee and community at IWOMP 2012.

[Research Profile]

Dr. Kshitij Mehta's research focusses on evolution of models and paradigms for managing large scale scientific data. He worked on developing a programming interface for parallel I/O for shared memory programming models such as OpenMP as part of his PhD. He conducted research on important factors and metrics for I/O in shared memory programming models. After his PhD, Dr. Mehta worked with a leading oil and gas company in the research department on accelerating seismic applications for accelerator based systems. He worked as a visiting scientist at Oak Ridge National Lab conducting research on directive-based programming models such as OpenACC and OpenMP and their scalability on leadership class supercomputers.

[Relevant Publications]

[1] Mehta K., Gabriel E., Chapman B. (2012) Specification and Performance Evaluation of Parallel I/O Interfaces for OpenMP. In: Chapman B.M., Massaioli F., Müller M.S., Rorro M. (eds) OpenMP in a Heterogeneous World. IWOMP 2012. Lecture Notes in Computer Science, vol 7312. Springer, Berlin, Heidelberg

[2] K. Mehta, J. Bent, A. Torres, G. Grider and E. Gabriel, "A Plugin for HDF5 Using PLFS for Improved I/O Performance and Semantic Analysis," *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, Salt Lake City, UT, 2012, pp. 746-752.
doi: 10.1109/SC.Companion.2012.102
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6495884&isnumber=6495777

[3] Mehta, K. & Gabriel, E. "Multi-Threaded Parallel I/O for OpenMP Applications", International Journal of Parallel Programming (2015) 43: 286. doi:10.1007/s10766-014-0306-9

[Synergistic Activities]

- Member of the OpenACC technical committee
- Played a leading role in the collaboration between industry (TOTAL E&P, Oil & Gas) and Oak Ridge National Laboratory for adopting directive-based programming for accelerating seismic applications on supercomputers such as Titan

[Collaborators]

Bent, John, EMC Corp.
Bernholdt, David E., ORNL
Calandra, Henri, TOTAL E&P
Chapman, Barbara, Stony Brook University
Gabriel, Edgar, University of Houston
Grider, Gary, LANL
Hernandez, Oscar, LANL
Hugues, Maxime, TOTAL E&P
Torres, Aaron, LANL

[Graduate Advisor]

Gabriel, Edgar, University of Houston

Kathryn Mohror

Center for Applied Scientific Computing, L-557
Lawrence Livermore National Laboratory
Livermore, CA 94551

Phone: **(925) 423-2997**
Fax: **(925) 422-6287**
E-mail: **kathryn@llnl.gov**

Research Interests

- Scalable Parallel I/O Systems
- Parallel I/O Performance Measurement
- Scalable Application Performance Measurement
- Parallel Performance Analysis and Visualization
- Fault Tolerance
- Parallel Computing Paradigms

Education

Ph.D	Computer Science, Portland State University, Portland OR, March 2010.
M.S.	Computer Science, Portland State University, Portland OR, March 2004.

Professional Experience

9/12 – present	Computer Scientist, Center for Applied Scientific Computing (CASC), Lawrence Livermore National Laboratory (LLNL), Livermore CA
1/10-9/12	Postdoctoral Research Staff Member, Center for Applied Scientific Computing (CASC), Lawrence Livermore National Laboratory (LLNL), Livermore CA

Synergistic Activities

- MPI Forum Tools Working Group Co-lead, MPI Standard Tools Chapter Chair
- Dagstuhl Seminar on User Level File Systems for HPC 2017 Co-organizer
- EuroMPI/USA 2017 Posters Chair
- WHPC Careers in HPC Workshop 2017 Co-organizer
- SC 2017 Program Committee Area Co-chair (Data, Analytics, and Visualization)
- CLUSTER 2017 Finance Chair
- PDSW-DISCS 2017 PC Co-chair
- HIPS 2013 Co-chair

Selected Relevant Publications

Ignacio Laguna, David F. Richards, Todd Gamblin, Martin Schulz, Bronis R. de Supinski, Kathryn Mohror, and Howard Pritchard, "Evaluating and Extending User-Level Fault Tolerance in MPI", LLNL-JRNL-663434, in International Journal of High Performance Computing Applications, 30(3):305-319, 2016.

Kathryn Mohror, Adam Moody, Greg Bronevetsky, Bronis R. de Supinski, "Detailed Modeling and Evaluation of a Scalable Multilevel Checkpointing System," in Transactions on Parallel and Distributed Systems, LLNL-JRNL-564721, Transactions on Parallel and Distributed Systems, 25(9):2255-2263, Sept. 2014.

Tanzima Zerin Islam, Kathryn Mohror, Saurabh Bagchi, Adam Moody, Bronis R. de Supinski, Rudolf Eigenmann, "McrEngine: A Scalable Checkpointing System Using Data-Aware Aggregation and Compression," LLNL-CONF-554251, Scientific Programming, 21(3):149-163, 2013.

Teng Wang, Kathryn Mohror, Adam Moody, Kento Sato, Weikuan Yu, “An Ephemeral Burst-Buffer File System for Scientific Applications,“ LLNL-CONF-681480, Supercomputing 2016, Salt Lake City Utah, November 2016 (Acceptance rate 18%).

Lee Savoie, David K. Lowenthal, Bronis R. de Supinski, Tanzima Islam, Kathryn Mohror, Barry Rountree, and Martin Schulz, “I/O Aware Power Shifting,” LLNL-CONF-669729, IPDPS 2016, Chicago, IL, May 2016.

Kento Sato, Adam Moody, Kathryn Mohror, Todd Gamblin, Bronis R. de Supinski, Naoya Maruyama, Satoshi Matsuoka, “A User-level Infiniband-based File System and Checkpoint Strategy for Burst Buffers,” LLNL-CONF-645876, 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2014), Chicago, IL, May 2014.

Kento Sato, Adam Moody, Kathryn Mohror, Todd Gamblin, Bronis R. de Supinski, Naoya Maruyama, Satoshi Matsuoka, "FMI: Fault Tolerant Messaging Interface for Fast and Transparent Recovery," LLNL-CONF-645209, 28th IEEE International Parallel & Distributed Processing Symposium (IPDPS 2014), Phoenix, AZ, May 2014.

Raghunath Raja Chandrasekar, Adam Moody, Kathryn Mohror, Dhabaleswar K. Panda, “A 1 PB/s File System to Checkpoint Three Million MPI Tasks,” LLNL-CONF-592884, International Symposium on High Performance Distributed Computing 2013, New York City, NY, June 2013.

Kento Sato, Adam Moody, Kathryn Mohror, Todd Gamblin, Bronis R. de Supinski, Naoya Maruyama, Satoshi Matsuoka, "Design and Modeling of a Non-blocking Checkpointing System," LLNL-CONF-554431, Supercomputing 2012, Salt Lake City, UT, November 2012.

Adam Moody, Greg Bronevetsky, Kathryn Mohror, Bronis R. de Supinski, "Design, Modeling, and Evaluation of a Scalable Multi-level Checkpointing System," LLNL-CONF-427742, Supercomputing 2010, New Orleans, LA, November 2010.

Dmitriy Morozov

Data Analytics and Visualization Group
Computational Research Division
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

www.mrzv.org
Phone: (510) 486-4292
E-mail: dmorozov@lbl.gov

Professional Experience

2016–present	Data Science Fellow, Berkeley Institute for Data Science, UC Berkeley
2013–present	Research Scientist, Lawrence Berkeley National Laboratory
2011–2013	Postdoc, Computational Research Division, Lawrence Berkeley National Laboratory
2008–2010	Postdoc, Departments of Computer Science and Mathematics, Stanford University

Education

Duke University	Computer Science	Ph.D.	2008
NC State University	Applied Math and Computer Science	B.S.	2003
09/2007–08/2008	Visiting Ph.D. Student, Berlin Mathematical School, Berlin, Germany		

Select Recent Publications

1. Dmitriy Smirnov and Dmitriy Morozov. Triplet Merge Trees. *Workshop on Topology-based Methods in Visualization*, 2017. (Best paper award.)
2. Dmitriy Morozov and Tom Peterka. Efficient Delaunay Tessellation through K-D Tree Decomposition. *Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, 2016.
3. Dmitriy Morozov and Tom Peterka. Block-Parallel Data Analysis with DIY2. *Proceedings of the Symposium on Large Data Analysis and Visualization*, 2016.
4. Dmitriy Morozov and Zarija Lukić. Master of Puppets: Cooperative Multitasking for In Situ Processing. *Proceedgins of the Annual Symposium on High-Performance Parallel and Distributed Computing*, pages 285–288, 2016.
5. Ryan Lewis and Dmitriy Morozov. Parallel Computation of Persistent Homology using the Blowup Complex. *Proceedings of the Annual Symposium on Parallelism in Algorithms and Architectures*, pages 323-331, 2015.
6. Tom Peterka, Dmitriy Morozov, and Carolyn Phillips. High-Performance Computation of Distributed-Memory Parallel 3D Voronoi and Delaunay Tessellation. *Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, pages 997–1007, 2014.
7. Dmitriy Morozov and Gunther Weber. Distributed Merge Trees. *Proceedings of the Annual Symposium on Principles and Practice of Parallel Programming*, pages 93–102, 2013.
8. Paul Bendich, Herbert Edelsbrunner, Dmitriy Morozov, and Amit Patel. Homology and Robustness of Level and Interlevel Sets. *Homology, Homotopy, and Applications*, vol. 15, pages 51–72, 2013.
9. Leonidas Guibas, Quentin Mérigot, and Dmitriy Morozov. Witnessed k-Distance. *Discrete and Computational Geometry*, vol. 49, pages 22–45, 2013.
10. Daniela M. Ushizima, Dmitriy Morozov, Gunther H. Weber, Andrea G.C. Bianchi, James A. Sethian, and E. Wes Bethel. Augmented Topological Descriptors of Pore Networks for Material Science. *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE Vis 2012)*, 18:2041–2050, 2012.
11. Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. Persistent Cohomology and Circular Coordinates. *Discrete and Computational Geometry*, vol. 45, pages 737–759, 2011.

Synergistic Activities

1. Editorial board membership: Journal of Applied and Computational Topology (managing editor), SIAM Journal on Applied Algebra and Geometry (associate editor).
2. Program Committee: ACM Symposium on Computational Geometry (2013), SIAM Conference on Applied Algebraic Geometry (2017).
3. Awards: Outstanding PhD disseration award, Dept. of Computer Science, Duke University, 2008.
4. Conference and journal reviewer: Discrete and Computational Geometry, Computational Geometry: Theory and Applications, Foundations of Computational Mathematics, Geometry and Topology, SIAM Journal on Computing (SICOMP), Inverse Problems, SIAM Journal on Imaging Sciences (SIIMS), Symposium on Discrete Algorithms, Foundations of Computer Science, Symposium on Computational Geometry, Symposium on Geometry Processing, International Parallel & Distributed Processing Symposium, TopoInVis, EuroVis.
5. Co-organizer: "Generalized Persistence" at the American Institute of Mathematics (2014); "Applied and Computational Topology" at SIAM Conference on Applied Algebraic Geometry (2013); "Topological Data Analysis and Machine Learning Theory" at Banff International Research Stations (2012); "Computational Topology" at Symposium on Computational Geometry (2012); "Persistent Homology: Applications, Computation, Recent Developments" at SIAM Conference on Applied Algebraic Geometry (2011).

Graduate and Postdoctoral Advisors and Advisees

- Ph.D. Advisor: Herbert Edelsbrunner (Duke University)
- Postdoctoral Advisors: Gunnar Carlsson (Stanford University), Leonidas Guibas (Stanford University), Gunther Weber (LBNL)
- Graduate students: Michael Lesnick (Ph.D. committee, Stanford University), Clément Maria (Ph.D. committee, INRIA Sophia-Antipolis), Melissa Yeung (DOE CSGF practicum, LBNL), Patrick O'Neil (graduate summer intern, LBNL), Pawin Vongmasa (Ph.D. committee, Stanford University), Ryan Lewis (Ph.D. committee, Stanford University), Dmitriy Smirnov (undergraduate summer intern, LBNL), Erich Lohrmann (graduate summer intern, LBNL).

Collaborators and co-editors:
Utkarsh Ayachit (Kitware), Andrew Bauer (Kitware), Vin de Silva (Pomona College), Earl P.N. Duque (Intelligent Light), Greg Eisenhauer (Georgia Tech), Dmitry Feichtner-Kozlov (U Bremen), Nicola Ferrier (ANL), Kenneth E. Jansen (U Colorado Boulder), Michael Kerber (TU Graz), Kevin Knudson (U Florida), Ryan Lewis (Ayasdi), Suresh Menon (Georgia Tech), Quentin Mérigot (CNRS), Arnur Nigmetov (TU Graz), Patrick O'Leary (Kitware), Rateesh Ranjan (Georgia Tech), Michel Rasquin (U Colorado Boulder), Carolyn Phillips (Neurensic), Christopher P. Stone (Computational Science and Engineering, LLC), Bernd Sturmfels (UC Berkeley), Sara Kališnik Verovšek (Brown U), Venkat Vishwanath (ANL), Shmuel Weinberger (U Chicago), Brad Whitlock (Intelligent Light), Matthew Wolf (Georgia Tech), Weixing Wang (PPPL), Stephane Ethier (PPPL), Stephen C. Jardin (PPPL), Francesca Poli (PPPL), Lois Curfman McInnes (ANL), Shashi Aithal (ANL), Todd Munson (ANL), David Eder (LLNL), Xueqiao Xu (LLNL), Ahmed Hassanein (Purdue U), Tatyana Sizyuk (Purdue U)

Boyana Norris
University of Oregon, Department of Computer and Information Science, Eugene, OR 97403
Phone: 541-346-4413; Email: `norris@cs.uoregon.edu`

Education and Training

- Argonne National Laboratory, Postdoctoral Fellow, 1999– 2001
- University of Illinois at Urbana-Champaign, Computer Science, Ph.D., 2000
- Wake Forest University, B.S. *summa cum laude* in Computer Science, 1995

Research and Professional Experience

- *Associate Professor*, University of Oregon, Department of Computer and Information Science, 2015 – present
- *Assistant Professor*, University of Oregon, Department of Computer and Information Science, 2013 – 2015
- *Computer Scientist*, Argonne National Laboratory, Mathematics and Computer Science Division, 2006 – 2013
- *Assistant Computer Scientist*, Argonne National Laboratory, Mathematics and Computer Science Division, 2001 – 2006
- *Senior Fellow*, Computation Institute, University of Chicago and Argonne National Laboratory, 2004 – 2013

Selected Publications Relevant to this Proposal

1. E. Jessup, P. Motter, B. Norris, and K. Sood. Performance-based numerical solver selection in the Lighthouse framework. *SIAM Journal on Scientific Computing* 38, 11 2016,
2. M. Rashti, G. Sabin, and B. Norris. Power and energy analysis and modeling of high performance computing systems using WattProf. *Proceedings of the 2015 IEEE National Aerospace and Electronics Conference (NAECON)*, July 2015.
3. T. Nelson, A. Rivera, M. Hall, P. Hovland, E. Jessup, B. Norris, and P. Balaprakash. Generating efficient tensor contractions for GPUs. *Proceedings of The 44th International Conference on Parallel Processing (ICPP)*, 2015.
4. A. Farzad, B. Norris, and M. Rashti. Portable power/performance benchmarking and analysis with WattProf. *Proceedings of the 6th International Workshop in Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS15)*, Nov 2015.
5. T. Nelson, G. Belter, J. Siek, E. Jessup, and B. Norris. Reliable generation of high-performance matrix algebra. *ACM Transactions on Mathematical Software* 41, June 2015.
6. R. Lim, A. Malony, B. Norris, and N. Chaimov. Identifying optimization opportunities within kernel execution in GPU codes. *Proceedings of the Thirteenth International Workshop on Algorithms, Models and Tools for Parallel Computing on Heterogeneous Platforms (HeteroPar'2015).*
7. X. Dai, B. Norris, and A. D. Malony. Autoperf: Workflow support for performance experiments. *Proceedings of the Workshop on Challenges in Performance Methods for Software Development (WOSP-C'15), January 31, 2015, Austin, Texas*, 1 2015.
8. N. Chaimov, B. Norris, and A. Malony. Toward multi-target autotuning for accelerators. *Proceedings of the 20th IEEE International Conference on Parallel and Distributed Systems, December 16-19, 2014, Hsinchu, Taiwan*, 2014.

9. P. Balaprakash, D. Buntinas, A. Chan, A. Guha, S. H. K. Narayanan, A. A. Chien, P. Hovland, and B. Norris. Exascale workload characterization and architecture implications. *Proceedings of the 21st High Performance Computing Symposium (HPC), San Diego*, April 2013, (**Best paper award**).
10. S. Muszala, J. Amundson, L. C. McInnes, and B. Norris. Two-tiered component design and performance analysis of Synergia2 accelerator simulations. *Proceedings of the 2009 Workshop on Component-Based High Performance Computing (CBHPC 2009)*, Nov. 2009.

Selected Synergistic Activities

- Lead architect and co-developer of the Orio autotuning framework (`http://brnorris03.github.io/Orio/`); and co-designer and developer of source analysis and transformation tools for performance modeling (PBound) and automatic differentiation (ADIC).
- Lead architect of Lighthouse (`http://lighthousehpc.github.io/lighthouse/`), a searchable taxonomy for numerical linear algebra software.
- Designer and co-developer of multiple tools for improving application developer productivity (ADIC, Bocca, CCA infrastructure used in the NSF CSDMS project).
- Argonne PI for the PERI and SUPER DOE SciDAC projects (2006–2013).

Coauthors (48 months) and Co-editors (24 months)

- D. Bernholdt (Oak Ridge National Laboratory), S. Bhowmick (University of Nebraska at Omaha), G. Cerati (Fermilab), C. Choudary (Dimensional Fund Advisors), P. Diniz (USC Information Sciences Institute), A. Dubey (Argonne National Laboratory), R. Fowler (University of North Carolina), W. Gropp (University of Illinois at Urbana-Champaign), M. Hall (University of Utah), M. Heroux (Sandia National Laboratory), J. Hollingsworth (University of Maryland), P. Hovland (Argonne National Laboratory), E. Jessup (University of Colorado at Boulder), R. Lucas (University of Southern California), V. Mahadevan (Argonne National Laboratory), A. Malony (University of Oregon), S. Moore (University of Texas El Paso), L. McInnes (Argonne National Laboratory), R. Narayanan (RNet), L. Oliker (LBNL), S. Peckham (University of Colorado Boulder), P. Sadayappan (Ohio State University), S. Shende (University of Oregon), J. Siek (Indiana University), B. Smith (Oak Ridge National Laboratory), B. de Supinski (Lawrence Livermore National Laboratory), B. Van Straalen (Lawrence Berkeley National Laboratory), S. Williams (Lawrence Berkeley National Laboratory), P. Worley (Oak Ridge National Laboratory).

Advisors

- M. Heath (University of Illinois at Urbana-Champaign, Doctoral)
- W. Gropp (Argonne National Laboratory; now UIUC, Postdoctoral)

Graduate/Postdoctoral Advisees

- Li Li (Unknown), Sri Hari Krishan Narayanan (ANL), Quian Zhu (Accenture), Azamat Mametjanov (ANL)

Manish Parashar
parashar@rutgers.edu
Rutgers, The State University of New Jersey
96 Frelinghuysen Road, Piscataway, New Jersey 08854

Education and Training:

Bombay University, Bombay, India	Electronics & Telecommunications	B.E.	1988
Syracuse University, Syracuse, NY	Computer Engineering	M.S.	1994
Syracuse University, Syracuse, NY	Computer Engineering	Ph.D.	1994
University of Texas at Austin, Austin, TX	Computer Science	Post-doc	1994-95

Research and Professional Experience:

2015-present	Distinguished Professor, Rutgers, The State University of New Jersey
2014-present	Assoc. Vice President, Rutgers Office of Advanced Research Computing
2014-present	Faculty Member, Rutgers Cancer Institute of New Jersey
2012-present	Rutgers Discovery Informatics Institute (RDI^2)
2009-2011	Program Director, OCI, US National Science Foundation
2008-present	Co-Director, Cloud & Autonomic Computing Center, Rutgers University
2008-present	Associate Director, Center for Information Assurance, Rutgers University
2005-2015	Professor I, Rutgers, The State University of New Jersey
2002-2005	Associate Professor, Rutgers, The State University of New Jersey
1997-2002	Assistant Professor, Rutgers, The State University of New Jersey
1996-1997	Adjunct Assistant Professor, University of Texas at Austin
1995-2005	Research Associate, CSM/ICES, University of Texas at Austin

Publications:

"Scheduling and Flexible Control of Bandwidth and In-transit Services for End-to-End Application Workflows," M. Fatih-Aktas, G. Haldeman and M. Parashar, Future Generation Computer Systems, The International Journal of eScience, North-Holland, Vol. 56, pp. 284—294, 2016.

"ActiveSpaces: Exploring Dynamic Code Deployment for Extreme Scale Data Processing," C. Docan, F. Zhang, T. Jin, H. Bui, Q. Sun, J. Cummings, N. Podhorszki, S. Klasky, and M. Parashar, Concurrency and Computation: Practice and Experience, John Wiley & Sons, DOI: 10.1002/cpe.3407, 2014.

"Persistent Data Staging Services for Data Intensive In-Situ Scientific Workflows," M. Romanus, H. Bui, F. Zhang, T. Jin, Q. Sun, J. Choi, S. Janhunen, R. Hager, S. Klasky, C-S Chang and M. Parashar, Proceedings of the Seventh International Workshop on Data Intensive Distributed Computing (DIDC 2016). Kyoto, Japan, June 2016.

"Scalable Run-time Data Indexing and Querying for Scientific Simulations," Q. Sun, F. Zhang, T. Jin, H. Bui, K. Wu, A. Shoshani, H. Kolla, S. Klasky, J. Chen and M. Parashar, 5th International Workshop on Big Data Analytics: Challenges, and Opportunities (BDAC-14), SC'14, New Orleans, LA, USA, Nov. 2014.

Deployed the CometCloud software systems for enabling dynamic software defined infrastructure across federated infrastructure and supporting application workflows in such environments (cometcloud.org); and the DataSpaces data management substrate for enabling extreme scale coupled workflows with in-situ/in-transit analytics and data staging (dataspaces.org).

Other Significant Products:

"CometCloud: Enabling Software-Defined Federations for End-to-End Application Workflows," J. Diaz-Montes, M. AbdelBaky, M. Zou, and M. Parashar. IEEE Internet Computing 19, 69-73 (2015), DOI: 10.1109/MIC.2015.4

"Federated Computing for the Masses – Aggregating Resources to Tackle Large-Scale Engineering Problems," J. Diaz-Montes Y. Xie, I. Rodero, J. Zola, B. Ganapathysubramanian, and M. Parashar, IEEE Computing in Science and Engineering (CiSE) Magazine, IEEE Computer Society Press and AIP, 16, 62-72, 2014. DOI: 10.1109/MCSE.2013.134

"Cloud Paradigms and Practices for Computational and Data-Enabled Science and Engineering,"

M. Parashar, M. Abdelbaky, I. Rodero, IEEE Computing in Science and Engineering (CiSE) Magazine, 15, 10-18 (2013). DOI: 10.1109/MCSE.2013.49

"Market Models for Federated Clouds," I. Petri, J. Diaz-Montes, M. Zou, T. Beach, O. Rana, M. Parashar, IEEE Transactions on Cloud Computing, Special Issue on Cloud Economics, 2015.

Deployed several software systems including DART for high-throughput, low latency data extraction and streaming, AutoMate/Accord/Meteor to support autonomics, GrACE/DAGH and MACE/Seine for dynamically adaptive and coupled simulations.

Synergistic Activities:

1. Assoc. Vice President for Research Computing, Rutgers Office of Advanced Research Computing;
2. Founder of the Rutgers Discovery Informatics Institute; Co-Director of NSF Industry/University Cloud and Autonomic Computing Center; Co-founder of International Conference on Autonomic Computing (ICAC).
3. Leading Advanced Cyberinfrastructure strategic planning at Rutgers. PI/O&M lead in Cyberinfrastructure for the Oceans Observatory Initiative.
4. Co-Editor in Chief, ACM Transactions on Autonomous and Adaptive Systems (TAAS), Associate Editorial in Chief of IEEE TPDS, editorial board of 22 international journals including ACM Computing Surveys, IEEE-TCC, IEEE-TSC, IEEE TBD, IEEE CiSE, IEEE IoT-J, JPDC and CCPE, steering committee of 10 conferences including SC, HPDC, ISC, CAC, HiPC and CCGrid.
5. Program chair (44): MS 2016, I2CE 2015, IC3 2012-2015, UCC 2015, General chair (32): CCGrid 2017, IPDPS 2014, BDC 2014, HPDC 2013; Organizing committees for over 300 conferences.

Collaborators and Co-Editors: (92)

H. Abbasi (ORNL), S. Abdelwahed (MSU), M. Adams (Columbia), G. Agha (UIUC), G. Agnew, (Rutgers), A. Agrawal, (NWU), S.Ahern (ORNL), J. Ahrens (LANL), A. Aiken (Stanford), A. Akoglu (U.of Arizona), D. Bader (Georgia Tech), I. Banicescu (MSU), J. Bell (LBNL), J. Bennett (SNL), H. Berman (Rutgers), W. Bethel (LBNL), P. T. Bremer (LLNL), J. Browne (U. of TX, Austin), E. Brugger (LLNL),U.Catalyurek (OSU), C. S. Chang (PPPL), J. Chen (Sandia) H, Childs (LBNL), A. Choudhary (NWU), J. Cummings (Cal Tech), S. Ethier (PPPL), E. Feibush (PPPL), R. Figueiredo (U. of FL), L. Fong (IBM),J. Fortes (U. of FL), B. Ganapathysubramanian (ISU), B. Geveci (Kitware), M. Greenwald (MIT), P. Hanrahan (Stanford), C. Hansen (U.Utah), S. Hanson (Rutgers), S. Hariri (U. of Arizona), J. Hesthaven (Brown), J. Huang (U. Tenn. Knoxville), R. Jantz (Rutgers), C. Johnson (U. Utah),K. Jordon (IBM), K. Joy (U. C. Davis), D. Katz (U of Chicago), S. Klasky (ORNL), H. Kolla (SNL),A. Kritz (Lehigh), T. Kurc (Emory), G. Laskaris (NJEdge.net), R. Latham (ANL), W. Liao (NWU), J. Lofstead (SNL), J. Logan (ORNL), K. L. Ma (U.C. Davis), C. Maher (Rutgers),P. McCormick (LANL), L. Michelson (Rutgers),K. Moreland (Sandia), R. Moser (U. of TX, Austin), T. Nguyen (Rutgers), R. A. Oldfield (SNL), M. Papka (ANL), S. Parker (U of CO), V. Pascucci (U Utah), T. Peterka (ANL), V. Prasanna (USC), N. Podhorszki (ORNL), S. Poole (ORNL), D. Quinlan (LLNL), R. Ross (ANL), J. Saltz (Emory), N. Samatova (NCSU), K. Schwan (Georgia Tech), J. Sexton (IBM), J. Shalf (LBNL), H-W. Shen (OSU), M. Shephard (RPI), A. Shoshani (LBNL), D. Smith (Rutgers) S. Srinivasan (U. of TX, Austin), L. Sugiyama (MIT), G. Tynan (UCSD), C. Varela (RPI), V. Vishwanath (ANL), J. Weissman (UMN), M. F. Wheeler, (U of TX, Austin), M. Wolf (Georgia Tech), J. Woodring (LANL), P. Worley (ORNL), K. Wu (LBNL), H. Yu (U. of Nebraska-Lincoln)

Graduate and Postdoctoral Advisors:

Graduate Advisor: (1) Salim Hariri, University of Arizona

Postdoctoral Advisors: (2) James Browne & Richard Matzner, University of Texas at Austin

Thesis Advisor and Postgraduate-Scholar Sponsor:

Post-Doctoral: (4) H. Bui, J. Diaz, V. Potluru, I. Rodero

Ph.D. Students: (13) M. AbdelBaky, M. F. Aktas, M. Gamell, G. Haldeman, T. Jin, A. Pelaez, J.Ren, Q.Sun, S. Swami, D. Wang, A. Zamani, M. Zou

Past Ph.D. Students/Post-Docs: (15) V. Bhat (Yahoo), S. Chandra (Bloomberg), C. Docan (Google), A. Hernandez (Xerox), N. Jiang (Microsoft), H. Kim (Xerox), X. Li (Univ. of FL), Z. Li (Ayasdi), H. Liu (Microsoft), V. Matossian (Altair), C. Schmidt (Google), M. Wang (Google), F. Zhang (Two Sigma Investments LLC.), G. Zhang (FactSet), L. Zhang (Bloomberg), J. Zola (Univ. of Buffalo, NY)

Other advisees: MS Thesis – 51 (48 completed), Undergraduate – 3

SHIRLEY V. MOORE

EDUCATION

B. A.	Indiana University, Bloomington, IN	Mathematics and Chemistry	1978
M.Ed.	University of Illinois, Urbana-Champaign, IL	Science Education	1980
M.S.	Wichita State University, Wichita, KS	Mathematics	1985
Ph.D.	Purdue University, West Lafayette, IN	Computer Sciences	1990

EXPERIENCE

Oak Ridge National Laboratory	Senior Computer Scientist	July 2016 – present
University of Texas at El Paso	Affiliate Graduate Faculty	July 2016 -- present
University of Texas at El Paso	Associate Professor	July 2012 – June 2016
University of Tennessee at Knoxville	Associate Research Professor	July 2010 – Jun 2012
Lindblom Math & Science Academy Chicago, IL	High School Teacher	Aug 2007 -- June 2010
University of Tennessee at Knoxville	Associate Director of Research	June 1998 – July 2007
University of Tennessee at Knoxville	Research Associate	Jan 1993 – May 1998
Hope College, Holland, MI	Assistant Professor	Aug 1990 – Dec 1992
Wichita State University	Instructor	Aug 1981 – July 1985

SELECTED PUBLICATIONS

1. Christopher D. Carothers, Jeremy S. Meredith, Mark P. Blanco, Jeffrey S. Vetter, Misbah Mubarak, Justin M. LaPre, Shirley Moore. Durango: Scalable synthetic workload generation for extreme-scale application performance modeling and simulation. SIGSIM-PADS 2017: 97-108.
2. Mariam Umar, Shirley V. Moore, Jeremy S. Meredith, Jeffrey S. Vetter, and Kirk Cameron. Aspen-based performance and energy modeling frameworks. Journal of Parallel and Distributed Computing 96 (2017) (to appear).
3. Charles W. Lively, Valerie E. Taylor, Xingfu Wu, Hung-Ching Chang, Chun-Yi Su, Kirk W. Cameron, Shirley Moore, Daniel Terpstra. E-AMOM: an energy-aware modeling and optimization methodology for scientific applications. Computer Science - R&D 29(3-4): 197-210 (2014)
4. Charles W. Lively, Xingfu Wu, Valerie E. Taylor, Shirley Moore, Hung-Ching Chang, Chun-Yi Su, Kirk W. Cameron. Power-aware predictive models of hybrid (MPI/OpenMP) scientific applications on multicore systems. Computer Science - R&D 27(4): 245-253 (2012)
5. Charles W. Lively, Xingfu Wu, Valerie E. Taylor, Shirley Moore, Hung-Ching Chang, Kirk W. Cameron. Energy and performance characteristics of different parallel implementations of scientific applications on multicore systems. IJHPCA 25(3): 342-350 (2011)
6. Karl Fürlinger, Shirley Moore. Recording the control flow of parallel applications to determine iterative and phase-based behavior. Future Generation Comp. Syst. 26(1): 162-166 (2010)
7. Karl Fürlinger, Shirley Moore. Capturing and Analyzing the Execution Control Flow of OpenMP Applications. International Journal of Parallel Programming 37(3): 266-276 (2009)
8. Rogelio Long and Shirley Moore, Countering the noise-induced critical path problem, Twelfth Workshop on High Performance Power Aware Computing (HPPAC 2016), Chicago, IL, May 2016.
9. Rogelio Long, Shirley Moore, and Barry Rountree. Iso-power-efficiency: an approach to scaling application codes with a power budget. Eleventh Workshop on High Performance Power Aware Computing (HPPAC 2015), Hyderabad, India, May 2015.
10. Vincent Weaver, Daniel Terpstra, and Shirley Moore. “Nondeterminism and Overcount in Hardware Counter Implementations”. In Proceedings of the 2013 IEEE International Symposium on

Performance Analysis of Systems and Software (ISPASS), Austin, TX, USA, April 21-23, 2013.

SYNERGISTIC ACTIVITIES

1. Co-PI on the DOE SciDAC Performance Evaluation Research Center (PERC), Performance Engineering Research Institute (PERI), and Sustained Performance, Energy, and Resilience (SUPER) Institute, 2001-2016.
2. Co-PI on the AFOSR Hardware-Software Co-Design Approach to Advances in Software and Hardware (CoDAASH) project, 2012-2017.
3. Organizer and chair of First (2014) and Second (2015) International Workshops on Hardware-software Co-design for High Performance Computing (Co-HPC).
4. Developer of the ASPEN performance modeling tool, Oak Ridge National Laboratory, funded by the Department of Energy Exascale Computing Project Design Space Evaluation, 2017-2019.
5. Co-PI on the Department of Energy Exascale Computing Project Application Development project "Enabling GAMESS for Exascale Computing in Chemistry & Materials"

GRADUATE AND POSTDOCTORAL ADVISORS

Bharat Bhargava, Purdue University; Jack Dongarra, University of Tennessee

GRADUATE STUDENTS AND POSTDOCTORAL RESEARCHERS ADVISED

Karl Fürlinger, LMU-Munich; Madhu Hari, University of Utah; Matthew Johnson, Garmin; Kevin London, Microsoft; Rogelio Long, University of Texas at El Paso; Marco Lopez, University of Texas at El Paso; Daniel Lucio, Oak Ridge National Laboratory; Henry Moncada, University of Texas at El Paso; Snigdha Mudunuri, Texas Tech University; Umayanganie Munipala Klaassen, University of Texas at El Paso; Sonish Shrestha, Grand Central Networks; Fengguang Song, Indiana Univ.-Purdue Univ. Indianapolis; Joseph Thomas, University of Nevada-Las Vegas; Mariam Umar, Virginia Tech University; Vincent Weaver, University of Maine; Felix Wolf, University of Darmstadt.

Total Graduate Students Advised: 12. Total Postdoctoral Scholars Sponsored: 4.

COLLABORATORS AND CO-EDITORS

Mark Blanco, Rensselaer Polytechnic Institute; Kirk Cameron, Virginia Tech University; Christopher Carothers, Rensselaer Polytechnic Institute; Laura Carrington, San Diego Supercomputer Center; Eric Freudenthal, University of Texas at El Paso; Mark Gordon, Iowa State University/Ames Laboratory; Heike Jagode, University of Tennessee; Umayanganie Munipala Klaassen, University of Texas at El Paso; Justin LaPre, Rensselaer Polytechnic Institute; Rogelio Long, University of Texas at El Paso; Ivonne Lopez, University of Texas at El Paso; Misbah Mubarak, Argonne National Laboratory; Jeremy Meredith, Google; Philip Roth, Oak Ridge National Laboratory; Barry Rountree, Lawrence Livermore National Laboratory; Sarat Sreepathi, Oak Ridge National Laboratory; Valerie Taylor, Texas A&M University; Ananta Tiwari, San Diego Supercomputer Center; Daniel Terpstra, University of Tennessee (retired); Jeffrey Vetter, Oak Ridge National Laboratory; Vincent Weaver, University of Maine; Theresa Windus, Iowa State University; Xingfu Wu, Texas A&M University.

Leonid Oliker

Lawrence Berkeley National Laboratory, Computational Research Division
One Cyclotron Road, M/S 59R4104, Berkeley, CA 94720
EMail: loliker@lbl.gov

Education and Training

Univ. of Pennsylvania	Computer Engineering	Bachelor Science	magna cum laude	1991
Univ. of Pennsylvania	Computer Engineering	Bachelor Science	magna cum laude	1991
University of Colorado	Computer Science	Masters		1994
University of Colorado	Computer Science	PhD		1997

Research and Professional Experience

Lawrence Berkeley Nat'l Lab	Senior Computer Scientist	2010– present

Areas of research include:
Optimization of scientific methods on emerging multi-core systems;
Ultra-efficient designs of domain-optimized computational platforms;
Performance evaluation of extreme-scale applications on leading supercomputers.
Compiler-driven autotuning for stencil computations.

Lawrence Berkeley Nat'l Lab	Staff Computer Scientist	1999 – 2010
NASA Ames Research Center	Post-Doctoral Scientist	1997 – 1998
NASA Ames Research Center	Visiting Doctoral Researcher	1994 – 1997
University of Colorado	Research Assistant	1993 – 1994
David Sarnoff Research Center	Senior Technical Associate	1991 – 1992

Selected Publications

1. William Tang, Bei Wang, Stephane Ethier, Grzegorz Kwasniewski, Torsten Hoefler, Khaled Z. Ibrahim4, Kamesh Madduri, Samuel Williams, Leonid Oliker, Carlos Rosales-Fernandez, Tim Williams, "Extreme Scale Plasma Turbulence Simulations on Top Supercomputers Worldwide", Supercomputing, November 2016.

2. E. Georganas, A. Buluc, J. Chapman, S. Hofmeyr, C. Aluru, R. Egan, L. Oliker, D. Rokhsar, K. Yelick, "HipMer: an extreme-scale de novo genome assembler,", Supercomputing, November 2013.

3. S. Kamil, C. Chan, S. Williams, L. Oliker, J. Shalf, M. Howison, W. Bethel, and Prabhat. *" A Generalized Framework for Auto-tuning Stencil Computations* . In Cray User Group Conference, Atlanta, GA, May 2009. **Best paper award**.

4. S. Williams, J. Carter, L. Oliker, J. Shalf, K. Yelick, *"Lattice Boltzmann Simulation Optimization on Leading Multicore Platforms"*, International Parallel & Distributed Processing Symposium (IPDPS) 2008. **Best paper award**.

5. L. Oliker, A. Canning, J. Carter, C. Iancu, M. Lijewski, S. Kamil, J. Shalf, H. Shan, E. Strohmaier, S. Ethier, T. Goodale,*"Scientific Application Performance on Candidate PetaScale Platforms"*, International Parallel & Distributed Processing Symposium (IPDPS) 2007. **Best paper award**.

6. J. Carter, M, Soe, L. Oliker, Y. Tsuda, G.Vahala, L. Vahala, A. Macnab, *"Magnetohydrodynamic Turbulence Simulations on the Earth Simulator Using the lattice Boltzmann Method"*, SC05, 2005. **Gorden Bell Finalist**

7. J. Borrill, J. Carter, L. Oliker, D. Skinner, R. Biswas, *"Integrated Performance Monitoring of a Cosmology Application on Leading HEC Platforms"*,International Conference on Parallel Processing: ICPP 2005. **Nominated: Best paper award**

8. L. Oliker, A. Canning, J. Carter, J. Shalf, S. Ethier, *"Scientific Computations on Modern Parallel Vector Systems"*, SC04, 2004. **Nominated: Best paper award**

9. H. Shan, J. Singh, L. Oliker, R. Biswas, *"A Comparison of Three Programming Models for Adaptive Applications on the Origin2000"*, SC00, 2000. **Best student paper award**

10. L. Oliker, R. Biswas, *"Parallelization of a Dynamic Unstructured Application using Three Leading Paradigms"*,SC99, 1999. **Best paper award**

Research Interest and Expertise

Dr. Oliker has published over 150 peer-reviewed publications, including five best paper awards, in the areas of performance evaluation and optimization, co-design for scientific computing, sparse matrix optimization, autotuning, large-scale genomic optimization, and power-efficient architectural designs. His work has investigated a broad range of international supercomputering systems as well as detailed exploration of emerging architectural designs, including multicore, manycore, highly multithreaded, and power-efficient platforms, and has led numerous multi-institutional large-scale studies.

Synergistic Activities

Director SUPER SciDAC Institute for Computer Science Performance and Optimization

Program committee member: SC05, SC06, SC07, SC08, SC10, SC11, SC12, SC13, SC14, SC16, ICS 2008, ICS 2009, ICS 2010, ICS 2012, ICS 2013, ICPP11, ICPP 2014, E2SC 2013, E2SC 2014, E2SC 2015, E2SC 2016, Vice-Chair IPDPS11, Co-Program Chair IPDPS12,

Associated Editor JPDC 2013-present

PI: INCITE PEAC End Station, 100M hours for over 150 users 2013-present

Guest Editorship: IJHPCA 2008, ParCo Special Issue "Revolutionary Technologies for Acceleration of Emerging Petascale Applications" 2009, ParCo "Emerging Programming Paradigms for Large-Scale Scientific Computing" 2011, IJHPCA "Special Issue on Autotuning" 2013, ParCo Special Issue "Graph Analysis for Scientific Discovery", ParCo Special Issue "High-End Computing for Next-Generation Scientific Discovery".

Workshop Co-Organizer: Exascale Computing Systems Productivity Workshop (2014), PEEPS: Programming Environments for Emerging Parallel Systems, 2010. Modeling and Simulation at the Exacscale for Energy & the Environment (E3) 2007 , DOE/DOD Emerging Architectures & Applications 2007, Programming Models for HPCS Ultra-Scale Applications 2005

External Collaborators (48 months) and Co-Editors (24 months)

K. Bergman (Columbia), R. Biswas (NASA), L. Carrington (UCSD) J. Chame (USC) J. Demmel (UCBerkeley), S. Ethier (PPPL), T. Gamblin (LLBNL) J. Gilbert (UCSB), M. Hall (U Utah), J. Hollingsworth (UMD), K. Huck (U Oregon), D. Jacobsen (LANL), S. Jegelka (MIT), S. Kamil (Adobe), R. Lucas (ISI), A. Malony (U Oregon), O. McBryan (PhD Advisor), K. Madduri (Penn State), S. Moore (UTK), B. Norris (U Oregon), P.Roth (ORNL), T. Ringler (LANL), B. de Supinski (LLNL), W. Tang (PPPL), R. Vuduc (Georgia Tech), P. Worley (ORNL).

John M. Patchett

Los Alamos National Laboratory
Computer, Computational, and Statistical Sciences Division
PO Box 1663, Mail Stop B287
Los Alamos, NM 87544
patchett@lanl.gov

Education and Training

2017	Doctoral Candidate	Computer Science	TU Kaiserslautern
2011	MSc	Computer Science	University of New Mexico
2003	BA	Anthropology	University of New Mexico

Research and Professional Experience

Los Alamos National Laboratory	Staff Scientist	Oct 2005 – present
Advanced Computing Laboratory (LANL)	Contract Systems Administrator	Oct 1999 – Oct 2005
Protection Technology Los Alamos	Systems Administrator	Oct 1997 – Oct 1999
The PC Place	Assistant Sales Manager	Jun 1995 – Sep 1997

Selected Publications

- John Patchett, Boonthanome Nouanesengsy, Galen Gisler, James Ahrens, and Hans Hagen, "In situ and post processing workflows for asteroid ablation studies", 19th EG/VGTC Conference on Visualization, 2017.
- J. Ahrens, S. Jourdain, P. OLeary, J. Patchett, D. Rogers, M. Petersen, "An image-based approach to extreme scale in situ visualization and analysis", SC '14 Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis Pages 424-434
- Bent, John, Sorin Faibish, Jim Ahrens, Gary Grider, John Patchett, Percy Tzelnic, and Jon Woodring. "Jitter-free co-processing on a prototype exascale storage stack." In Mass Storage Systems and Technologies (MSST), 2012 IEEE 28th Symposium on, pp. 1-5. IEEE, 2012.
- Samsel, Francesca, John M. Patchett, David Honegger Rogers, and Karen Tsai. "Employing Color Theory to Visualize Volume-rendered Multivariate Ensembles of Asteroid Impact Simulations." In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1126-1134. ACM, 2017.
- Woodring, Jonathan, Mark Petersen, Andre Schmeier, John Patchett, James Ahrens, and Hans Hagen. "In situ eddy analysis in a high-resolution ocean climate model." IEEE transactions on visualization and computer graphics 22, no. 1 (2016): 857-866.
- Brownlee, Carson, John Patchett, Li-Ta Lo, David DeMarle, Christopher Mitchell, James Ahrens, and Charles D. Hansen. "A study of ray tracing large-scale scientific data in two widely used parallel visualization applications." (2012).
- Ahrens, James, John Patchett, Andrew Bauer, Sbastien Jourdain, David H. Rogers, Mark Petersen, Benjamin Boeckel, Patrick OLeary, Patricia Fasel, and Francesca Samsel. "In situ mpas-ocean image-based visualization." In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Visualization and Data Analytics Showcase. 2014.
- Hamilton, S., R. Burns, C. Meneveau, P. Johnson, P. Lindstrom, J. Patchett, and A. Szalay. "Extreme Event Analysis in Next Generation Simulation Architectures." In High Performance Computing: 32nd International Conference, ISC High Performance 2017, Frankfurt, Germany, June 18-22, 2017, Proceedings, vol. 10266, p. 277. Springer, 2017.
- Stompel, Aleksander, Kwan-Liu Ma, Eric B. Lum, James Ahrens, and John Patchett. "SLIC: sched-

uled linear image compositing for parallel volume rendering." In Proceedings of the 2003 IEEE Symposium on Parallel and Large-Data Visualization and Graphics, p. 6. IEEE Computer Society, 2003.

- Nouanesengsy, Boonthanome, Jonathan Woodring, John Patchett, Kary Myers, and James Ahrens. "Adr visualization: A generalized framework for ranking large-scale scientific data using analysis-driven refinement." In Large Data Analysis and Visualization (LDAV), 2014 IEEE 4th Symposium on, pp. 43-50. IEEE, 2014.

Synergistic Activities

LANL ASC CSSE Production Visualization Lead, 2017

Collaborators (last 48 months) and Co-Editors (last 24 months)

Samsel, Francesca(U Texas Austin), Hagen, Hans (TU Kaiserslautern), Gisler, Galen (LANL), Green, Jennifer (LANL), Cone, Giovanni (LANL), Hamilton, Stephen (Johns Hopkins University), Burns, Randal (Johns Hopkins University), Meneveau, Charles (Johns Hopkins University), Lindstrom, Peter (LLNL), Szalay, Alexander (Johns Hopkins University), Rogers, David (LANL), Tsai, Karen (LANL), Nouanesengsy, Boonthanome (LANL), Abram, Gregory (Texas Advanced Computing Center), Turton, Terece (U Texas Austin), Petersen, Mark (LANL), Schmeisser, Andre (Fraunhofer ITWM), Lang, Michael (LANL)

Graduate and Postdoctoral Advisors and Advisees:

Hans Hagen (TU Kaiserslautern– Thesis advisor)

ROBERT M. PATTON BIOGRAPHICAL SKETCH

Education and Training

- University of Central Florida, Computer Engineering, Ph.D., 2002.
- University of Central Florida, Computer Engineering, M.S., 2001.
- University of Central Florida, Computer Engineering, B.S., 1999.

Research and Professional Experience

- Nature Inspired Machine Learning Team Leader, Oak Ridge National Laboratory, 7/2016–present.
  At ORNL, I lead a machine learning research team with emphasis on nature inspired techniques such as evolutionary optimization, deep learning, and neuromorphic computing with broad scientific applications.
- Research Staff, Oak Ridge National Laboratory, 5/2003–present.
  At ORNL, my research focuses on statistical data analysis, evolutionary optimization, and deep learning on high performance computing platforms with broad scientific applications.

Publications

A. M. Terwilliger, G. N. Perdue, D. Isele, **R. M. Patton** and S. R. Young, "Vertex reconstruction of neutrino interactions using deep learning," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 2017.

S. R. Young, D. C. Rose, T. P. Karnowski, S-H. Lim, and **R. M. Patton**. "Optimizing deep learning hyper-parameters through an evolutionary algorithm", In Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments (MLHPC '15). ACM, New York, NY, USA. 2015.

Synergistic Activities

Patents:
Chad A. Steed, **Robert M. Patton**, Paul L. Bogen, Thomas E. Potok, Christopher T. Symons. "Interactive Visual Analytics for Situational Awareness of Social Media". Application No. 14/476,252, filed 3 Sep. 2014.

Thomas E. Potok, **Robert Patton**, Chad A. Steed. "Method and System to Discover and Recommend Interesting Documents". Application No. 13/737,652, filed 9 Jan. 2013.

Awards:
• R&D 100 Award (see patent No. 13/737,652 above) (2013)

Member: ACM, ACM SIGHPC, ACM SIGEVO, IEEE, and IEEE Computer Society.

Organizing Committee: Workshop on Machine Learning in HPC Environments at The International Conference for High Performance Computing, Networking, Storage and Analysis (2015, 2016, 2017), Neuromorphic Computing Symposium (2017).

Swann Perarnau

Education

Grenoble University/INRIA/LIG, France	Computer Science	Ph.D., 2011
Grenoble University, France	Computer Science	M.Sc., 2008
Grenoble University, France	Computer Science	B.Sc., 2006

Professional Experience

09/2016–present	Assistant Computer Scientist	Mathematics and Computer Science Division, Argonne National Laboratory
10/2014–08/2016	Postdoctoral Researcher	Mathematics and Computer Science Division, Argonne National Laboratory
03/2012–09/2014	Postdoctoral Researcher	Advanced Institute for Computational Science RIKEN, Japan
12/2011–02/2012	Expert Engineer	MOAIS Research Team INRIA Grenoble, France
09/2008–09/2011	Teaching Fellow	Ensimag/Grenoble University Grenoble, France

Select Publications

Swann Perarnau, Judicael A. Zounmevo, Matthieu Dreher, Brian C. Van Essen, Roberto Gioiosa, Kamil Iskra, Maya Gokhale, Kazutomo Yoshii, and Pete Beckman. Argo NodeOS: Toward Unified Resourcement Management for Exascale. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2017.

Swann Perarnau, Judicael A. Zounmevo, Balazs Gerofi, Kamil Iskra, Pete Beckman. Exploring Data Migration for Future Deep-Memory Many-Core Systems. In *IEEE Cluster Conference*, 2016.

Swann Perarnau and Mitsuhisa Sato. Discovering sector cache optimizations on the K computer. In *Asia-Pacific Programming Languages and Compilers Workshop (APPLC)*, 2013.

Swann Perarnau, Marc Tchiboukdjian, and Guillaume Huard. Controlling cache utilization of HPC applications. In *International Conference on Supercomputing (ICS)*, 2011.

Synergistic Activities

1. Developed cache optimization methods based on intra-process cache partitioning for Intel and SPARCVIIIfx architectures. Transferred the resulting knowledge to Fujitsu's internal compiler development division.
2. Lead integrator and key researcher on resource management for the DOE-funded Argo OS/R project.
3. Program Committee member of REPPAR 2017 and ICPP 2017.
4. Reviewer for JPDC (since 2015), JoS 2011, SC 2016, IPDPS 2012 and 2016, ROME 2017, and CCgrid 2011 and 2016.

Collaborations

Collaborators and Co-Editors: This list does not not include institutional conflicts.

Mitsuhisa Sato (RIKEN), Edgar Leon (LLNL), Balazs Gerofi (RIKEN) Yutaka Ishikawa (RIKEN), George Bosilca (UTK), Jack Dongarra (UTK), Thomas Herault (UTK), Henry Hoffmann (UChicago), Laxmikant Kale (UIUC), Sriram Krishnamoorthy (PNNL), Allen Malony (UOregon), Barry Rountree (LLNL), Martin Schulz (LLNL), Sameer Shende (UOregon), Daniel Ellsworth (LLNL), Tapasya Patki (LLNL), Leonardo

Bautista-Gomez (BSC) Ana Gainaru, (UIUC) Devesh Tiwari (ORNL) Saurabh Gupta (ORNL) Christian Engelmann (ORNL) Marc Snir (UIUC), Roberto Gioiosa (PNNL), Maya Gokhale (LLNL), Erik Saule (UNC Charlotte) Arnaud Legrand (CNRS, LIG, France), Olivier Richard (LIG, France), Eric Rutten (Gipsa, France) Bogdan Robu (Gipsa, France)

Graduate Advisors and Postdoctoral Sponsors: Guillaume Huard (LIG, France), Denis Trystram (LIG, France), Mitsuhisa Sato (RIKEN, Japan), Kamil Iskra (ANL), Pete Beckman (ANL)

Tom Peterka

Professional Preparation

- **Doctor of Philosophy** Computer Science Engineering. University of Illinois at Chicago (UIC), Chicago, IL. July, 2007. Dissertation: Dynallax: Dynamic Parallax Barrier Autostereoscopic Display.
- **Master of Science** Computer Science Engineering. University of Illinois at Chicago (UIC), Chicago, IL. December, 2003. Thesis: Scientific Visualization of N-Dimensional Attainable Regions.
- **Bachelor of Science** Computer Science Engineering. University of Illinois at Chicago (UIC), Chicago, IL. June, 1987.

Academic / Professional Appointments

- **Argonne National Laboratory** Argonne IL, 2015 - present. Computer Scientist, Mathematics and Computer Science Division.
- **Northwestern University Argonne National Laboratory Institute of Science and Engineering** Evanston IL, 2014 - present. Fellow.
- **University of Chicago Computation Institute** Chicago IL, 2009 - present. Fellow.
- **University of Illinois at Chicago** Chicago IL, 2009 - present. Adjunct Assistant Professor, Electronic Visualization Laboratory.

Related Publications

- Peterka, T., Croubois, H., Li, N., Rangel, S., Cappello, F.: Self-Adaptive Density Estimation of Particle Data. SIAM Journal on Scientific Computing SISC Special Section on CSE15: Software and Big Data, 2016.
- Peterka, T., Morozov, D., Phillips, C.: High-Performance Computation of Distributed-Memory Parallel 3D Voronoi and Delaunay Tessellation. Proceedings of SC14, New Orleans, LA, 2014.
- Peterka, T., Ross, R., Kendall, W., Gyulassy, A., Pascucci, V., Shen, H.-W.: Scalable Parallel Building Blocks for Custom Data Analysis. Proceedings of Large Data Analysis and Visualization Symposium (LDAV'11), IEEE Visualization Conference, Providence RI, 2011.
- Peterka, T., Ross, R., Nouanesengsey, B., Lee, T.-Y., Shen, H.-W., Kendall, W., Huang, J.: A Study of Parallel Particle Tracing for Steady-State and Time-Varying Flow Fields. Proceedings of IPDPS'11, Anchorage AK, 2011.
- Peterka, T., Goodell, D., Ross, R., Shen, H.-W., Thakur, R.: A Configurable Algorithm for Parallel Image-Compositing Applications. Proceedings of SC09, Portland OR, November 2009.

Other Significant Publications

- Ewa Deelman and Tom Peterka and Ilkay Altintas and Christopher Carothers and Kerstin Kleese van Dam and Kenneth Moreland and Manish Parashar and Lavanya Ramakrishnan and Michela Taufer and Jeffrey Vetter: The Future of Scientific Workflows. International Journal of High Performance Computing Applications, 2017.
- Dreher, M., Peterka, T.: Bredala: Semantic Data Redistribution for In Situ Applications. Proceedings of IEEE Cluster 2016, Taipei, Taiwan, 2016.

- Morozov, D., Peterka, T.: Efficient Delaunay Tessellation through K-D Tree Decomposition. Proceedings of SC16, Salt Lake City, UT, 2016.
- Morozov, D., Peterka, T.: Block-Parallel Data Analysis with DIY2. Proceedings of LDAV'16, Baltimore, MD, 2016.
- Guo, H., Glatz, A., Peterka, T.: In Situ Magnetic Flux Vortex Visualization in Time-Dependent Ginzburg-Landau Superconductor Simulations. Proceedings of IEEE Pacific Visualization, Seoul, South Korea, 2017.

Synergistic Activities

- **Research Collaborations:** Northwestern University Northwestern-Argonne Institute of Science and Engineering, University of Chicago Computation Institute, University of Illinois at Chicago Electronic Visualization Laboratory
- **Award Recognitions:** Best Paper, IEEE Cluster 2016, Best Poster, IEEE LDAV 2013; Best Student Paper, SC11; Best Paper, EGPGV 2008; Best Paper, IEEE VR 2007; University Fellowship, University of Illinois at Chicago, 2006
- **Education and Outreach:** SC 2012-2017 technical program committee member; IEEE Vis 2015-2016 organizing committee member, EuroPar 2016 organizing committee member; LDAV 2016 program committee member; IEEE SciVis 2015 international program committee member; IEEE Cluster 2015 data storage, analysis and visualization area co-chair; ICPP 2014 technical program committee member; EGPGV 2013-2015 international program committee member; UltraVis 2013 technical program committee member; SC 2012 ACM student research competition posters judge; CScADS Workshop on Data Analysis and Visualization 2008-2012 co-organizer
- **Reviews:** IEEE Cluster, IEEE Pacific Visualization, IEEE International Parallel and Distributed Processing Symposium, Eurovis, IS&T / SPIE, IEEE Visualization, ACM SIGGRAPH, ACM SIGGRAPH-Asia, IEEE-ACM Supercomputing, Eurographics Parallel Graphics and Visualization, International Conference on Parallel Processing, IEEE Transactions on Visualization and Computer Graphics, IEEE Computer Graphics & Applications

Collaborators

Gagan Agrawal, OSU, Ilkay Altintas, SDSC, Jim Ahrens, LANL, Wes Bethel, LBNL, Chris Carothers, RPI, Emily Casleton, LANL, Hank Childs, LBNL, Alok Choudhary, NU, Kerstin Kleese van Dam, BNL, Ewa Deelman, ISI, Berk Geveci, Kitware, Jian Huang, UTK, Scott Klasky, ORNL, Bill Kramer, NCSA, Wei-Keng Liao, NU, Jay Lofstead, SNL, Kwan-Liu Ma, UCD, Kenneth Moreland, SNL, Dmitriy Morozov, LBNL, Youssef Nashed, NU, Manish Parashar, RU, David Pugmire, ORNL, Lavanya Ramakrishnan, LBNL, Han-Wei Shen, OSU, Marc Snir, UIUC, Michela Taufer, UD, Jeff Vetter, ORNL, Patrick Widener, SNL, Joanne Wendelberger, LANL, Matthew Wolf, ORNL, Jon Woodring, LANL

Graduate and Postdoctoral Advisors

Postdoctoral: Robert Ross (ANL). Phd: Andrew Johnson, Jason Leigh, Dan Sandin, Thomas DeFanti, (UIC), Jurgen Schulze (UCSD). MS: Peter Nelson, John Bell (UIC). Advisees: None.

Thesis Advisor and Postgraduate-Scholar Sponsor

Youssef Nashed, Hanqi Guo, Matthieu Dreher

Biographical Sketch for Norbert Podhorszki

Oak Ridge National Laboratory **Phone: (865)-574-7159**
Oak Ridge, TN **Fax: (865) 241-4811**
E-mail: pnorbert@ornl.gov

Education

PhD in Information Science and Technology, Eötvös Loránd University of Budapest, Hungary, 2005.
MS in Computer Science, Eötvös Loránd University of Budapest, Hungary, 1995.

Professional Experience

2008 – present, Research Scientist, Oak Ridge National Laboratory, Oak Ridge, TN.
2006 – 2007, Post-doctorate Researcher, University of California, Davis
1995 – 2005, Research Scientist Computer and Automation Research Institute, Budapest, Hungary

Professional Expertise

Dr. Podhorszki has over 20 years of experience in large-scale scientific applications, high performance I/O, code coupling, workflow automation, Grid computing, Desktop Grid computing, performance analysis and logic programming. Dr. Podhorszki develops and maintains the ADIOS software since the first public release version 1.0.

Select Publications

1. Lion Krischer, James Smith, Wenjie Lei, Matthieu Lefebvre, Youyi Ruan, Elliott Sales de Andrade, Norbert Podhorszki, Ebru Bozdag, Jeroen Tromp. An Adaptable Seismic Data Format. *Geophysical Journal International* GJI-S-16-0264, 2016
2. S. Herbein, S. McDaniel, N. Podhorszki, J. Logan, S. Klasky, and M. Taufer. 2016. Performance characterization of irregular I/O at the extreme scale. *Parallel Comput.* 51, C (January 2016), 17-36.
3. Norbert Podhorszki, Scott Klasky, Qing Liu, Yuan Tian, Manish Parashar, Karsten Schwan, Matthew Wolf, Sriram Lakshminarasimhan, "ADIOS", In *High Performance Parallel I/O*, eds: Prabhat, Q. Koziol. 2014 Chapman & Hall/CRC Computational Science Series, ISBN 9781466582347
4. Q. Liu, J. Logan, Y. Tian, H. Abbasi, N. Podhorszki, J-Y. Choi, S. Klasky, R. Tchoua, J. Lofstead, R. Oldfield, M. Parashar, N. Samatova, K. Schwan, A. Shoshani, M. Wolf, K. Wu, W. Yu: Hello ADIOS: the challenges and lessons of developing leadership class I/O frameworks. Concurrency and Computation: Practice and Experience, 2013. doi: 10.1002/cpe.3125.
5. Tong Jin, Fan Zhang, Qian Sun, Hoang Bui, Manish Parashar, Hongfeng Yu, Scott Klasky, Norbert Podhorszki, Hasan Abbasi. "Using cross-layer adaptations for dynamic data management in large scale coupled scientific workflows", In Proceedings of SC'13. ACM, New York, NY, USA, Article 74
6. Qing Liu, Norbert Podhorszki, Jeremy Logan, Scott Klasky. "Runtime I/O Re-Routing + Throttling on HPC Storage," In USENIX HotStorage, 2013
7. B. Xie, J. Chase, D. Dillow, O. Drokin, S. Klasky, S. Oral, N. Podhorszki: "Characterizing output bottlenecks on a supercomputer". In Proceedings of SC '12. IEEE Computer Society Press, Los Alamitos, CA, USA, Article 8
8. K. Moreland, R. Oldfield, P. Marion, S. Jourdain, N. Podhorszki, C. Docan, M. Parashar, M. Hereld, M. E. Papka and S. Klasky. "Examples of In Transit Visualization", Proceedings of the Workshop on Petascale Data Analytics: Challenges and Opportunities (PDAC-11), in conjunction with ACM/IEEE SC11, Seattle, WA, USA, November 2011
9. J. Chen, Choudhary A, Supinski B, DeVries M, Hawkes E, Klasky S, Liao W, Ma K, Mellor-Crummey J, Podhorszki N, Sankaran R, Shende S and Yoo C. Terascale direct numerical simulations of turbulent combustion using S3D. *Computational Science and Discovery*, 2 015001 (31pp), Jan 2009
10. J. Cummings, A. Pankin, N. Podhorszki, G. Park, S. Ku, R. Barreto, S. Klasky, C. S. Chang, H. Strauss, L. Sugiyama, P. Snyder, D. Pearlstein, B. Ludaescher, G. Bateman, A. Kritz and the CPES Team. Plasma edge kinetic-MHD modeling in tokamaks using Kepler workflow for code coupling, data management and visualization. *Commun. Comput. Phys.,* 4 (2008), pp. 675-702.

Select Synergistic Activities

OLCF. My role in the Oak Ridge Leadership Facility is to help applications use better I/O techniques and to develop new solutions for their needs. **DOE ASCR SciDAC Scalable Data Analysis and Visualization.** My ADIOS support for applications is partially funded by this center. **DOE SciDAC for: Center for Plasma Edge Simulation.** My role was to develop scientific workflows for fusion code coupling and later to develop and use the ADIOS framework for code coupling. **DOE SciDAC Scientific Data Management Center:** I developed scientific workflow automation for high-performance computing applications. Presented Kepler scientific workflow tutorials at Supercomputing and SciDAC conferences.

Collaborators and Co-editors

Manish Parashar (Rutgers), Greg Eisenhauer (Georgia Tech), Mark Ainsworth (Brown), John K. Wu (LBNL), Nagiza Samatova (NCSU), Jeroen Tromp (Princeton), Lion Krischer (LMU Munich), Ebru Bozdag (Colorado School of Mines), Michela Taufer (U. Delaware), Jackie Chan (Sandia), Hemanth Kolla (Sandia), Hank Childs (U. Oregon), Xiaosong Ma (NCSU), Prabhat (LBNL), Quincy Koziol (LBNL), Chuck Atkins (Kitware)

Graduate and Postdoctoral Advisors and Advisees

Ph.D.:	Prof. Péter Kacsuk, Computer and Automation Research Institute, Budapest, Hungary and
	Katalin Pásztorné Varga, Ph.D., Eötvös Loránd University of Budapest, Hungary
Post-doctorate:	Bertram Ludaescher, Ph.D., University of California, Davis, CA

Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831
(865) 241-8990 pugmire@ornl.gov

David R. Pugmire, Ph. D.

Education

Undergraduate	University of Utah	Computer Science	BS	1992
Graduate	University of Utah	Computer Science	PhD	2000

Research and Professional Experience

2016-present	Joint Faculty Professor, Electrical Engineering and Computer Science, University of Tennessee
2015-present	Senior Computer Scientist, Visualization Lead, Scientific Data Group, Oak Ridge National Laboratory
2013-2015	Visualization Lead, Scientific Data Group, Oak Ridge National Laboratory
2011-2013	Visualization Lead, Scientific Computing Group, Oak Ridge National Laboratory
2007-2011	Computer Scientist, Scientific Computing Group, Oak Ridge National Laboratory
2003-2007	Computer Scientist, High Performance Computing Group, Los Alamos National Laboratory
1997-2003	Design Manager, Lead Software Engineer, IronCAD, LLC, Atlanta, GA.

Synergistic Activities

ORNL PI for Exascale Computing Projects Software Technologies Project
Program Committee for Supercomputing, IEEE Visualization, IEEE Cluster
Facilities Chair for SciDAC Institute for Scientific Data Analysis and Visualization

Selected Recent Publications

J. Kress, R. Churchill, S. Klasky, M. Kim, H.Childs, D. Pugmire, "**Preparing for In Situ Processing on Upcoming Leading-edge Supercomputers**", Supercomputing Frontiers and Innovations, Dec, 2016.

J. Choi, T. Kurc, J. Logan, N. Podhorszki, M. Wolf, D. Pugmire, E. Suchyta, E. Byun, S. Klasky, **"Stream Processing for Near-Real-Time Data Analysis"**, 2016 New York Scientific Data Summit (NYSDS), New York, NY, Dec, 2016.

J. Kress, D. Pugmire, S. Klasky, H. Childs, **"Visualization and Analysis Requirements for In Situ Processing for Large-Scale Fusion Simulation Code"**, Proceedings of the Second Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (ISAV), held in conjunction with SC16, Nov, 2016.

E. Bozdag, D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, D. Pugmire, "**Global Adjoint Tomography: First-Generation Model"**, Geophysical Journal International, 207(3):1739-1766 2016.

M. Larsen, C. Harrison, J. Kress, D. Pugmire, J. Meredith, H. Childs, "**Performance Modeling of In Situ Rendering**." At the International Conference for High Performance Computing, Networking, Storage and Analysis (SC16), Salt Lake City, UT, Nov. 2016. Best Paper Finalist.

K. Moreland, C. Sewell, W. Usher, L. Lo, J. Meredith, D. Pugmire, J. Kress, H. Schroots, K.- L. Ma, H. Childs, M. Larsen, C.-M. Chen, R. Maynard, and B. Geveci. **VTK-m: Accelerating the Visualization Toolkit for Massively Threaded Architectures**. IEEE Computer Graphics and Applications (CG&A), 36(3):48–58, May/June 2016.

D. Pugmire, J. Kress, J. Choi, S. Klasky, T. Kurc, R. Churchill, M. Wolf, G. Eisenhauer, H. Childs, K. Wu, A. Sim, J. Gu, J. Low, "**Visualization and Analysis for Near-Real-Time Decision Making in Distributed Workflows**", High Performance Data Analysis and Visualization (HPDAV), IPDPS 2016, Chicago, IL. May 2016.

D. Pugmire, J. Kress, J. Meredith, H. Childs, M. Larsen, S. Klasky, J. Choi, and N. Podhorszki. "**Visualization Plugins using VTKm for In-Transit Visualization with ADIOS"**. In Supercomputing Frontiers, Singapore, Mar. 2016.

J. Kress, D. Pugmire, H. Childs, S. Klasky, N. Podhorzski, J. Choi, "**Loosely Coupled In Situ Visualization: A Perspective on Why it's Here to Stay**", In-Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization, Supercomputing 2015, Austin TX.

Jong Choi, Yuan Tian, Gary Liu, Norbert Podhorszki, David Pugmire, Scott Klasky, Eun-Kyu Byun, Soonwook Hwang, Alex Sim, Lingfei Wu, John Wu, Mehmet Aktas, Manish Parashar, Michael Churchill, C.S. Chang, Tahsin Kurc, Xinyan Yan, Matthew Wolf,. **"ICEE: Enabling Data Stream Processing For Remote Data Analysis Over Wide Area Networks,"** In *Supercomputing Frontiers*, 2015.

M. Lefebvre, E. Bozdag, H. Calandra, J. Hill, W. Lei, D. Peter, n. Podhorszki, D. Pugmire, H. Rusmanugroho, J. Smith, J. Tromp, "**A Data Centric View of Large-Scale Seismic Imaging Workflows"**, 4th Supercomputing Workshop on Petascale Data Analytics, Nov 2013, Denver CO.

Dean Williams, Charles Doutriaux, John Patchett, Sean Williams, Galen Shipman, Ross Miller, Chad Steed, Harinarayan Krishnan, Claudio Silva, Aashish Chaudhary, Peer-Timo Bremer, David Pugmire, E. Wes Bethel, Hank Childs, Prabhat, Berk Geveci, Andrew Bauer, Alexander Pletzer, Jorge Poco, Tommy Ellqvist, Emanuele Santos, Gerald Potter, Brian Smith, Thomas Maxwell, David Kindig, and David Koop. "**Ultrascale Visualization of Climate Data"**. IEEE Computer, 46(9):68–76, September 2013.

D. Camp, H. Krishnan, D. Pugmire, C. Garth, I. Johnson, E. Bethel, K. Joy, H. Childs, "**GPU Acceleration of Particle Advection Workloads in a Parallel, Distributed Memory Setting**", Proceedings of Eurographics Symposium on Parallel Graphics and Visualization (EGPGV), May 2013, Girona, Spain.

D. Pugmire, T. Peterka, C. Garth, "**Parallel Integral Curves**". In High Performance Visualization: Enabling Extreme-Scale Scientific Insight, E. Bethel, H. Childs, C. Hansen, eds., Chapman & Hall/ CRC Press, 2012.

E. Bethel, D. Camp, H. Childs, C. Garth, M. Howison, K. Joy, D. Pugmire, "**Hybrid Parallelism**". In High Performance Visualization: Enabling Extreme-Scale Scientific Insight, E. Bethel, H. Childs, C. Hansen, eds., Chapman & Hall/ CRC Press, 2012.

H. Childs, D. Pugmire, S. Ahern, B. Whitlock, M. Howison, Prabhat, G. Weber, E. Bethel, "**Visualization at Extreme-Scale Concurrency**". In High Performance Visualization: Enabling Extreme-Scale Scientific Insight, E. Bethel, H. Childs, C. Hansen, eds., Chapman & Hall/ CRC Press, 2012.

H. Childs, et. al., "**VisIt: An End-User Tool Visualizing and Analyzing Very Large Data**". In High Performance Visualization: Enabling Extreme-Scale Scientific Insight, E. Bethel, H. Childs, C. Hansen, eds., Chapman & Hall/ CRC Press, 2012.

D. Camp, H. Childs, C. Garth, K. Joy, D. Pugmire, "**Parallel Stream Surface Computation for Large Data Sets**", In Proceedings of the IEEE Symposium on Large-Scale Data Analysis and Visualization (LDAV) 2012.

J.S. Meredith, S. Ahern, D. Pugmire, R. Sisneros, "**EAVL: The Extreme-scale Analysis and Visualization Library"**, Eurographics Symposium on Parallel Graphics and Visualization (EGPGV) in association with Eurographics, 2012.

J.S. Meredith, R. Sisneros, D. Pugmire, S. Ahern, "**A Distributed Data-Parallel Framework for Analysis and Visualization Algorithm Development**" , Fifth Workshop on General Purpose Processing on Graphics Processing Units (GPGPU5), 2012.

Robert B. Ross
9700 S. Cass Ave.
Building 240
Argonne, IL 60439-4844

email: robertbross@acm.org
email: rross@mcs.anl.gov

I am a computer scientist and engineer fascinated with the design, implementation, and deployment of resilient and high performance distributed systems. My focus area for nearly two decades has been data and communication system software for high performance computing. Over that time I have participated in and lead the development of a number of algorithms and open source software packages for scientific computing.

Professional Preparation

Clemson University, Clemson, SC	Computer Engineering, Focus: Computer Architecture	Ph.D., 2000
Clemson University, Clemson, SC	Computer Engineering, Minor in Math Science	B.S., 1994

Appointments

2014–present	Senior Computer Scientist	Mathematics and Computer Science Division, Argonne National Laboratory
2012–present	Senior Fellow	Northwestern–Argonne Institute for Science and Engineering, Northwestern University and Argonne National Laboratory
2011–present	Senior Fellow	Computation Institute, The University of Chicago and Argonne National Laboratory
2004–present	Adjunct Assistant Professor	Department of Electrical and Computer Engineering, Clemson University
2016–2017	Interim Division Director	Mathematics and Computer Science Division, Argonne National Laboratory
2004–2014	Computer Scientist	Mathematics and Computer Science Division, Argonne National Laboratory
2004–2011	Fellow	Computation Institute, The University of Chicago and Argonne National Laboratory

Related Publications

[1] H. Luu, M. Winslett, W. Gropp, R. Ross, P. Carns, K. Harms, M. Prabhat, S. Byna, and Y. Yao. A multiplatform study of I/O behavior on petascale supercomputers. In *Proceedings of the 24th International Symposium on High-Performance Parallel and Distributed Computing*, pages 33–44. ACM, June 2015.

[2] S. Snyder, P. Carns, J. Jenkins, K. Harms, R. Ross, M. Mubarak, and C. Carothers. A case for epidemic fault detection and group membership in HPC storage systems. In *Proceedings of the 5th International Workshop on Performance Modeling, Benchmarking, and Simulation of High Performance Computer Systems (PMBS14)*, November 2014.

[3] M. Dorier, G. Antoniu, R. Ross, D. Kimpe, and S. Ibrahim. CALCioM: Mitigating I/O interference in HPC systems through cross-application coordination. In *Proceedings of the International Parallel and Distributed Processing Symposium*, May 2014.

[4] J. Soumagne, D. Kimpe, J. Zounmevo, M. Chaarawi, Q. Koziol, A. Afsahi, and R. Ross. Mercury: Enabling remote procedure call for high-performance computing. In *Proceedings of the IEEE Cluster Conference*, September 2013.

[5] M. Mubarak, C. D. Carothers, R. B. Ross, and P. Carns. Modeling a million-node dragonfly network using massively parallel discrete event simulation. In *Proceedings of the 3rd International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS12) held as part of SC12*, November 2012.

[6] N. Liu, J. Cope, P. Carns, C. Carothers, R. Ross, G. Grider, A. Crume, and C. Maltzahn. On the role of burst buffers in leadership-class storage systems. In *Proceedings of the 2012 IEEE Conference on Massive Data Storage*, Pacific Grove, CA, April 2012.

[7] S. W. Son, S. Lang, P. Carns, R. Ross, R. Thakur, B. Ozisikylimaz, P. Kumar, W.-K. Liao, and A. Choudhary. Enabling active storage on parallel I/O software stacks. In *Proceedings of the IEEE Symposium on Mass Storage Systems and Technologies*, May 2010.

[8] N. Ali, P. Carns, K. Iskra, D. Kimpe, S. Lang, R. Latham, R. Ross, L. Ward, and P. Sadayappan. Scalable I/O forwarding framework for high-performance computing systems. In *IEEE International Conference on Cluster Computing (Cluster 2009)*, New Orleans, LA, September 2009.

[9] D. Kimpe, R. Ross, S. Vandewalle, and S. Poedts. Transparent log-based data storage in MPI-IO applications. In *Proc. of the 14th European PVM/MPI Users' Group Meeting (Euro PVM/MPI 2007)*, September 2007.

[10] P. H. Carns, W. B. Ligon III, R. B. Ross, and R. Thakur. PVFS: A parallel file system for Linux clusters. In *Proceedings of the 4th Annual Linux Showcase and Conference*, pages 317–327, Atlanta, GA, October 2000. USENIX Association.

Synergistic Activities

DOE Exascale Computing Project Software Technology Area

Level 3 Lead for Data Management and Workflows, providing guidance and coordinating research activities in this area.

SciDAC Scalable Data Management, Analysis, and Visualization Institute (SDAV)

Deputy Director, working with computational scientists and other experts in scientific data management and analysis to support data management solutions for DOE computational science applications. Project ended in 2017.

Argonne Math and Computer Science Division Data Strategy

Strategic Area Lead, working with computer scientists and applied mathematicians to develop division strategy for Big Data and data-driven science research and development in support of the DOE mission.

Darshan I/O Characterization Tool

Lightweight tool for observing the I/O behavior of HPC applications at the largest scales that has been used in production to capture information on hundreds of thousands of HPC jobs.

CODES Simulation Framework

Parallel discrete event simulation framework, building on RPI's ROSS system, that enables high fidelity simulation of complex distributed storage systems.

Philip C. Roth

Education and Training:

University of Iowa, Iowa City, Iowa	B.S.	1992	Comp. Sci. and Math
University of Illinois, Urbana-Champaign, Illinois	M.S.	1996	Comp. Sci.
University of Wisconsin, Madison, Wisconsin	Ph.D.	2005	Comp. Sci.

Research and Professional Experience:

2004 – present	Research and Development Staff Member (2007-present), Associate (2004-2007), Future Technologies Group, Oak Ridge National Laboratory, Oak Ridge, TN. Researching techniques for performance analysis/optimization and application characterization with emphasis on large-scale parallel computation.
1998 – 2004	Research Assistant, Paradyn Project, University of Wisconsin–Madison, Madison, WI. Part of team that developed and evaluated MRNet scalable tool infrastructure, and developed and evaluated MRNet-based scalable performance diagnosis and performance diagnosis results visualization techniques.
1994 – 1998	Software Developer, MCSB Technology (formerly known as CHEN Systems Corp.), Eau Claire, WI, Developed UnixWare and Windows system administration tools for award-winning CHEN 1000 server system.
1992 – 1994	Research Assistant, Pablo Project, University of Illinois at Urbana-Champaign, Urbana, IL. Developed method for reducing event trace data using statistical data clustering, and developed new components for Pablo performance tool.
1992	Teaching Assistant, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL. Taught discussion sessions on introduction to computing for behavioral and social sciences.
1990 – 1992	Undergraduate Research Assistant, ECE Parallel Processing Laboratory, University of Iowa, Iowa City, IA. Part of team that developed parallel software performance prediction tool using timed Petri nets.
1988 – 1990	Undergraduate Research Assistant, Physics Department, University of Iowa, Iowa City, IA. Participated in building of particle detector and in data analysis for high energy particle physics experiment at Fermi National Accelerator Laboratory.

Selected Publications:

W. Xie, Y. Chen, and P.C. Roth. "Parallel-DFTL: a Flash Translation Layer That Exploits Internal Parallelism in Solid State Drives." In: *Proceedings of the 11th IEEE International Conference on Networking, Architecture, and Storage (NAS'16)*. Long Beach, California, USA, Aug. 2016. *Best Paper Award nominee.*

P.C. Roth, J.S. Meredith, and J.S. Vetter. "Automated Characterization of Parallel Application Communication Patterns." In: *Proceedings of the 24th International Symposium on High-Performance Parallel and Distributed Computing (HPDC '15)*. Portland, Oregon, USA, 2015, pp. 73–84.

P. C. Roth and J. S. Meredith. "Value Influence Analysis for Message Passing Applications." In: *Proceedings of the 28th ACM International Conference on Supercomputing (ICS '14)*. Munich, Germany, 2014, pp. 145–154.

P. C. Roth. "Tracking a Value's Influence on Later Computation." In: *Proceedings of the 6th Workshop on Productivity and Performance (PROPER 2013)*. Aachen, Germany, Aug. 2013.

K. Spafford, J. S. Meredith, S. Lee, D. Li, P. C. Roth, and J. S. Vetter. "The Tradeoffs of Fused Memory Hierarchies in Heterogeneous Architectures." In: *Proceedings of the 2012 ACM Computing Frontiers (CF)*. Cagliari, Italy, 2012.

J. S. Meredith, P. C. Roth, K. L. Spafford, and J. S. Vetter. "Performance Implications of Non-Uniform Device Topologies in Scalable Heterogeneous Architectures." IEEE Micro **31**, 5 (2011), pp. 66–75.

X. Wu, K. Vijayakumar, F. Mueller, X. Ma, and P. C. Roth. "Probabilistic Communication and I/O Tracing with Deterministic Replay at Scale." In: *Proceedings of the 2011 International Conference on Parallel Processing (ICPP '11)*. Taipei, Taiwan, 2011, pp. 196–205.

P. C. Roth and B. P. Miller. "On-line Automated Performance Diagnosis on Thousands of Processes." In: *Proceedings of the eleventh ACM SIGPLAN symposium on Principles and practice of parallel programming (PPoPP '06)*. New York, New York, USA, 2006, pp. 69–80.

P.C. Roth, D.C. Arnold, and B.P. Miller. "MRNet: ASoftware-Based Multicast/Reduction Network for Scalable Tools." In: *Proceedings of the 2003 International Conference for High Performance computing, Networking, Storage and Analysis*. Phoenix, Arizona, USA, 2003, pp. 21–36. *Best Student Paper award nominee.*

P. C. Roth and B. P. Miller. "Deep Start: A Hybrid Strategy for Automated Performance Problem Searches." In: *Proceedings of the 8th International Euro-Par Conference on Parallel Processing. (Euro-Par '02)*. Paderborn, Germany, 2002, pp. 86–96.

Synergistic Activities:

- General Chair, Intl. Workshop on Data-Intensive Scalable Computing Systems (DISCS), 2015; Program Committee Co-Chair, DISCS, 2013 and 2014.
- Tutorials Chair, SC16; Birds-of-a-Feather Chair, SC13; Awards Vice Chair SC12; Posters Co-Chair, SC11.
- Guest editor, Parallel Computing: Systems & Applications, 2014 and 2017; Subject Area Editor, International Journal of High Performance Computing Applications, 2015.
- Reviewer, SC Technical Papers, 2006, 2007, 2010, 2013; SC Technical Posters 2005, 2012, 2013, 2014; Intl. Conf. on Supercomputing, 2011, 2013, 2014; Intl. Conf. on Parallel Processing 2007, 2010; Intl. Conf. on Cluster Computing 2010, 2015, 2016; Intl. Parallel and Distributed Processing Symposium 2011, 2012; Symposium on High-Performance Parallel and Distributed Computing 2011.

Identification of Potential Conflicts of Interest or Bias in Selection of Reviewers
Collaborators and Co-editors:
Jim Ahrens, Los Alamos National Laboratory; Nikolas Antolin, Ohio State U.; Dorian Arnold, Emory U.; Sarala Arunagiri, SMART High Reliability Solutions; Sophie Blondel, Unversity of Tennessee-Knoxville; Ronald Brightwell, Sandia National Laboratories; Yong Chen, Texas Tech U.; Junjie Chen, Texas Tech U.; Bronis de Supinski, Lawrence Livermore National Laboratory; Jack Dongarra, Unversity of Tennessee-Knoxville; Richard Glassbrook, Capital Advanced Technology Services; Karl Hammond, U. of Missouri; Jeffrey Hollingsworth, U. of Maryland; Mitchel Horton, Unversity of Tennessee-Knoxville; Paul Hovland, Argonne National Laboratory; Crystal Jernigan, Leidos; Jungho Lee, Argonne National Laboratory; Dong Li, U. of California, Merced; Gerald Lofstead, Sandia National Laboratories; Robert Lucas, Retired; Allen Malony, U. of Oregon; Gabriel Marin, Google; Jeremy Meredith, Google; Barton Miller, U. of Wisconsin; Leonid Oliker, Lawrence Berkeley National Laboratory; Daniel Reed, U. of Iowa; David Riegner, Ohio State U.; Kenny Roche, Pacific Northwest National Laboratory; Evan Samanas, U. of Wisconsin; Hongzhang Shan, Lawrence Berkeley National Laboratory; Barry Smith, Argonne National Laboratory; Kyle Spafford, Accelerate Diagnostics; Xianzhu Tang, Los Alamos National Laboratory; Patricia Teller, U. of Texas-El Paso; Daniel Terpstra, Retired Richard Vuduc, Georgia Tech, Samuel Williams, Lawrence Berkeley National Laboratory; Stephan Wild, Argonne National Laboratory; Wolfgang Windl, Ohio State U.; Brian Wirth, Unversity of Tennessee-Knoxville; Patrick Worley, Retired; Danqing Wu, Argonne National Laboratory; Wei Xie, Texas Tech U.; Donghua Xu, Oregon State U.; Jeffrey Young, Georgia Tech; Timothy Younkin, Unversity of Tennessee-Knoxville;

Graduate Advisors: Barton P. Miller (U. of Wisconsin-Madison), Daniel A. Reed (U. of Iowa)
Postdoctoral Advisee: Yong Chen (Texas Tech U.)

Oliver Rübel
Lawrence Berkeley National Laboratory (LBNL)
1 Cyclotron Road
Mail Stop 50F-1650
Berkeley, CA 94720, U.S.A

E-mail: oruebel@lbl.gov
Telephone: +1-510-486-4064
URL: http://vis.lbl.gov/~oruebel/
.

Education and Training:

- **Ph.D.**, Computer Science, Department of Computer Science, University of Kaiserslautern, Germany, November 2009. Advisors: Dr. H. Hagen (University of Kaiserslautern, Germany), Dr. B. Hamann (University of California, Davis), and Dr. G. H. Weber (LBNL).
- **M.S.**, Computer Science, Department of Computer Science, University of Kaiserslautern, Germany, January 2006. Advisors: Dr. H. Hagen (University of Kaiserslautern, Germany), Dr. B. Hamann and Dr. G. H. Weber (University of California, Davis).
- **B.S.**, Computer Science, Department of Computer Science, University of Kaiserslautern, Germany, July 2002.

Research and Professional Experience:

- Computer Research Scientist (Career) at the Visualization Group of the Computational Research Division at Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA, USA March 2011 – present.
- Member of the Data and Analytics Services Group of the National Energy Research Scientific Computing Center (NERSC), LBNL. March 2011 – Nov. 2016
- Postdoctoral researcher at the Data Analysis Group, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA, USA. March 2010 – March 2011.
- Student assistant at the Visualization Group, Computational Research Division, LBNL, Berkeley, CA, USA. July 2006 – March 2010.
- Collegiate of the International Research Training Group "Visualization of Large and Unstructured Data Sets Applications in Geospatial Planning, Modeling, and Engineering," University of Kaiserslautern, Germany. November 2005 – November 2009 (Alumni since November 2009).
- Visiting scholar at the Institute for Data Analysis and Visualization (IDAV), Department of Computer Science, University of California, Davis, CA, USA. March 2005 – November 2009.

Selected Publications:

- **O. Rübel**, M. Dougherty, Prabhat, P. Denes, D. Conant, E. F. Chang, K. Bouchard, "Methods for Specifying Scientific Data Standards and Modeling Relationships with Applications to Neuroscience," *Frontiers in Neuroinformatics*, vol. 10, Oct. 2016, doi:10.3389/fninf.2016.00048
- **O. Rübel**, B. Loring, J.-L. Vay, D. P. Grote, R. Lehe, S. Bulanov, H. Vincenti, and E. W. Bethel, "WarpIV: In Situ Visualization and Analysis of Ion Accelerator Simulations," *IEEE Computer Graphics & Applications, Scientific Visualization*, pp. 22-35, May/June, 2016. LBNL-1005718. (Also invited for presentation at IEEE VisWeek 2016)
- T. A. O'Brien, W. D. Collins, K. Kashinath, **O. Rübel**, S. Byna, J. Gu, H. Krishnan, and P. A. Ullrich, "Resolution dependence of precipitation statistical fidelity in hindcast simulations," *Journal of Advances in Modeling Earth Systems*, May, 2016. doi:10.1002/2016MS000671
- **O. Rübel**, C. G. R. Geddes, M. Chen, E. Cormier-Michel and E. W. Bethel, "Feature-based Analysis of Plasma-based Particle Acceleration Data," *IEEE Transactions on Visualization and Computer Graphics*, 20(2):196210, February 2014., LBNL-6333E.
- **O. Rübel**, A. Greiner, S. Cholia, K. Louie, E. W. Bethel, T. R. Northen, and B. P. Bowen, "OpenMSI:

A high-performance web-based platform for mass spectrometry imaging," *ACS Analytical Chemistry,* October, 2013. LBNL-6477E. DOI: 10.1021/ac402540a.

- S. Byna, J. Chou, **O. Rübel**, Prabhat, H. Karimabadi, W. S. Daughton, V. Roytershteyn, E. W. Bethel, M. Howison, K.-J. Hsu, K.-W. Lin, A. Shoshani, A. Uselton, and K. Wu, "Parallel I/O, Analysis, and Visualization of a Trillion Particle Simulation," *SuperComputing 2012 (SC12)*, Salt Lake City, Utah, Nov. 10-16, 2012. LBNL-5832E.
- **O. Rübel**, E. W. Bethel, Prabhat, and K. Wu, "Query-driven Visualization and Analysis," in E. W. Bethel, H. Childs and C. Hansen Eds., *High Performance Visualization*, Series: Chapman & Hall/CRC Computational Science, CRC Press, November 2012, pp.117-144.
- **O. Rübel**, G. H. Weber, M.-Y. Huang, E. W. Bethel, M. D. Biggin, C. C. Fowlkes, C. L. Luengo Hendriks, S. V. E. Keränen, M. B. Eisen, D. W. Knowles, J. Malik, H. Hagen, and B. Hamann, "Integrating data clustering and visualization for the analysis of 3d gene expression data," *IEEE Transactions on Computational Biology and Bioinformatics*, Vol.7, Num.1, pp 64–79, 2010, LBNL-382E.

Synergisitc Activities

- Co-founder and compute lead of the OpenMSI project and project lead for WarpIV, BASTet, NWB-4-HPC, and 2 SDAV-EFCR collaborations.
- Lead personal and contributor to major DOE programs including SciDAC - Scalable Data Management, Analysis, and Visualization (SDAV) and ECP - Algorithms and Infrastructure for In Situ Visualization and Analysis (ALPINE) and others.
- Reviewer for top journals and conference, e.g., EuroVis, IEEE VisWeek, IEEE TVCG, IEEE CG&A and others.

Selected Awards

- LBNL Lab Director Award for Early Career Achievement, November, 2016.
- R&D 100 Award for the development of OpenMSI, selected as "One of the 100 Most Technologically Significant New Products of the Year in Software/Services." November 13, 2015.
- In the timeframe of January, 2015 – present, Dr Rübel has received 4 LBNL SPOT awards.
- Sparkassen foundation award for outstanding masters thesis, Kaiserslautern, Germany, 2007

Collaborators and Co-editors (Non-LBNL only)

J. Ahrens (LANL), P.-T. Bremer (LLNL), E. Bruger (LLNL), M. Chen (Shanghai Jiao Tong University, PRC.), J. Chhugani (HiPerform Inc.), H. Childs (UO), J. Chou (National Tsing Hua University, Taiwan), W.S. Daughton (LANL), D. Doralyn (WSU), H. Hagen (University of Kaiserslautern, Germany), B. Hamann (UC Davis), W. Harvey William (OSU), H. Karimabadi (UC San Diego), V. Krishnamurthy (Cray Inc.), J. Kim (KAIST, Korea), D. Laney (LLNL), N. Lewis (WSU), M. Mahoney (UCB), J. Meredith (Google), V. Pascucci, (UU), D. Pugmire (ORNL), M.F. Ringenburg (Cray Inc.), V. Roytershteyn (UC San Diego), A. R. Sanderson (UU), M. Singh (Pandora), B. Smit (UCB), Y. Wang (OSU), R.T. Whitaker (UU), B. Whitlock (IntelligentLight), Y. Yao (SalesForce)

Han-Wei Shen

Professional Preparation

National Taiwan University, Taiwan	Computer Science	BS, 1988
University of New York, Stony Brook	Computer Science State	MS, 1992
University of Utah	Computer Science	PhD, 1998

Appointments

Professor, Computer Science and Engineering, The Ohio State University 2012-present
Associate Professor, Computer Science and Engineering, The Ohio State University 2005-2012
Assistant Professor, Computer Science and Engineering, The Ohio State University 1999-2005
Visiting Scientist, Argonne National Laboratory 10/08-6/09
Research Scientist, NASA Ames Research Center 9/96–8/99

Five Related Products

1. DUTTA, S., CHEN, C. M., HEINLEIN, G., SHEN, H. W., AND CHEN, J. P. In situ distribution guided analysis and visualization of transonic jet engine simulations. IEEE Transactions on Visualization and Computer Graphics 22, (1) (2017).

2. WANG, J., LIU, X., SHEN, H.-W., AND LIN, G. Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. IEEE Transactions on Visualization and Computer Graphics 22, (1) (2017).

3 .BISWAS, A., LIN, G., LIU, X., AND SHEN, H. W. Visualization of time-varying weather ensembles across multiple resolutions. IEEE Transactions on Visualization and Computer Graphics 22, (1) (2017).

4. DUTTA, S., AND SHEN, H.-W. Distribution driven extraction and tracking of features for time-varying data analysis. IEEE Tranactions on Visualization and Computer Graphics 22, (1) (2016).

5. CHEN, C. M., DUTTA, S., LIU, X., HEINLEIN, G., SHEN, H. W., AND CHEN, J. P. Visualization and analysis of rotating stall for transonic jet engine simulation. IEEE Transactions on Visualization and Computer Graphics 22, 1 (2016), 847–856.

Five Other Products

1. Lu, K., Shen, H.-W. A compact multivariate histogram representation for query-driven visualization. In 5th IEEE Symposium on Large Data Analysis and Visualization, LDAV 2015, Chicago, IL, USA, October 25-26, 2015 (2015)

2. Liu, X. and Shen, H-W. Association Analysis for Visual Exploration of Multivariate Scientific Data Sets. IEEE Transactions on Visualization and Computer Graphics 22(1): 955-964 (2016)

3. He, W., Chen, C., Liu, X., Shen, H-W: A Bayesian approach for probabilistic streamline computation in uncertain flows. IEEE PacificVis 2016: 214-218

4. Chen, C., Biswas, A., Shen, H-W: Uncertainty modeling and error reduction for pathline computation in time-varying flow fields. IEEE PacificVis 2015: 215-222

5. Biswas, A., Dutta, S., Shen, H-W, Woodring, J., An Information-Aware Framework for Exploring Multivariate Data Sets, IEEE Transactions on Visualization and Computer Graphics, Vol. 19, No. 2, pp. 2683-2690, Dec. 2013(special issue of IEEE Visualization 2013).

Synergistic Activities

Committee Members:

- IEEE Visualization Executive Committee, 2016 - Present
- IEEE SciVis Steering Committee, 2015 – Present
- IEEE Pacific Visualization Program Committee
- EuroVis Program Committee
- IEEE SciVis Program Committee
- Pacific Graphics Program Committee
- IEEE Clusters Program Committee
- ACM Interactive 3D Graphics Program Committee

Editors:

- Associate Editor, IEEE Transactions on Visualization and Computer Graphics 2007-2011
- Associate Editor, Journal of Visualization 2010- Present
- IEEE Computer Graphics and Applications Guest Editors

Co-Chairs:

- IEEE Visualization SciVis Paper co-Chair, 2013, 2014
- IEEE Pacific Visualization Paper co-Chair 2009, 2010
- Program co-Chair, IEEE Visualization 2008-2009
- Eurographics Symposium on Parallel Graphics and Visualization (EGPGV) 1999, 2004

,

STEPHEN F. SIEGEL

Department of Computer & Information Sciences
101 Smith Hall
University of Delaware
Newark, DE 19716

email: siegel@udel.edu
web: http://vsl.cis.udel.edu/siegel.html
phone: (302) 831–0083
fax: (302) 831–8458

Education and Training

University of Chicago, Mathematics, B.A., 1988

Oxford University, Mathematics, M.Sc., 1989

University of Chicago, Mathematics, Ph.D., 1993

NSF Postdoctoral Fellow, Dept. of Mathematics, Northwestern University, 1993–1996

Research and Professional Experience

Dept. of Computer & Information Sciences and Dept. of Mathematical Sciences, Univ. Delaware: Assistant Professor (2006–2012), Associate Professor (2012–present)

Univ. Massachusetts Amherst, Dept. of Computer Science: Senior Research Scientist (2001–2006), Senior Software Engineer (1998–2001)

Univ. Massachusetts Amherst, Dept. of Mathematics & Statistics: Visiting Assistant Professor (1996–1998)

Publications

1. Manchun Zheng, John G. Edenhofner, Ziqing Luo, Mitchell J. Gerrard, Michael S. Rogers, Matthew B. Dwyer, and Stephen F. Siegel. CIVL: Applying a general concurrency verification framework to C/Pthreads programs (competition contribution). In *Tools and Algorithms for the Construction and Analysis of Systems – 22nd International Conference, TACAS 2016, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2016, Eindhoven, The Netherlands, April 2–8, 2016. Proceedings*, volume 9636 of *Lecture Notes in Computer Science*, pages 908–911. Springer, 2016.

2. Adam Hammouda, Andrew R. Siegel, and Stephen F. Siegel. Noise-tolerant explicit stencil computations for nonuniform process execution rates. *ACM Trans. Parallel Comput.*, 2(1):7:1–7:33, April 2015.

3. Stephen F. Siegel, Manchun Zheng, Ziqing Luo, Timothy K. Zirkel, Andre V. Marianiello, John G. Edenhofner, Matthew B. Dwyer, and Michael S. Rogers. CIVL: The Concurrency Intermediate Verification Language. In *SC15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, New York, NY, USA, Nov 2015. ACM. Article no. 61, pages 1–12.

4. Stephen F. Siegel and Timothy K. Zirkel. Loop invariant symbolic execution for parallel programs. In V. Kuncak and A. Rybalchenko, editors, *Verification, Model Checking, and Abstract Interpretation: 13th International Conference, VMCAI 2012*, *LNCS* **7148**, 412–427. Springer, 2012.

5. Stephen F. Siegel and Timothy K. Zirkel. TASS: The Toolkit for Accurate Scientific Software. *Mathematics in Computer Science*, 5(4):395–426, 2011.

6. Stephen F. Siegel and Timothy K. Zirkel. FEVS: A Functional Equivalence Verification Suite for high performance scientific computing. *Mathematics in Computer Science*, 5(4):427–435, 2011.

7. Stephen F. Siegel and Louis F. Rossi. Analyzing BlobFlow: A case study using model checking to verify parallel scientific software. In A. Lastovetsky, T. Kechadi, and J. Dongarra, eds., *Recent Advances in Parallel Virtual Machine and Message Passing Interface, 15th European PVM/MPI User's Group Meeting*, *LNCS* **5205**, 274–282. Springer, 2008.

8. Stephen F. Siegel, Anastasia Mironova, George S. Avrunin, and Lori A. Clarke. Combining symbolic execution with model checking to verify parallel numerical programs. *ACM Transactions on Software Engineering and Methodology*, 17(2):Article 10, 1–34, 2008.

9. Stephen F. Siegel. Model checking nonblocking MPI programs. In B. Cook and A. Podelski, eds., *Verification, Model Checking, and Abstract Interpretation: 8th International Conference, VMCAI 2007*, *LNCS* **4349**, 44–58. Springer, 2007.

10. Stephen F. Siegel and George S. Avrunin. Modeling wildcard-free MPI programs for verification. In *Proceedings of the ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, 95–106, Chicago, IL, June 2005.

Synergistic Activities

Principal Investigator, National Science Foundation Award NSF CCF-1319571, *SHF: Small: Contracts for Message-Passing Parallel Programs*, September 1, 2013 – August 31, 2018

Principal Investigator, Department of Energy Award DE-SC0012566, *Program Verification for Extreme-Scale Applications*, September 1, 2014 – August 31, 2017

Lead investigator of CIVL project on verification of concurrent programs: `http://vsl.cis.udel.edu/civl`

Lead developer, ABC compiler front-end: `http://vsl.cis.udel.edu/abc`

Lead developer, SARL: Symbolic Execution and Reasoning Library, `http://vsl.cis.udel.edu/sarl`

Other Professional Activities

Associate Editor, *IEEE Transactions on Software Engineering*, 2015–present

Program Committees: EuroPVM/MPI 2009–12; Tutorial Committee, International Conference for High Performance Computing, Networking, Storage, and Analysis ("Supercomputing", SC09); Technical Paper Committee, SC10, SC16; NASA Formal Methods Symposium, 2013; Exploiting Concurrency Efficiently and Correctly ($(EC)^2$), 2008–2012, a workshop at the Intl. Conference on Computer Aided Verification

Award: *Most Distinguished Tool Feature: Checking Functional Equivalence (CIVL)*, VerifyThis 2017

Invited Talks

- Correctness 2017 (First International Workshop on Software Correctness for HPC Applications, held in conjunction with SC17), *TBA*
- RERS 2017 (Rigorous Examination of Reactive Systems, co-located with ISSTA and SPIN), *Data race detection with CIVL*
- VMCAI 2011 (International Conference on Verification, Model Checking, and Abstract Interpretation), *Formal Analysis of Message Passing*
- EuroPVM/MPI 2009, *Formal Verification for Scientific Computing: Trends and Progress*

Co-developer, Memory-Aware Data Redistribution Engine (MADRE), an MPI-based library for memory-efficient parallel data redistribution: http://vsl.cis.udel.edu/madre

Collaborator with SHARP group at Argonne in the development of a state-of-the-art neutronics transport code (see Stephen F. Siegel, Andrew R. Siegel, and Cristian Rabiti, UNIC Code: Algorithmic specification of the Method of Long Characteristics, ANL/MCS-TM-301, 2008)

Co-developer of INCA: a toolset using integer programming techniques to verify concurrent systems: http://laser.cs.umass.edu/tools/inca.shtml

Conflicts of Interest

Collaborators and co-editors: Matthew Dwyer (U. Nebraska), John Edenhofner (U. Delaware), Maximilian Fecke (LLNL), Kyle Felker (Princeton U.), Mitchell Gerrard (U. Nebraska), Ganesh Gopalakrishnan (U. Utah), Adam Hammouda (Argonne), Paul Hovland (Argonne), Falk Howar (Clausthal U. Tech.), Marc Jasper (LLNL), Ziqing Luo (U. Delaware), Andre Marianiello (U. Delaware), Jeroen Meijer (U. Twente), Krishna Narayanan (Argonne), Dan Quinlan (LLNL), Zvonimir Rakamaric (U. Utah), Michael Rogers (U. Nebraska), Markus Schordan (LLNL), Andrew Siegel (Argonne), Bernhard Steffen (TU Dortmund), Rajeev Thakur (Argonne), Stefan Wild (Argonne), Manchun Zheng (Pure Storage), Timothy Zirkel (MITRE), Jaco van de Pol (U. Twente)

Graduate and Postdoctoral Advisors and Advisees: Jonathan Alperin (U. Chicago), Leonard Evens (Northwestern U.), Manchun Zheng (Pure Storage), Pujan Kafle, Timothy McClory, Yi Wei, Jiaqiang Xu, Timothy Zirkel (MITRE), Wenhao Wu (U. Delaware), Ziqing Luo (U. Delaware), Si Li, Yihao Yan (U. Delaware)

Alex Sim <ASim@LBL.GOV>, (510) 495-2290, http://www.lbl.gov/~asim
Scientific Data Management Group, Computational Research Division, Lawrence Berkeley National Laboratory

Education and Training

- M.S. in Computer Sciences, San Francisco State University, August 1995.
- M.A. in Statistics, University of California at Berkeley, May 1993.
- B.A. (High Distinction in General Scholarship - Magna Cum Laude, Departmental Honors in Statistics) in Applied Mathematics and Statistics (Double Major), University of California at Berkeley, May 1990.

Research and Professional Experience

- 07/2013–present, Senior Computing Engineer, Lawrence Berkeley National Laboratory, Berkeley, CA.
- 10/2002–06/2013, Computer Science Engineer IV, Lawrence Berkeley National Laboratory, Berkeley, CA.
- 04/2000–10/2002, Computer Science Engineer III, Lawrence Berkeley National Laboratory, Berkeley, CA.
- 07/1997– 04/2000, Computer Science Engineer II, Lawrence Berkeley National Laboratory, Berkeley, CA.
- 03/2000–07/2000, Technical Adviser, Verifia, Inc, Mountain View, CA.
- 10/1996–11/1998, Co-founder, IVision Corp, Alameda, CA.
- 08/1995–06/1997, Artificial Intelligence Software Engineer, Stottler Henke Associates Inc, San Mateo, CA.

Selected Publications

- "*Improving Statistical Similarity Based Data Reduction for Non-Stationary Data*", D. Lee, A. Sim, J. Choi, K. Wu, 29th International Conference on Scientific and Statistical Database Management (SSDBM2017)
- "*Parallel Variable Selection for Effective Performance Prediction*", J. Wang, W. Yoo, A. Sim, P. Nugent, K. Wu, the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid2017), 2017.
- "*Comparison of Clustering Techniques for Residential Energy Behavior using Smart Meter Data*", L. Jin, S. Lee, A. Sim, A. Borgeson, K. Wu, A. Spurlock, A. Todd, 2nd International Workshop on Artificial Intelligence for Smart Grids and Smart Buildings, 2017.
- "*Time-series Forecast Modeling on High-Bandwidth Wide Area Network Measurements*", W. Yoo, A. Sim, Journal of Grid Computing, Vol. 14, Issue 3, pp 463–476, doi:10.1007/s10723- 016-9368-9, 2016
- "*Performance Analysis Tool for HPC and Big Data Applications on Scientific Clusters*", W. Yoo, M. Koo, Y. Cao, A. Sim, P. Nugent, K. Wu, in Conquering Big Data Using High Performance Computing, edited by R. Arora, pp.139-161, Springer International Publishing, 2016.
- *"Towards Real-Time Detection and Tracking of Spatio-Temporal Features: Blob-Filaments in Fusion Plasma*", L. Wu, K. Wu, A. Sim, M. Churchill, J. Choi, A. Stathopoulos, C-S Chang, S. Klasky, IEEE Transactions on Big Data (TBD), Vol. 2, Issue 3, pp. 262-275, 2016.
- "*Best Predictive Generalized Linear Mixed Model with Predictive Lasso for High-Speed Network Data Analysis*", K. Hu, J. Choi, A. Sim, J. Jiang, International Journal of Statistics and Probability, Vol. 4, No. 2, p132-148, 2015.
- "Efficient Attribute-based Data Access in Astronomy Analysis", B. Ma, A. Shoshani, A. Sim, K. Wu, Y.-I. Byun, J. Hahm and M.-S. Shin, The 2nd International Workshop on Network-Aware Data Management Workshop (NDM2012), Nov. 2012.
- "*StorNet: Integrated Dynamic Storage and Network Resource Provisioning and Management for Automated Data Transfers*", J. Gu, D. Katramatos, X. Liu, V. Natarajan, A. Shoshani, A. Sim, D. Yu, S. Bradley and S. McKee, Journal of Physics: Conf. Ser. vol. 331, 2011.
- *"Grid Collector: Facilitating Efficient Selective Access from Data Grids",* K. Wu, J. Gu, J. Lauret, A. Poskanzer, A. Shoshani, A. Sim, W. Zhang, International Supercomputer Conference, Germany, 2005, Best Paper Award in Data Management on Distributed Systems and Grids.

Synergistic Activities

- Software distribution under open source license, that is used throughout the world for distributed data storage and transfer management: Implementation of Dynamic Extensible Adaptive Locally Exchangeable Measures (IDEALEM), Berkeley Storage Manager (BeStMan), Bulk Data Mover (BDM), DataMover-Lite

(DML), Memory-to-Memory Zero-Copy Network Channel Technology (MemzNet), SRM-client took kit, SRM-Lite, SRM-Tester, etc.

- Technical program committees for conferences, steering and advisory committees for conferences and journal editorial board member and reviewer
- Founding member: Storage Resource Management Collaboration Group
- Patents
  - US Patent Pending serial no. 14/555,365, "*Data Reduction Methods, Systems, and Devices*", filed on 11/26/2014.
  - US Patent 20,120,269,053. "*Co-scheduling of network resource provisioning and host-to-host bandwidth reservation on high-performance network and storage systems*", 2012.
- Awards
  - Best paper award, "*Extracting Baseline Electricity Usage Using Gradient Tree Boosting*", International Conference on Big Data Intelligence and Computing (DataCom 2015).
  - Federal Laboratory Consortium For Technology Transfer, FAR-WEST Region, Outstanding Partnership Award, *Earth System Grid Federation*, 10/21/2013.
  - Best paper award, "*Grid Collector: Facilitating Efficient Selective Access from Data Grids*", International Supercomputer Conference, Heidelberg, Germany, 2005.
  - Hottest Infrastructure Award, *A Data Management Infrastructure for Climate Modeling Research* - Super Computing 2000 Network Challenge, 2000.
  - High Distinction in General Scholarship (Magna Cum Laude), U.C. Berkeley, 1990
  - Departmental Honors in Statistics, U.C. Berkeley, 1990

Recent Collaborators outside of LBNL

Dimitrios Katramatos, Dantong Yu (BNL); Inder Monga, Brian Tierney (ESnet); Constantine Dovrolis (Georgia Tech); Dean Williams (LLNL); Jong Choi, Scott Klasky, G. Ostrouchov (ORNL); C.S. Chang (PPPL); Jinoh Kim, Dongeun Lee (Texas A&M); Jiming Jiang (UC Davis); Jaesik Choi (UNIST)

A. Sim <ASim@lbl.gov> 510-495-2290

Han-Wei Shen

Professional Preparation

National Taiwan University, Taiwan	Computer Science	BS, 1988
University of New York, Stony Brook	Computer Science State	MS, 1992
University of Utah	Computer Science	PhD, 1998

Appointments

Professor, Computer Science and Engineering, The Ohio State University 2012-present
Associate Professor, Computer Science and Engineering, The Ohio State University 2005-2012
Assistant Professor, Computer Science and Engineering, The Ohio State University 1999-2005
Visiting Scientist, Argonne National Laboratory 10/08-6/09
Research Scientist, NASA Ames Research Center 9/96–8/99

Five Related Products

1. DUTTA, S., CHEN, C. M., HEINLEIN, G., SHEN, H. W., AND CHEN, J. P. In situ distribution guided analysis and visualization of transonic jet engine simulations. IEEE Transactions on Visualization and Computer Graphics 22, (1) (2017).

2. WANG, J., LIU, X., SHEN, H.-W., AND LIN, G. Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. IEEE Transactions on Visualization and Computer Graphics 22, (1) (2017).

3 .BISWAS, A., LIN, G., LIU, X., AND SHEN, H. W. Visualization of time-varying weather ensembles across multiple resolutions. IEEE Transactions on Visualization and Computer Graphics 22, (1) (2017).

4. DUTTA, S., AND SHEN, H.-W. Distribution driven extraction and tracking of features for time-varying data analysis. IEEE Tranactions on Visualization and Computer Graphics 22, (1) (2016).

5. CHEN, C. M., DUTTA, S., LIU, X., HEINLEIN, G., SHEN, H. W., AND CHEN, J. P. Visualization and analysis of rotating stall for transonic jet engine simulation. IEEE Transactions on Visualization and Computer Graphics 22, 1 (2016), 847–856.

Five Other Products

1. Lu, K., Shen, H.-W. A compact multivariate histogram representation for query-driven visualization. In 5th IEEE Symposium on Large Data Analysis and Visualization, LDAV 2015, Chicago, IL, USA, October 25-26, 2015 (2015)

2. Liu, X. and Shen, H-W. Association Analysis for Visual Exploration of Multivariate Scientific Data Sets. IEEE Transactions on Visualization and Computer Graphics 22(1): 955-964 (2016)

3. He, W., Chen, C., Liu, X., Shen, H-W: A Bayesian approach for probabilistic streamline computation in uncertain flows. IEEE PacificVis 2016: 214-218

4. Chen, C., Biswas, A., Shen, H-W: Uncertainty modeling and error reduction for pathline computation in time-varying flow fields. IEEE PacificVis 2015: 215-222

5. Biswas, A., Dutta, S., Shen, H-W, Woodring, J., An Information-Aware Framework for Exploring Multivariate Data Sets, IEEE Transactions on Visualization and Computer Graphics, Vol. 19, No. 2, pp. 2683-2690, Dec. 2013(special issue of IEEE Visualization 2013).

Synergistic Activities

Committee Members:

- IEEE Visualization Executive Committee, 2016 - Present
- IEEE SciVis Steering Committee, 2015 – Present
- IEEE Pacific Visualization Program Committee
- EuroVis Program Committee
- IEEE SciVis Program Committee
- Pacific Graphics Program Committee
- IEEE Clusters Program Committee
- ACM Interactive 3D Graphics Program Committee

Editors:

- Associate Editor, IEEE Transactions on Visualization and Computer Graphics 2007-2011
- Associate Editor, Journal of Visualization 2010- Present
- IEEE Computer Graphics and Applications Guest Editors

Co-Chairs:

- IEEE Visualization SciVis Paper co-Chair, 2013, 2014
- IEEE Pacific Visualization Paper co-Chair 2009, 2010
- Program co-Chair, IEEE Visualization 2008-2009
- Eurographics Symposium on Parallel Graphics and Visualization (EGPGV) 1999, 2004

,

CHAD A. STEED BIOGRAPHICAL SKETCH

Education and Training

- Mississippi State Univ., Computer Science (Computer Graphics and Visualization), Ph.D., 2008.
- University of Southern Mississippi, Hydrographic Science, M.S., 2004.
- University of Southern Mississippi, Software Engineering (minor in Fine Art), B.S., 1999.

Research and Professional Experience

- Senior Research Staff and Team Leader, Oak Ridge National Laboratory, 5/2010–present.
  At ORNL, I lead a data science research team. My research focuses on interactive data visualization, visual analytics, and statistical data analysis with broad scientific applications.
- Joint Faculty Appointment, University of Tennessee, Dept. of Electrical Engineering and Computer Science, Knoxville, TN, 1/2013–present.
  At UT, my duties include teaching, faculty collaboration, and advising students.
- Computer Scientist, Naval Research Laboratory, Stennis Space Center, MS, 7/2001–5/2010.
  At NRL, my research focused on interactive data visualization, databases, data analytics, and acoustic sonar processing with oceanographic and geoacoustic applications.
- Software Engineer, Lockheed Martin, Stennis Space Center, MS, 5/1999–7/2001.
  I developed and maintained bathymetric databases and cartographic applications for the Navy.

Publications (10 Most Related)

Chad A. Steed. "Interactive Data Visualization". In *Data Analytics for Intelligent Transportation*. Eds. M. Chowdhury, A. Apon, and K. Dey. pp. 165–190, 2017.

Chad A. Steed, William Halsey, Ryan Dehoff, Sean L. Yoder, Vincent Paquit, and Sarah Powers. "Falcon: Visual Analysis of Large, Irregularly Sampled, and Multivariate Time Series Data in Additive Manufacturing". *Computers & Graphics*, 63:50–64, 2017.

Chad A. Steed, Jamison Daniel, Margaret Drouhard, Thomas Proffen, and Steven Hahn. "Immersive Visual Analytics for Transformative Neutron Scattering Science". In *Proceedings of the 1st Immersive Analytics Workshop at IEEE VR 2016*, Mar. 2016.

Blake Haugen, Stephen Richmond, Jakub Kurzak, **Chad A. Steed**, and Jack Dongarra. "Visualizing Execution Traces with Task Dependencies". In *Proceedings of the 2nd Workshop on Visual Performance Analysis at SC '15*, Austin, TX, Nov. 2015.

Chad A. Steed, Katherine J. Evans, John F. Harney, Brian C. Jewell, Galen Shipman, Brian E. Smith, Peter E. Thornton, and Dean N. Williams. "Web-based Visual Analytics for Extreme Scale Climate Science". In *Proceedings of the IEEE International Conference on Big Data*, pp. 383–392, Oct. 2014.

Chad A. Steed, Daniel M. Ricciuto, Galen Shipman, Brian Smith, Peter E. Thornton, Dali Wang, and Dean N. Williams. "Big Data Visual Analytics for Earth System Simulation Analysis". *Computers & Geosciences*, 61:71–82, 2013. ***Best Paper of 2013 Award***

Dean N. Williams, ..., **Chad A. Steed**, ..., and Thomas P. Maxwell. "The Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT): Data Analysis and Visualization for Geoscience Data". *IEEE Computer*. 46(9):68–76, 2013.

Chad A. Steed, J. Edward Swan II, Patrick J. Fitzpatrick, and T.J. Jankun-Kelly. "A Visual Analytics Approach for Correlation, Classification, and Regression Analysis". In *Innovative Approaches of Data Visualization and Visual Analytics*. Eds. M. Huang and W. Huang. pp. 25–45, 2013.

Chad A. Steed, Galen Shipman, Peter Thornton, Daniel Ricciuto, David Erickson, and Marcia Branstetter. "Practical Application of Parallel Coordinates for Climate Model Analysis". In *Proceedings of the International Conference on Computational Science*, pp. 877–886, June 2012.

Chad A. Steed, J. Edward Swan II, T.J. Jankun-Kelly, and Patrick J. Fitzpatrick. "Guided Analysis of Hurricane Trends using Statistical Processes Integrated with Interactive Parallel Coordinates". In *Proceedings of Symposium on Visual Analytics Science and Technology*, Atlantic City, NJ, pp. 19–26, 2009.

Synergistic Activities

Awards:
- ORNL Best Director's Research & Development Poster Award (2016)
- ORNL/UT-Battelle Early-Career Researcher Award (2014)
- ORNL Technology Commercialization Award (2013, 2014)
- ORNL Significant Event Award (2013, 2014)
- Elsevier Computers & Geosciences Journal Best Paper Award (2013)
- R&D 100 Award (see patent No. 13/737,652 above) (2013)
- 2005 Naval Research Laboratory Technology Transfer Award
- 2005 Naval Research Laboratory Select Graduate

Patents:
Chad A. Steed, Robert M. Patton, Paul L. Bogen, Thomas E. Potok, Christopher T. Symons. "Interactive Visual Analytics for Situational Awareness of Social Media". Application No. 14/476,252, filed 3 Sep. 2014.

Thomas E. Potok, Robert Patton, **Chad A. Steed**. "Method and System to Discover and Recommend Interesting Documents". Application No. 13/737,652, filed 9 Jan. 2013.

Chad A. Steed, J. Edward Swan II, T.J. Jankun-Kelly, and Patrick J. Fitzpatrick. "Information Assisted Visual Interface, System, and Method for Identifying and Quantifying Multivariate Associations". 1 Jan. 2013.

Senior Member: ACM, ACM SIGGRAPH, ACM SIGCHI, IEEE, and IEEE Computer Society.

Program Committee: IEEE Visualization in Data Science Symposium (2017), Visualization and Data Analysis Conference (2012, 2013–2017), IEEE VIS Arts Program (2015–2017).

Organizing Committee: IEEE Workshop on Interactive Visual Text Analysis (2013).

Session Chair: ACM Intelligent Use Interfaces Conference (2015), Visualization and Data Analysis Conference (2012), MTS/IEEE Oceans (2009).

Journal Reviewer: ACM Trans. on Applied Perception (2016, 2017), Neurocomputing (2016), ACM Trans. on Interactive Intelligent Systems (2015), GeoInformatica (2015), IEEE Computer Graphics and Applications (2013), Journal of Computational and Graphical Statistics (2013), WIREs Computational Statistics (2012, 2013), Computers & Geosciences Journal (2012), Cartography and Geographic Information Science (2011), International Journal of Geographical Information Science (2009).

Conference Reviewer: IEEE VIS Arts Program (2015, 2016), ACM CHI (2015), ACM Intelligent User Interfaces (2015), IEEE Symposium on Biological Data Visualization (2014), SPIE Visualization and Data Analysis Conference (2012–2017), IEEE EuroVis Conference (2012), Hawaii International Conference on System Science (2012), IEEE Visualization Conference (2008–2017), IEEE Information Visualization Conference (2009–2011, 2013–2017), IEEE Visual Analytics Science and Technology Conference (2010, 2011, 2013–2017), IEEE VisWeek Poster Session (2010), MTS/IEEE Oceans Conference (2009), IEEE Pacific Visualization Symposium (2009), IEEE Virtual Reality Conference (2007), International Symposium on Mixed and Augmented Reality (2007).

Review Panel: NSF Information & Intelligent Systems Division (IIS) Panel (2013, 2014, 2015, 2016), DOE SBIR Review (2011, 2012, 2014), NSF Arctic Ocean Visualization (2010), SPAWAR SBIR Optimal Seafloor Mapping Technologies (2009).

Collaborators and Co-editors for prior 2 years (alphabetical by last name)

Amy Apon, Clemson University
Michael Berry, University of Tennessee
Aashish Chaudhary, Kitware
Mark Dean, University of Tennessee
Alex Endert, Georgia Tech University
Christopher Healey, North Carolina State University
Donald House, Clemson University
Jian Huang, University of Tennessee
Joshua Levine, University of Arizona
Chris North, Virginia Tech University
Kevin Shaw, Naval Research Laboratory
Galen Shipman, Los Alamos National Laboratory
Dean Williams, Lawrence Livermore National Laboratory

Graduate and Postdoctoral Advisors and Advisees

Patrick Fitzpatrick, Mississippi State University
T.J. Jankun-Kelly, Mississippi State University
Robert Moorhead, Mississippi State University
J. Edward Swan II, Mississippi State University

Shinjae Yoo

Brookhaven National Laboratory

Education and Training:

Ph.D., Carnegie Mellon University, Language Technologies Institute, School of Computer Science, 2010

Masters of Language Technologies, Carnegie Mellon University, Language Technologies Institute, School of Computer Science, 2005, Pittsburgh, USA

M.S., Seoul National University, Computer Science, 2002, Seoul, Korea

B.E., Soongsil University, Computer Science, 2000, Seoul, Korea

Research and Professional Experience:

10/2016 – present	Scientist, Computational Science Initiative, BNL
10/2014 – present	Adjunct Assistant Prof., IACS, Stony Brook University
10/2013 – 09/2016	Associate Scientist, Computational Science Initiative, BNL
10/2011 – 09/2013	Assistant Scientist, Computational Science Center, BNL
06/2010 – 09/2010	Research Associate, Computational Science Center, BNL
05/2006 – 08/2006	Engineering Intern, Google Inc., Mountain View, CA

Selected Publications:

1. Shinjae Yoo, Hao Huang, Shiva Kasiviswanathan, "Streaming Spectral Clustering". IEEE International Conference on Data Engineering (ICDE), 2016.
2. Jin Xu, Shinjae Yoo, John Heiser, Paul Kalb, "Sensor Network Based Solar Forecasting Using a Local Vector Autoregressive Ridge Framework". ACM SAC '16.
3. Hao Huang, Shinjae Yoo, Dantong Yu, Hong Qin, "Diverse Power Iteration Embeddings: Theory and Practice". IEEE TKDE.
4. Hao Huang, Shinjae Yoo, Shiva Kasiviswanathan, "Unsupervised Feature Selection on Data Streams". ACM CIKM, 2015.
5. Shun Yao, Shinjae Yoo, Dantong Yu, "Prior knowledge driven Granger causality analysis on gene regulatory network discovery". BMC Bioinformatics, 16:273, August 2015.
6. Hao Huang, Shinjae Yoo, Dantong Yu, Hong Qin, "Density-aware Clustering based on Aggregated Heat Kernel and Its Transformation". ACM TKDD, Volume 9 Issue 4, 2015.
7. Hao Huang, Shinjae Yoo, Dantong Yu, Hong Qin, "Diverse Power Iteration Embeddings and Its Applications". IEEE International Conference on Data Mining (ICDM), 2014.
8. Shinjae Yoo, Yiming Yang, Jaime Carbonell, “Modeling Personalized Email Prioritization: Classification-based and Regression-based Approaches”, ACM CIKM, 2011.
9. Shinjae Yoo, Yiming Yang, Frank Lin and Il-Chul Moon, “Mining Social Networks for Personalized Email Prioritization”, ACM SIGKDD, 2009
10. Yiming Yang, Shinjae Yoo, Jian Zhang, Bryan Kisiel, “Robustness of adaptive filtering methods in a cross-benchmark evaluation”, ACM SIGIR, 2005

Synergistic Activities:

Reviewer/Program Committee: ACM SAC, ACM TOIT (Transactions on Internet Technology), Journal of Computer Science and Technology, DOE SBIR, ACM SIGKDD, SocialCom Workshop in IEEE Conference on Social Computing, International Conference on Digital Government Research, ACM SIGIR

Professional membership: ACM, IEEE, ISCB

Identification of Potential Conflicts of Interest or Bias in Selection of Reviewers:

Collaborators and Co-editors:

Hao Huang (GE Research), Shiva Kasiviswanathan (Samsung Research), Fei He (UIUC, University of Illinois at Urbana-Champaign), Daifeng Wang (Yale University), Sunita Kumari (CSHL, Cold Spring Harbor Lab), Mark Gerstein (Yale University), Dorren Ware (CSHL), Sergei Maslov (UIUC), Dantong Yu (BNL), Hong Qin (SBU, Stony Brook University), Jin Xu (SBU), John Heiser (BNL), Paul Kalb (BNL), Shun Yao (SBU), Zhenzhou Peng (SBU), Adrian Soto Cambres (SBU), Deyu Lu (BNL), Maria Fenandez-Serra (SBU), Hayan Lee (LBL, Lawrence Berkeley National Laboratory), Yangang Liu (BNL), Dmitri Zakharov (BNL), Eric Stach (BNL), Dimitrios Katramatos (BNL), Kerstin Kleese van Dam (BNL), James Gurtowski (CSHL), Shoshana Marcus (City University of NY), Richard McCombie (CSHL), Michael Schatz (Johns Hopkins University), Huan Feng (Montclair University), Jun Wang (BNL), Jia-Jun Wang (BNL), Christopher Eng (BNL), Chang-Jun Liu (BNL), Ryan Tappero (BNL), Adam Arkin (LBL), Ricky Stevens (ANL, Argone National Laboratory), Bob Cottingham (ORNL, Oakridge National Laboratory), Thomas Brettin (ANL), Dan Murhpy-Olson (ANL), Shane Richard Canon (LBL), Christopher Henry (ANL), Manoj Mahajan (National Geospatial-intelligence Agency)

Graduate and Postdoctoral Advisors and Advisees:

Yiming Yang (Carnegie Mellon University) Dantong Yu (BNL), Jin Xu (Stony Brook University), Mingshen Chen (Stony Brook University), Xi Zhang (Stony Brook University)

Samuel W. Williams

Computational Research Division, Lawrence Berkeley National Laboratory
One Cyclotron Road, Mail Stop 59R4104, Berkeley, CA 94720
SWWilliams@lbl.gov

Education and Training

2008	PhD	Computer Science	University of California, Berkeley
2003	MS	Computer Science	University of California, Berkeley
1999	BSEE	Electrical Engineering	Southern Methodist University
1999	BS	Physics	Southern Methodist University
1999	BS	Mathematics	Southern Methodist University

Research and Professional Experience

Lawrence Berkeley National Lab	Staff Scientist	Jan 2009 – present
Lawrence Berkeley National Lab	Graduate Student Researcher	Jan 2005 – Jan 2009
University of California, Berkeley	Graduate Student Researcher	Jul 1999 – Jan 2005
Cyrix/National Semiconductor	intern	Jun 1996 – Jul 1999 (alternating semesters)

Selected Publications

- Pieter Ghysels, Xiaoye S. Li, François-Henry Rouet, Samuel Williams, Artem Napov, "An Efficient Multicore Implementation of a Novel HSS-Structured Multifrontal Solver Using Randomized Sampling", SIAM J. Sci. Comput. 38-5, pp. S358-S384.
- William Tang, Bei Wang, Stephane Ethier, Grzegorz Kwasniewski, Torsten Hoefler, Khaled Z. Ibrahim, Kamesh Madduri, Samuel Williams, Leonid Oliker, Carlos Rosales-Fernandez, Tim Williams, "Extreme Scale Plasma Turbulence Simulations on Top Supercomputers Worldwide", Supercomputing, November 2016.
- B. Wang, S. Ethier, W. Tang, T. Williams, K. Ibrahim, K. Madduri, S. Williams, L. Oliker, "Kinetic Turbulence Simulations at Extreme Scale on Leadership-Class Systems", Supercomputing, 2013.
- S. Williams, J. Carter, L. Oliker, J. Shalf, K. Yelick, "Lattice Boltzmann Simulation Optimization on Leading Multicore Platforms", International Parallel & Distributed Processing Symposium (IPDPS), 2008. **BEST PAPER, applications track**
- K. Datta, M. Murphy, V. Volkov, S. Williams, J. Carter, L. Oliker, D. Patterson, J. Shalf, K. Yelick, "Stencil Computation Optimization and Autotuning on State-of-the-Art Multicore Architectures", Supercomputing, 2008.
- S. Williams, D. Kalamkar, A. Singh, A. Deshpande, B. Van Straalen, M. Smelyanskiy, A. Almgren, P. Dubey, J. Shalf, L. Oliker, "Optimization of Geometric Multigrid for Emerging Multi- and Manycore Processors", Supercomputing, 2012.
- S. Williams, E. Carson, M. Lijewski, N. Knight, A. Almgren, J. Demmel, B. Van Straalen, "s-step Krylov Subspace Methods as Bottom Solvers for Geometric Multigrid", International Parallel and Distributed Processing Symposium (IPDPS), 2014.
- S. Williams, L. Oliker, R. Vuduc, J. Shalf, K. Yelick, J. Demmel, "Optimization of Sparse Matrix-Vector Multiplication on Emerging Multicore Platforms", Supercomputing, 2007.
- S. Williams, A. Waterman, D. Patterson, "Roofline: An Insightful Visual Performance Model for Floating-Point Programs and Multicore Architectures", Communications of the ACM, 2009.
- S. Williams, J. Shalf, L. Oliker, S. Kamil, P. Husbands, K. Yelick, "The Potential of the Cell Processor for Scientific Computing", ACM International Conference on Computing Frontiers, 2006.

Synergistic Activities

Exascale Requirements Review for Advanced Scientific Computing Research, 2016
Exascale Requirements Review for Biological and Environmental Research, 2016
Advancing X-cutting Ideas for Computational Climate Science (AXICCS) Program Committee
CoHPC 2014-15 Program Committee

Collaborators (last 48 months) and Co-Editors (last 24 months)

H.M. Aktulga (Michigan State), G. Ballard (Wake Forest), A. Barker (LLNL), S. Biedron (CSU), G. Bosilca (UTK), J. Brown (CU), P. Caldwell (LLNL), L. Carrington (UCSD), N. Chaimov (UO), J. Chame (USC), J. Chen (SNL), K. Clark (NVIDIA), B. de Supinski (LLNL), B. Debusschere (SNL), E. Epifanovsky (Q-Chem), S. Ethier (PPPL), K. Evans (ORNL), R. Falgout (LLNL), R. Fowler (UNC), J. Gilbert (UCSB), N. Gorelenkov (PPPL), L. Grigori (INRIA), R. Grout (NREL), M. Hall (Utah), R. Heikes (CSU), O. Hernandez (ORNL), T. Hoefler (ETH), F. Hoffman (ORNL), J. Hollingsworth (UMD), P. Hovland (ANL), AP. Hynninen (ORNL), C. Jablonowski (Michigan), C. Jackson (Texas), D. Jacobsen (Intel), S. Jardin (PPPL), C. Johnson (SDSU), P. Jones (LANL), D. Kalchev (LLNL), G. Kwasniewski (ETH), A. Kamil (Michigan), S. Kamil (Adobe), S. Klasky (ORNL), A. Krylov (USC), E. Lawrence (LANL), R. Leung (PNNL), Z. Lin (UC Irvine), B. Lucas (ISI/USC), A. Lugowski (UCSB), K. Madduri (Penn State), A. Malony (UO), P. Maris (Iowa State), S. Moore (UTK), R. Moser (Texas), B. Norris (ANL), W.E. Ormand (LLNL), G. Ostrouchov (ORNL), J. Owens (UCD), D. Quinlan (LLNL), D. Randall (CSU), T. Ringler (LANL), C. Rosales-Fernandez (TACC), O. Schwartz (Hebrew University), M. Shephard (RPI), B. Smith (ANL), D. Spong (ORNL), T. Straatsma (ORNL), A. Tharrington (ORNL), S. Toledo (Tel Aviv), R. Tuminaro (SNL), P. Ullrich (UC Davis), M. Umansky (LLNL), E. Valeev (VT), K. Van Dam (BNL), J. Vary (Iowa State), P. Vassilevski (LLNL), R. Waltz (General Atomics), S. Wild (ANL), T. Williams (ANL), P. Worley (ORNL), U. Yang (LLNL), Y. Zheng (Google)

Graduate and Postdoctoral Advisors and Advisees:

D.A. Patterson (University of California, Berkeley – Thesis advisor)
Protonu Basu (Lawrence Berkeley National Lab – Postdoctoral advisee)

Jeffrey S. Vetter

Oak Ridge National Laboratory
One Bethel Valley Road
Oak Ridge, TN 37831-6173

vetter@computer.org
https://ft.ornl.gov/~vetter
+1-720-239-2472

1 Professional Preparation

1. Postdoctoral research associate. University of Illinois at Urbana-Champaign. 1998-9.
2. Ph.D. in Computer Science. Georgia Institute of Technology. Atlanta. 1998.
3. M.S. in Computer Science. Georgia Institute of Technology. Atlanta. 1994.
4. B.S. in Electrical/Computer Engineering with Honors. Georgia Institute of Technology. Atlanta. 1989.

2 Appointments

1. Oak Ridge National Laboratory.
   (a) Distinguished R&D Staff Member and Group Leader. November 2009 - present.
   (b) Founding Group Leader of ORNL Future Technologies Group. 2004 - present.
   (c) Founding Director of Experimental Computing Laboratory (ExCL). 2004 - present.
   (d) Senior R&D Staff Member. November 2003 - November 2009
2. University of Tennessee - Knoxville, Department of Electrical Engineering and Computer Science.
   (a) Joint Professor. February 2016 - present
3. Georgia Institute of Technology, College of Computing.
   (a) Joint Professor. October 2006 - 2016.
   (b) Principal Investigator and Director, NVIDIA CUDA Center of Excellence. 2010 - 2016.
   (c) Principal Investigator and Project Director, NSF Keeneland Project. 2009 - 2015.
   (d) Adjunct Professor. May 2005 - October 2006.
4. Lawrence Livermore National Laboratory. Computer Scientist and Project Leader. 1999 - 2003.
5. Los Alamos National Laboratory. Graduate research intern. Summers of 1996, 1997.
6. Intel Corporation. Graduate research intern. Sept 1991 - Dec 1992, Summer 1993.
7. Hewlett Packard Corporation (now Agilent Technologies). Technical staff member. 1989 - 1991.

3 Products (Most publications available at Google Scholar.)

3.1 Closely-related Publications

1. J. Kim, K. Sajjapongse, S. Lee, and J. S. Vetter. Design and implementation of papyrus: Parallel aggregate persistent storage. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, IPDPS '17, Orlando, 2017. ACM
2. J. E. Denny, S. Lee, and J. S. Vetter. Language-based optimizations for persistence on nonvolatile main memory systems. In *International Parallel and Distributed Systems Symposium (IPDPS)*, Orlando, 2017. IEEE
3. J. E. Denny, S. Lee, and J. S. Vetter. NVL-C: Static analysis techniques for efficient, correct programming of non-volatile main memory systems. In *Proceedings of the 25th ACM International Symposium on High-Performance Parallel and Distributed Computing*, pages 125–136, Kyoto, Japan, 2016. ACM. doi:10.1145/2907294.2907303
4. J. S. Vetter and S. Mittal. Opportunities for nonvolatile memory systems in extreme-scale high performance computing. *Computing in Science and Engineering*, 17(2):73–82, 2015. doi:10.1109/mcse.2015.4
5. S. Mittal and J. S. Vetter. A survey of software techniques for using non-volatile memories for storage and main memory systems. *IEEE Transactions on Parallel and Distributed Systems*, 27(5):1537–1550, 2016. doi:10.1109/TPDS.2015.2442980

3.2 Other Significant Publications

1. J. S. Vetter, editor. *Contemporary High Performance Computing: From Petascale Toward Exascale*,

volume 2 of *CRC Computational Science Series*. Taylor and Francis, Boca Raton, 1 edition, 2015

2. J. S. Vetter, R. Glassbrook, K. Schwan, S. Yalamanchili, M. Horton, A. Gavrilovska, M. Slawinska, J. Dongarra, J. Meredith, P. C. Roth, K. Spafford, S. Tomov, and J. Wynkoop. Keeneland: Computational science using heterogeneous GPU computing. In Jeffrey S. Vetter, editor, *Contemporary High Performance Computing: From Petascale Toward Exascale*, volume 1 of *CRC Computational Science Series*, page 730. Taylor and Francis, Boca Raton, 1 edition, 2013
3. L. Yu, D. Li, S. Mittal, and J. S. Vetter. Quantitatively modeling application resilience with the data vulnerability factor (best student paper finalist). In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 695–706, New Orleans, Louisiana, 2014. IEEE Press. doi:10.1109/sc.2014.62
4. K. L. Spafford and J. S. Vetter. Aspen: A domain specific language for performance modeling. In *SC12: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–11, Salt Lake City, 2012. doi:10.1109/SC.2012.20
5. S. Lee, J. Kim, and J. S. Vetter. OpenACC to FPGA: A framework for directive-based high-performance reconfigurable computing. In *IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, Chicago, 2016. IEEE. doi:10.1109/IPDPS.2016.28

4 Recent Synergistic Activities

1. International Workshop on Post Moores Supercomputing, Co-founder and Co-chair, 2016-7.
2. SC15, Technical Program Chair, 2015.
3. ACM Journal on Emerging Technologies in Computing (JETC), Associate Editor, 2016 - present.
4. ISC, Steering Committee, 2016 - present.
5. Have mentored over 50 students for internships in HPC.

5 Collaborators

This list omits organizational conflicts at ORNL, UTK, and Georgia Tech.

Alam, Sadaf R. (CSCS); Ang, James (Sandia); Anitescu, Mitescu (ANL); Beckman, Pete (ANL); Belak, James (LLNL); Biros, George (Texas); Brightwell, Ron (Sandia); Campbell, Dan (Georgia Tech Research Institute); Carothers, Christopher (Rensselaer Polytechnic Institute); Dally, William (NVIDIA); Davis, Keith (LANL); Deelman, Ewa (ISI); Dongarra, Jack (U. Tennessee); Dubey, A. (ANL); Fahey, Mark R. (ORNL/UTK); Faraboschi, Paolo (HP); Fischer, Paul (ANL); Germann, Timothy (LANL); Ghattas, Omar (Texas); Hanrahan, Pat (Stanford); Harrison, Robert J. (BNL); Heroux, Michael (Sandia); Hoisie, Adolfy (PNNL); Hovland, Paul (ANL); Hwu, Wen-mei (UIUC); Johnansen, Hans (LBL); Jouppi, Norm (Google); Karavanic, Karen (U. Portland); Keasler, Jeffrey (LLNL); Keckler, Steven (NVIDIA); Kuehn, Jeffrey (LANL); Li, Xiaoye (Sherry) (LBL); Lookman, T. (LANL); Lucas, B. (ISI); Malony, Allan (U. Oregon); Mandal, Anirban (RENCI/UNC Chapel Hill); McCormick, Patrick (LANL); McInnes Curfman, Lois (ANL); Mohd-Yusof, J. (LANL); Moulton, David (LANL); Mudge, Trevor (U. Michigan); Mueller, Frank (NCSU); Norris, Boyana (University of Oregon); Ortiz, Michael (CalTech); Owhadi, H. (CalTech); Pancake, Cherri (Oregon State); Peterka, T. (ANL); Poole, Stephen W. (LANL); Ragahavan, Padma (Penn State); Reed, Daniel (Iowa); Richards, David (LLNL); Rodrigues, Arun (Sandia); Rosner, Robert (ANL); Salinger, Andrew (ANL); Scheibe, Timothy (PNNL); Schreiber, Rob (HP); Schulthess, Thomas C. (ETH Zurich); Schulz, Martin (LLNL); Shalf, John (LBNL); Shende, Sameer (U. Oregon); Siegel, Andrew (ANL); Smith, Kord (MIT); Smith, M. (ANL); Steefel, Carl (LBL); Sterling, Thomas (Indiana); Swaminarayan, S. (LANL); Tierney, Brian (LBL); Warburton, Timothy (Rice); Yang, Ulrike (LLNL); Yelick, Kathy (LBNL); Yu, Weikuan (Auburn);

Biographical Sketch

GUNTHER H. WEBER

July, 2017

Computational Research Division
Lawrence Berkeley National Laboratory (LBNL)
One Cyclotron Road, MS: 50F-1650
Berkeley, CA 94720

http://crd.lbl.gov/q/ghweber
Voice: (510) 486-5871
Fax: (510) 486-5812
E-mail: GHWeber@lbl.gov

BIOGRAPHICAL SKETCH

Weber is a Staff Scientists in the Lawrence Berkeley National Laboratory's (LBNL) Computational Research Division and an Adjunct Associate Professor of Computer Science at the University of California, Davis. Weber has been PI/Co-PI on several DOE ASCR and NSF projects. His research interests include computer graphics, scientific visualization, data analysis, topological data analysis methods, parallelization of visualization algorithms, hierarchical data representation methods, and bioinformatics. He has extensive experience in working with researchers from numerous science and engineering domains including applied numerical algorithms, combustion simulations, gene expression, medicine, civil engineering, cosmology, climate and particle accelerator modeling. Weber has over 70 publications and has served as reviewer for funding agencies (DOE, NSF) as well as on international program committees of internationally recognized conferences.

EDUCATION AND TRAINING

University of California, Davis	Visualization and Data Analysis	Postdoc	2003 – 2006
University of Kaiserslautern	Computer Science	Ph.D.	2003
University of Kaiserslautern	Computer Science (minor: Physics)	M.S.	1999
University of Kaiserslautern	Computer Science (minor: Physics)	B.S.	1994

RESEARCH AND PROFESSIONAL EXPERIENCE

01/2015 – Present	Staff Scientist/Engineer, LBNL, Berkeley, CA.
07/2014 – Present	Adjunct Associate Professor, University of California, Davis, CA.
02/2008 – 12/2014	Computer Research Scientist/Engineer, LBNL, Berkeley, CA.
07/2008 – 06/2014	Adjunct Assistant Professor, University of California, Davis, CA.
12/2006 – 12/2008	Participating Guest, Lawrence Livermore National Lab., Livermore, CA.
01/2007 – 01/2008	Computer Systems Engineer, LBNL, Berkeley, CA.
08/2006 – 12/2006	Assistant Project Scientist, UC Davis, CA.
09/2003 – 12/2006	Guest Researcher, LBNL, Berkeley, CA.
09/2003 – 07/2006	Postdoctoral Researcher, UC Davis, CA.
10/2002 – 08/2003	Visiting Scholar, University of California, Davis, CA.
09/2001 – 09/2002	Research Assistant, University of Kaiserslautern, Germany.
02/2001 – 07/2001	Visiting Scholar, University of California, Davis, CA.
10/2000 – 08/2003	Guest Student Assistant, LBNL, Berkeley, CA.
10/2000 – 01/2001	Research Assistant, University of Kaiserslautern, Germany.
07/2000 – 09/2000	Student Employee, LBNL, Berkeley, CA.
02/2000 – 06/2000	Visiting Scholar, University of California, Davis, CA.
09/1999 – 01/2000	Researcher, German Research Center for Artificial Intelligence GmbH, Kaiserslautern, Germany.

SELECT RECENT PUBLICATIONS

1. J. Lukasczyk, R. Maciejewski, G. H. Weber, C. Garth, and H. Leitte. Nested tracking graphs. *Computer Graphics Forum (Special Issue, Proceedings Eurographics/IEEE Symposium on Visualization)*, 36(3), 2017. In press.
2. H. A. Carr, G. H. Weber, C. M. Sewell, and J. P. Ahrens. Parallel peak pruning for scalable SMP contour tree computation. In *Proceedings of the 6th IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pages 75–84, Oct. 2016. Best paper award.
3. S. Murugesan, K. Bouchard, J. A. Brown, B. Hamann, W. W. Seeley, A. Trujillo, and G. H. Weber. Brain Modulyzer: Interactive visual analysis of functional brain connectivity. *IEEE Transactions on Computational Biology and Bioinformatics*, 2016. In press.

4. P. Oesterling, C. Heine, G. H. Weber, D. Morozov, and G. Scheuermann. Computing and visualizing time-varying merge trees for high-dimensional data. In H. Carr, C. Garth, and T. Weinkauf, editors, *Topological Methods in Data Analysis and Visualization IV: Theory, Algorithms, and Applications*, Mathematics and Visualization. Springer-Verlag, 2016. Best paper award, in press.
5. D. Morozov and G. H. Weber. Distributed merge trees. In *Proceedings of the 18th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP'13)*, pages 93–102, New York, NY, USA, 2013. ACM.

SELECT SYNERGISTIC ACTIVITIES

1. Panelist and technical proposal reviewer for U. S. Department of Energy and National Science Foundation.
2. Organizing Committee: In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization 2015–2017; ISC Workshop on In-Situ Visualization 2017.
3. Posters Co-Chair, IEEE VIS (SciVis) 2016/2017; Visualization Co-Chair, 11th International Symposium on Visual Computing, 2015; Workshop Co-Chair, Visualization in Environmental Sciences (EnvirVis), 2015.
4. Program Committee: IEEE Vis 2011–2013; EuroVis 2009–2011 & 2014–2016; EuroVis Short Papers 2013–2017; Topology-Based Methods in Visualization (TopoInVis) 2007/2009/2011/2013/2015/2017; ISC Workshop on In-Situ Visualization 2016; IEEE Workshop on High Performance Computing and Big Data in Molecular Engineering 2016; Symposium on Visualization in High Performance Computing at SIGGRAPH Asia 2015; EnvirVis 2013/2014; Symposium on Biological Data Visualization 2012–2014, SPIE Visualization and Data Analysis 2012, Eurographics Workshop on Visual Computing for Biology and Medicine 2012.
5. Technical and journal paper review: IEEE Visualization, IEEE InfoVis, IEEE VAST, IEEE Transactions on Visualization and Computer Graphics, Joint Eurographics/IEEE Symposium on Visualization, PacificVis, Pacific Graphics, The Visual Computer, Computer Graphics and Applications, Workshop on Visualization in Medicine and Life Sciences, Journal of Earth Simulator, International Journal of Software and Informatics, Concurrency and Computation: Practice and Experience, IASTED International Conference on Computer Graphics, International Journal of High Performance Computing, SciDAC conference, International Conference on Numerical Modeling of Space Plasma Flows.

COLLABORATORS AND C-EDITORS

J. Ahrens (LANL), A. Almgren (LBNL), K. Antypas (LBNL), U. Ayachit (Kitware), A. Bauer (Kitware), V. Beckner (LBNL), J. Bell (LBNL), E.W. Bethel (LBNL), K. Bouchard (LBNL), J. Brown (UCSF), E. Brugger (LLNL), H. Carr (Univ. of Leeds), H. Childs (Univ. of Oregon), P. Colella (LBNL), M. Day (LBNL), E. Duque (Intelligent Light), A. Ebert (TU Kaiserslautern), H. Elgammal (LBNL), P. Fasel (LANL), N. Ferrier (ANL), B. Friesen (LBNL), C. Garth (TU Kaiserslautern), B. Geveci (Kitware), H. Hagen (TU Kaiserslautern), B. Hamann (UC Davis), M. Haranczyk (LBNL), V. Hendrix (LBNL), C. Heine (TU Kaiserslautern), K. Joy (UC Davis), D. Knowles (LBNL), H. Leitte (TU Kaiserslautern), T. Liebmann (Univ. of Leipzig), T. Ligocki (LBNL), J. Lukasczyk (TU Kaiserslautern), Z. Lukic (LBNL), J. Meredith (ORNL, now Google), D. Morozov (LBNL), P. Nugent (LBNL), P. Oesterling (Univ. of Leipzig), P. OLeary (Kitware), A. Ovsyannikov (LBNL), V. Pascucci (Univ. of Utah), Prabhat (LBNL), D. Pugmire (ORNL), B. Quiter (LBNL), L. Ramakrishnan (LBNL), A. Sanderson (Univ. of Utah), C. Sewell, G. Scheuermann (Univ. of Leipzig), W. Seeley (UCSF), C. Sewell (LANL), D. Trebotich (LBNL), V. Vishwanath (ANL), B. Van Straalen (LBNL), D. Ushizima (LBNL), B. Whitlock (Intelligent Light), M. Wolf (ORNL), K. Wu (LBNL)

GRADUATE AND POSTDOCTORAL ADVISORS AND ADVISEES

- *PhD. Advisors:* Hans Hagen (University of Kaiserslautern), Bernd Hamann (UC Davis), Gerik Scheuermann (University of Kaiserslautern; now University of Leipzig)
- *Postdoctoral Advisor:* B. Hamann
- *Graduate students co-supervised:* K. Beketayev (Ph.D., UC Davis), S. Dillard (Ph.D., UC Davis), M.-Y. Huang (Ph.D., UC Davis), S. Murugesan (Ph.D., UC Davis), O. Rübel (Ph.D., University of Kaiserslautern);
- *Postdoctoral Scholars supervised and co-supervised:* D. Morozov (LBNL), M. Hlawitschka (UC Davis)

Kesheng Wu
(John)

One Cyclotron Road, Berkeley, CA 94720
KWu@lbl.gov, (510)486-6609

Biographical Sketch

Dr. Wu works actively on a number of topics in data management, data analysis, and high-performance computing. His algorithmic research work includes bitmap indexing techniques for searching large datasets, statistical methods for extract features from a variety of data, and restarting strategies for computing extreme eigenvalues. He authored and coauthored more than 100 technical publications, eight of which have more than 100 citations each.

He is the developer of a number of software packages, including, IDEALEM, SDS, FastBit and TRLan. Among them, the FastBit software for indexing large datasets has earned an R&D 100 Award, and is used by many organizations. For example, a German bioinformatics company uses FastBit to accelerate their molecular docking software by hundreds of times, and an US internet company uses it to sift through terabytes of advertisement related data daily.

He leads the Scientific Data Management research group at LBNL and is on the board of directors at Global Algorithmic Instutite. He is a Distinguished Scientist of ACM and a Senior Member of IEEE.

Education and Training

University of Minnesota, Minneapolis, MN	Computer Science	PhD 1997
University of Wisconsin, Milwaukee, WI	Physics	MS 1990
Nanjing University, Nanjing, China	Physics	BS 1988

Research and Professional Experience

2015 – present	Senior Computer Scientist	Berkeley Lab
2004 – 2015	Staff Computer Scientist	Berkeley Lab
1999 – 2004	Computer System Engineer	Berkeley Lab
1997 – 1999	Postdoctoral Fellow	Berkeley Lab
1996 – 1997	Software Engineer	Guidant Inc

Selected Publications Google Scholar Profile at `http://tinyurl.com/johnspub`

1. U Ayachit, el al. Performance Analysis, Design Considerations, and Applications of Extreme-scale in Situ Infrastructures. In *SC'16*, pp. 79:1–79:12. 2016.
2. B Dong, S Byna, and K Wu. SDS-Sort: Scalable Dynamic Skew-aware Parallel Sorting. In *HPDC'16*, pp. 57–68. 2016.
3. W Zhao, F Rusu, B Dong, and K Wu. Similarity Join over Array Data. In *SIGMOD 2016*, pp. 2007–2022. 2016.
4. S Blanas, K Wu, S Byna, B Dong and A Shoshani. Parallel Data Analysis Directly on Scientific File Formats. In *SIGMOD 2014*, pp. 385–396. 2014.
5. S Byna, J Chou, O Rübel, Prabhat, H Karimabadi, W S Daughton, V Roytershteyn, E W Bethel, M Howison, K-J Hsu, K-W Lin, A Shoshani, A Uselton, and K Wu. Parallel I/O, analysis, and visualization of a trillion particle simulation. *SC'12*, Article # 59, 2012.
6. K Wu, E Otoo, and K Suzuki. Optimizing two-pass connected-component labeling algorithms. *Pattern Analysis & Applications*, 12(2):117–135, 2009.
7. K Wu, E Otoo, and A Shoshani. Optimizing bitmap indices with efficient compression. *ACM Transactions on Database Systems*, 31:1–38, 2006.
8. K Wu and H Simon. Thick-restart Lanczos method for large symmetric eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 22:602–616, 2000.

9. J R Chelikowsky, N Troullier, K Wu and Y Saad. Combining a higher-order finite-difference method with *ab Initio* pseudopotentials: application to diatomic molecule. *Phys. Rev. B*, 50:11355-64, 1994.
10. K Wu, R Savit, and W Brock. Statistical tests for deterministic effects in broad band time series. *Physica D*, 69:172–188, 1993.

Synergistic Activities

- Principal Investigator, "Scientific Data Services – Autonomous Data Management on Exascale Infrastructure," DOE ASCR. LBNL only, $800K/year, 07/2015-06/2018.
- Principal Investigator, "Open Framework for High-Performance Streaming Analytics," LBNL LDRD. Total $135K / year, 10/2015-9/2016.
- Principal Investigator, "International Collaboration Framework for Extreme Scale Experiments (ICEE)," DOE ASCR. Total $517K / year, LBNL $237K / year, 9/2012-8/2015.
- Co-Principal Investigator, "Runtime System for I/O Staging in Support of In Situ Processing on Extreme Scale Data," DOE ASCR. PI Scott Klasky (ORNL). LBNL $206K/year, 09/2013-08/2016.
- Senior Personnel, "Scalable Data Management, Analysis and Visualization Institute," DOE ASCR, PI Arie Shoshani (LBNL), LBNL data management portion $550K/year, 03/2012-02/2017.
- DOE ASCR workshops 2016: NETWORK2025, Streaming2016
- DOE ASCR Workshops 2015: Quantum Computing, Neurocomputing, SDAV, QUEST, ESNet requirments, workflows, FES Integrated Simulations workshop, SciDAC PI meeting, NGNS PI meeting, Experimental and Observational Data workshop, Streaming2015
- Invited presentations in 2015: International Supercomputing Frontiers Conference, Singapore; HPC China Big Data Forum, China
- Conference program committees: BDC 2015; BDAC 2015; ISAV 2015; SSDBM 2016, 2015, 2014, 2013, 2011, 2009

Postdoctoral mentor: Horst D. Simon (LBNL)
PhD thesis advisor: Yousef Saad (U. Minnesota)
Recent collaboraters: H. Abbasi (ORNL), S. Ahern (ORNL), B. Austin (LBNL) W. Bethel (LBNL), S. Byna (LBNL), Y. Byun (Yonsei University, S. Korea), H. Childs (LBNL), J. Choi (Ulsan National Institute of Science and Technology), J.Y. Choi (ORNL), J. Chou (National Tsinghua University, Taiwan), B. Dong (LBNL), A. Choudhary (Northwest), E. Cormier-Michel(LBNL), T. Critchlow (LLNL), C.G.R. Geddes (LBNL), M. Gu (UC Berkeley), W. Gu (UC Berkeley), J. Hahm (KISTI, S. Korea), M. Howison (LBNL), K. Joy (UC Davis), J. Kim (Texas A&M Univ.-Commerce), S. Klasky (ORNL), Q. Koziol (HDF Group), W.S. Ku (Auburn) D. Leinweber (LBNL), K. Lin (National Tsinghua University, Taiwan), Q. Liu (ORNL), J. Lofstead (Sandia), J. Logan (U of Tennessee), B. Ma (UC Berkeley), K.-L. Ma (UC DAvis), V. Markowitz (LBNL), K. Mavrommatis (LBNL) J. Meredith (ORNL), K. Moreland (Sandia), P. Nugent (LBNL), R. Oldfield (Sandia), G. Ostrouchov (ORNL), M. Papka (ANL), M. Parashar (Rutgers), V. Pascucci (U. Utah), N. Podhorszki (ORNL), E. Pourabbas (National Research Council, Roma, Italy), Prabhat (LBNL), J. Qiang (LBNL), O. Rübel (LBNL), A. Romosan (LBNL), D. Rotem (LBNL), N. Samatova (NCSU), K. Schwan (GATech), A. Sim (LBNL), F. Rusu (UC Merced), A. Shoshani (LBNL), R. Tchoua (ORNL), Y. Tian (ORNL), V. Vishwanath (ANL), M. Wehner (LBNL), M. Wolf (GATech), N. Wright (LBNL), I. Yamazaki (UT Knoxville), W. Yu (Auburn).

Appendix 2: Current and Pending Support

James Ahrens
Los Alamos National Laboratory

Current

Project title	Optimizing the Energy Usage and Cognitive Value of Extreme Scale Data Analysis Approaches
Role	Principal Investigator
Source of support	DOE ASCR program
Total award amount	$2,925,000
Total award period covered	1 AUG 2014–31 JUL 2017
Location of project	Los Alamos National Laboratory
Time commitment to project	23%
Project title	SciDAC Scalable Data Management and Analysis and Visualization (SDAV) Institute
Role	Investigator
Source of support	Department of Energy (SC/ASCR)
Total award amount	$2,000,000
Total award period covered	1 OCT 2012–30 SEP 2017
Location of project	Los Alamos National Laboratory
Time commitment to project	1%
Project title	SciDAC Dark Universe
Role	Investigator
Source of support	Department of Energy (SC/ASCR)
Total award amount	$345,000
Total award period covered	1 OCT 2015–30 SEP 2017
Location of project	Los Alamos National Laboratory
Time commitment to project	1%
Project title	ECP ALPINE: Algorithms and Infrastructure for In Situ Visualization and Analysis
Role	Principal Investigator
Source of support	Department of Energy (ECP/Software Technology)
Total award amount	$6,000,000
Total award period covered	15 SEPT 2016-14 SEP 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	10%
Project title	ECP Applications Proposal: Computing the Sky at Extreme Scales
Role	Investigator
Source of support	Department of Energy (ECP/Application)
Total award amount	$1,500,000
Total award period covered	15 SEPT 2016-14 SEP 2020
Location of project	Los Alamos National Laboratory
Time commitment to project	10%

Project title	ECP L3 Visualization and Analysis Lead
Role	Principal Investigator
Source of support	Department of Energy (ECP/NNSA)
Total award amount	$540,000
Total award period covered	October 1, 2016–September 30, 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	25%
Project title	ECP/ATDM Visualization and Analysis R&D project - Cinema
Role	Principal Investigator
Source of support	Department of Energy (ECP/NNSA/ATDM)
Total award amount	$3,600,000
Total award period covered	October 1, 2016–September 30, 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	10%
Project title	Real-time Adaptive Acceleration of Dynamic Experimental Science
Role	Principal Investigator
Source of support	DOE Los Alamos National Laboratory - Laboratory Directed Resarch and Development
Total award amount	$4,950,000
Total award period covered	October 1, 2016–September 30, 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	20%

Pending: Under review

Project title	A Continuously-running, Asynchronous Sampling Engine for the Perceptual and Cognitively-driven Visual Analysis of Massive Scientific Data
Role	Principal Investigator
Source of support	DOE ASCR program
Total award amount	$2,100,000
Total award period covered	1 October 2017–30 September 2020
Location of project	Los Alamos National Laboratory
Time commitment to project	23%
Project title	This proposal
Role	LANL Site PI
Source of support	DOE SCIDAC
Total award amount	$900,000
Total award period covered	September 30, 2017September 30, 2020
Location of project	Argonne National Laboratory
Time commitment to project	5%

Prasanna Balaprakash

Current Support

Award Source	Title	Months /Year	Overall PI	Period	Total Award
DOE ECP	Autotuning Compiler Technology for Cross-Architecture Transformation and Code Generation	3.6	Hall (Univ. Utah)	09/01/12 – 08/31/15	$4M
DOE ECP	CANDLE: Exascale Deep Learning and Simulation Enabled Precision Medicine for Cancer	2.4	Stevens (ANL)	10/1/16 – 9/30/20	$1.5M
DOE ASCR	APE - Benchmarking	1.6	Papka (ANL)	10/01/16 – 09/30/17	$67M
DOE ASCR	Accelerating HEP Science: Inference and Machine Learning at Extreme Scales	1.8	Habib (ANL)	09/17 – 08/22	$1.6M
Bloomberg City of Chicago	Automated Machine Learning Pipeline for City of Chicago's Smart Data Platform	0.6	Leyffer (ANL)	4/1/16 – 10/31/17	$250K

Pending Support

Sponsor	Number	Title	Months /Year	PI	Period	Total Award
NSF	NSF 16-610	S&CC-IRG Preliminary Proposal Track 1: Integrated Sensing, Prediction, and Communication to Prevent Flooding in Vulnerable Urban Communities	1.0	Packman (Northwest-ern)	09/17–08/22	$5.0M
DOE BER ASCR	LAB 17-1682	Application of Machine Learning Algorithms for Developing Fast Accurate High-Resolution Climate Simulations and Analysis	4.2	Kotamarthi (ANL)	09/17–08/20	$2.4M
DOE ASCR	LAB 17-1748	MMEFHeM: A Multiscale Modeling Environment for Flows through Heterocharged Media	1.9	de Pablo (ANL)	09/17–08/21	$6.4M
DOE ASCR	LAB 17-1748	A SciDAC Institute for Computer Science and Data (this proposal)	1.9	Ross (ANL)	09/17–08/20	$18M

David E. Bernholdt – Oak Ridge National Laboratory

Pending Support

Title	***Source***	***Dates***	***Requested Annual Funding***	***Effort***
AToM: Advanced Tokamak Modeling Environment	DOE FES & ASCR (SciDAC)	2017-08-01 2022-07-31	$3,720k	18%
Plasma Surface Interactions: Predicting the Performance and Impact of Dynamic PFC Surfaces	DOE FES & ASCR (SciDAC)	2017-09-01 2022-08-31	$4,413k	28%
Simulation of Fission Gas in Uranium Oxide Nuclear Fuel	DOE NE & ASCR (SciDAC)	2017-09-18 2022-09-17	$1,600k	13%
A SciDAC Institute for Computer Science and Data (current proposal)	DOE ASCR (SciDAC)	2017-09-30 2020-09-29	$6,000k	up to 30%

Current Support

Title	***Source***	***Dates***	***Annual Funding***	***Effort***
Plasma Surface Interactions: Bridging from the Surface to the Micron Frontier through Leadership Class Computing	DOE FES & ASCR (SciDAC)	2012-07-01 2017-08-31	$2,298k	10%
AToM: Advanced Tokamak Modeling	DOE FES & ASCR (SciDAC)	2014-08-01 2017-07-31	$1,250k	10%
Interoperable Design of Extreme-Scale Applications (IDEAS)	DOE ASCR & BER	2014-09-01 2017-08-31	$3,800k	10%
Advancing the Understanding of Fission Gas Behavior and Release in Nuclear Fuel	DOE NE & ASCR (SciDAC)	2016-08-01 2019-02-28	$3,000k (18-month total)	10%
Open MPI for Exascale (OMPI-X)	DOE ECP	2016-10-01 2019-09-30	$1,721k (average)	20%
Scaling OpenMP via LLVM for Exascale Performance and Portability (SOLLVE)	DOE ECP	2016-10-01 2019-09-30	$1,700k (average)	10%
IDEAS-ECP	DOE ECP	2017-01-01 2020-12-31	$2,500k	10%
Oak Ridge Leadership Computing Facility – Programming Environment and Tools	DOE ASCR Facilities	On-going	n/a	10%

Current and Pending Support
E. Wes Bethel, Lawrence Berkeley National Laboratory
July 2017

Current

Title:	Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery		
Sponsor:	DOE/ASCR	Dates: Aug 2014 – Jul 2017	
Annual funding:	$1.35M	Level of effort:	0.50FTE
Role:	PI		

Title:	Towards Exascale: High Performance Visualization and Analytics		
Sponsor:	DOE/ASCR	Dates: Aug 2014 – Jul 2017	
Annual funding:	$750K	Level of effort:	0.50FTE
Role:	PI		

Pending

Title:	Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery		
Sponsor:	DOE/ASCR	Dates: Aug 2017 – Jul 2020	
Annual funding:	$1.8M	Level of effort:	0.40FTE
Role:	PI		

Title:	Scalable Data-Computing Convergence and Scientific Knowledge Discovery		
Sponsor:	DOE/ASCR	Dates: Aug 2017 – Jul 2020	
Annual funding:	$750K	Level of effort:	0.40FTE
Role:	PI		

Title:	Whole Device Modeling at the Center for Effective Research in Fusion		
Sponsor:	DOE/ASCR	Dates: Aug 2017 – Jul 2022	
Annual funding:	$3M	Level of effort:	0.10FTE
Role:	SP		

Title:	Deep Learning for Image Search and Classification		
Sponsor:	Chan-Zuckerberg Foundation	Dates: Oct 2017 – Jul 2019	
Annual funding:	$350K	Level of effort:	0.15FTE
Role:	SP		

Title:	SciDAC Institute for Computer Science and Data (this proposal)		
Sponsor:	DOE/ASCR	Dates: pending	
Annual funding:	$6M	Level of effort:	0.10FTE
Role:	SP		

Currently Funded Projects for Eric Brugger

Funding Agency: DOE, NNSA, ASC IC
Level of effort: 0.85 FTE
Award Period: ongoing

ECP ALPINE: Algorithms and Infrastructure for In Situ Visualization and Analysis
Funding Agency: DOE
Level of effort: 0.15 FTE Award Period: 1/2017–12/2019

If this proposal is funded then his funding profile would become:

Funding Agency: DOE, NNSA, ASC IC
Level of effort: 0.75 FTE
Award Period: ongoing

ECP ALPINE: Algorithms and Infrastructure for In Situ Visualization and Analysis
Funding Agency: DOE
Level of effort: 0.15 FTE Award Period: 1/2017–12/2019

Funding Agency: DOE, SC, OASCR, SciDAC CS (this proposal)
Level of effort: 0.10 FTE
Award Period: FY18-FY20

Aydın Buluç – LBNL

Current Support

- **Project Title:** EDGAR: Energy-efficient Data and Graph Algorithms Research (ECRP)
  Principal Investigator: Aydın Buluç (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $500k/year (#DE-AC02-05CH11231)
  Funding level: 6 months
  Dates: 8/13 – 7/18
- **Project Title:** ExaGraph: Combinatorial Methods for Enabling Exascale Applications (ECP)
  Principal Investigator: Mahantesh Halappanavar (PNNL)
  Funding Agency: DOE/ASCR
  Total Award: $1M/year (#DE-AC02-05CH11231)
  Funding level: 2 months
  Dates: 5/17 – 4/21
- **Project Title:** ExaBiome: Exascale Solutions to Microbiome Analysis (ECP)
  Principal Investigator: Katherine Yelick (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $650K/year (#DE-AC02-05CH11231)
  Funding level: 0.6 months
  Dates: 9/16 – 8/20
- **Project Title:** BINOCULARS
  Principal Investigator: Katherine Yelick (LBNL)
  Funding Agency: DOD/NSA
  Total Award: $2M/year (#EAOCCC9979)
  Funding level: 2.4 months
  Dates: 1/17 – 12/17
- **Project Title:** THOR: Throughput-Oriented Runtime (ECP)
  Principal Investigator: Costin Iancu (LBNL)
  Funding Agency: DOD/NSA
  Total Award: $250K/year (#EAOC0012771)
  Funding level: 1 month
  Dates: 1/17 – 9/17

Pending Support

- **Project Title:** (proposed) Quantum Algorithms, Mathematics and Compilation Tools for Chemical Sciences
  Principal Investigator: Wibe Albert de Jong (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $1.5M/year
  Funding level: 2 months
  Dates: 9/17 – 8/22
- **Project Title:** (proposed) A SciDAC Institute for Computer Science and Data (SciDAC4)
  Principal Investigator: Robert Ross (ANL)
  Funding Agency: DOE/ASCR
  Total Award: $1766K/year
  Funding level: 2.6 months
  Dates: 10/17 – 9/20

Current and Pending Support
Investigator: Alok Choudhary

Current support

Project Title: Managing the Mosaic of Microstructure (Co-PI)
Project Dates: 9/30/12-11/30/17
Sponsor: Carnegie Mellon (originating sponsor: Air Force Research Laboratory)
Award Amount: $1,501,263
Location of Project: Northwestern University
Effort: 0.10 academic months

Project Title: Scalable Data Management, Analysis, and Visualization (SDAV)
Project Dates: 2/15/12-2/14/18
Sponsor: Department of Energy
Award Amount: $749,999
Location of Project: Northwestern University
Effort: 0.10 academic months

Project Title: Characterizing Scalable Data Mining Kernels/Primitives on SMP's
Project Dates: 8/15/03-7/31/18
Sponsor: Intel Corporation
Award Amount: $149,005
Location of Project: Northwestern University
Effort: 0.42 summer months

Project Title: Scalable Analytics and Data Mining for Big Data
Project Dates: 6/1/13-5/31/18
Sponsor: Intel Corporation
Award Amount: $30,000
Location of Project: Northwestern University
Effort: 0.09 academic months

Project Title: Advanced Materials Center for Excellence: Center for Hierarchical Materials Design (ChiMaD) (Co-PI)
Project Dates: 1/1/14-12/31/18
Sponsor: National Institute of Standards and Technology
Award Amount: $25,052,418
Location of Project: Northwestern University
Effort: 0.50 academic months / 0.50 summer months

Project Title: CDS&E: Black Holes in Dense Star Clusters (Co-PI)
Project Dates: 9/1/13-8/31/17
Sponsor: National Science Foundation
Award Amount: $448,377
Location of Project: Northwestern University
Effort: 0.09 academic months

Project Title: Scalable, In-situ Data Clustering data Analysis for Extreme Scale Scientific Computing
Project Dates: 8/1/15-7/31/18
Sponsor: Department of Energy
Award Amount: $1,219,900
Location of Project: Northwestern University
Effort: 1.00 summer month

Project Title: SHF:Medium:Collaborative Research: Scalable Algorithms for Spatio-temporal Data Analysis
Project Dates: 6/1/14-5/31/18
Sponsor: National Science Foundation
Award Amount: $709,342
Location of Project: Northwestern University
Effort : 0.45 academic months

Project Title: Data-driven Discovery of Novel Thermoelectric Materials
Project Dates: 6/22/15-9/21/18
Sponsor: QuesTek Innovations LLC (originating: Space and Naval Warfare Systems Center Pacific)
Award Amount: $1,427,501
Location of Project: Northwestern University
Effort : 0.40 summer months

Pending proposals

Project Title: DMREF: Development of Network Based Data Science Methods for Prediction of High-Electron Mobility Materials
Project Dates: 8/1/17-7/31/21
Sponsor: National Science Foundation
Award Amount: $1,597,831
Location of Project: Northwestern University
Effort: 0.50 summer months

Project Title: SI2-SSE: Parallel Machine Learning Algorithms for High-performance Computing Systems (Co-PI)
Project Dates: 9/1/17-8/31/20
Sponsor: National Science Foundation
Award Amount: $500,000
Location of Project: Northwestern University
Effort: 0.25 summer months

Project Title: Development of A Deep-Learning Based System for Automated Pavement Distress Detection and Classification (Co-PI)
Project Dates: 7/1/17-6/30/18
Sponsor: Transportation Research Board
Award Amount: $149,974
Location of Project: Northwestern University
Effort: 0.50 academic months

Project Title: A SciDAC Institute for Computer Science and Data (this proposal)
Project Dates: 9/30/17-9/29/20
Sponsor: Department of Energy
Award Amount: $360,000
Location of Project: Northwestern University
Effort: 0.20 academic months

Bronis R. de Supinski:

is currently funded as follows:

- 65% ASC CSSE Advanced Technology Systems Support
- 20% ASC ATDM Next Generation Computing Technology
- 10% ASCR SciDAC SUPER Project (ending at end of FY17)
- 5% SOLLVE (ECP ST project)

If this proposal is funded then he would be funded as follows:

- 65% ASC CSSE Advanced Technology Systems Support
- 20% ASC ATDM Next Generation Computing Technology
- 10% ASCR SciDAC IV CS Project (this proposal)
- 5% SOLLVE (ECP ST project)

Current and Pending Support for Anshu Dubey

Current

ACME-SM: A Global Climate Model Software Modernization Surge
Source: DOE Office of Science/BER **Award Period**: 08/01/2016 – 07/31/2019
Award Amount: $3,000,000 **PI**: Andy Salinger
Location: Argonne National Laboratory **Effort:** 4.8 person months/year
Research Description: Improvement of software engineering and verification of sections of ACME codebase.

Block Structured AMR Co-Design Center
Source: DOE Office of Science **Award Period**: 09/01/2016 – 08/31/2020
Award Amount: $1,255,087 **PI**: John Bell
Location: Argonne National Laboratory **Effort:** 2.4 person months/year
Research Description: Development of AMR infrastructure for exascale machines.

Exascale Models of Stellar Explosion: Quintessential Multiphysics Simulation
Source: DOE Office of Science **Award Period**: 09/01/2016 – 08/31/2020
Award Amount: $1,603,736 **PI**: Dan Kasen
Location: Argonne National Laboratory **Effort:** 2.4 person months/year
Research Description: Building an ecosystem of interoperating code modules for stellar explosion simulations

Advancing software productivity of ECP application projects
Source: DOE Office of Science **Award Period**: 02/02/2017 – 09/31/2020
Award Amount: $2,420,000 **PI**: Lois McInnes
Location: Argonne National Laboratory **Effort:** 1.2 person months/year
Research Description: Examining software practices in exascale applications and providing guidance for improving productivity, curate and generate resources for software teams

Code to Code Comparison Project
Source: DOE NNSA **Award Period**: 02/02/2017 – 09/31/2020
Award Amount: $200,000 **PI**: Donald Q. Lamb
Location: Argonne National Laboratory **Effort:** 0.6 person months/year
Research Description: Carry out comparison between two codes that implement AMR and physics solvers in different ways, quantify the pros and cons of both

Pending

Towards Exascale Astrophysics of Mergers and Supernovae
Source: DOE Office of Science **Award Period**: 02/02/2017 – 09/31/2020
Award Amount: $1,087,602 **PI**: Raph Hix
Location: Argonne National Laboratory **Effort:** 0.6 person months/year
Research Description: Simulations of supernovae

A SciDAC Institute for Computer Science and Data
Source: DOE Office of Science **Award Period**: 09/01/2017 – 08/31/2020
Award Amount: $18,000,000 **PI**: Rob Ross
Location: Argonne National Laboratory **Effort:** 3.6 person months/year
Research Description:

A Test Bed for Assessing the Inter-sectoral Cross-Scale Dependences for the Energy-Water Nexus in the Central United States
Source: DOE Office of Science **Award Period**: 10/01/2017 – 09/30/2020
Award Amount: $9,000,000 **PI**: Eugene Yan
Location: Argonne National Laboratory **Effort:** 3.0 person months/year
Research Description:

Berk Geveci

Current

Accelerating Community-Driven Medical Innovation with VTK
Supporting agency: NIBIB - NIH
Effort (Person-Months): 1.2
Total Funding: $2,462,200
Award Period: May 2013 - April 2018

ECP ALpINe & Algorithms and Infrastructure for In Situ Visualization and Analysis Funding Agency: DOE
Effort (Person-Months): 1.2
Total Funding: $930,000
Award Period: September 2016 - September 2019

A Unified Data-Driven Approach for Programming In Situ Analysis and Visualization Supporting agency: DOE
Effort (Person-Months): 1.1
Total Funding: $675,000
Award Period: September 2014 - August 2018

ECP VTK-m: Updating HPC Visualization Software for Exascale-Era Processors
Funding Agency: DOE
Effort (Person-Months): 1.2
Total Funding: $1,050K
Award Period: 9/2016–9/2019

Pending

A SciDAC Institute for Computer Science and Data (this proposal)
Funding Agency: DOE
Effort (Person-Months): 1.2
Total Funding: $318K (for Kitware)
Award Perido: 9/30/17-9/29/20

XVis 2: Analysis and Visualization of the NeXt Generation of Data
Supporting agency: DOE
Effort (Person-Months): 1.2
Total Funding: $1,500,000
Award Period: October 2018 - September 2021

Hanqi Guo

Current and Pending Support for Hanqi Guo

- (Current) Title: CODAR: A Co-design Center for Online Data Analysis and Reduction at the Exascale
  Source: Exascale Computing Project
  Award Amount: $4 M/yr
  Award Period: 10/01/2016 - 09/30/2020
  Location: Argonne National Laboratory
  PI: Ian Foster
  Effort: 40%

- (Current) Title: Catalog: Characterizing Faults, Errors, and Failures in Extreme-Scale Systems
  Source: DOE Office of Science
  Award Amount: $1.125 M/yr
  Award Period: 07/01/2015 - 06/30/2018
  Location: Argonne National Laboratory
  PI: Franck Cappello
  Effort: 50%

- (Current) Title: OSCon—Optimizing Superconductor Transport Properties Through Large-Scale Simulation
  Source: DOE Office of Science
  Award Amount: $760 K/yr
  Award Period: 09/01/2012 - 08/31/2017
  Location: Argonne National Laboratory
  PI: Andreas Glatz
  Effort: 10%

- (Pending) Title: SciDAC4 CS Institute
  Source: DOE Office of Science
  Award Amount: $6 M/yr
  Award Period: 10/01/2017 - 9/30/2020
  Location: Argonne National Laboratory
  PI: Rob Ross
  Effort: 30%

Overlap: If projects are fully funded, then effort will be adjusted accordingly.

Pascal Grosset
Los Alamos National Laboratory

Current

Project title	ECP Applications Proposal: Computing the Sky at Extreme Scales
Role	Investigator
Source of support	Department of Energy (ECP/Application)
Total award amount	$1,500,000
Total award period covered	15 SEPT 2016-14 SEP 2020
Location of project	Los Alamos National Laboratory
Time commitment to project	50%
Project title	LANL ASC Advanced Technology Development and Mitigation: Next-Generation Code project (NGC)
Role	Investigator
Source of support	Department of Energy (ECP/Application)
Total award amount	$9,000,000
Total award period covered	15 SEPT 2014-14 SEP 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	50%

Pending: Under review

Project title	ECP Applications Proposal: Computing the Sky at Extreme Scales
Role	Investigator
Source of support	Department of Energy (ECP/Application)
Total award amount	$1,500,000
Total award period covered	15 SEPT 2016-14 SEP 2020
Location of project	Los Alamos National Laboratory
Time commitment to project	35%
Project title	LANL ASC Advanced Technology Development and Mitigation: Next-Generation Code project (NGC)
Role	Investigator
Source of support	Department of Energy (ECP/Application)
Total award amount	$9,000,000
Total award period covered	15 SEPT 2014-14 SEP 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	35%
Project title	This proposal
Role	Investigator
Source of support	DOE SCIDAC
Total award amount	$900,000
Total award period covered	September 30, 2017September 30, 2020
Location of project	Argonne National Laboratory
Time commitment to project	30%

Current and Pending Support for Mary Hall

University of Utah

Current Support

ECP ST: AUTOTUNING COMPILER TECHNOLOGY FOR CROSS-ARCHITECTURE TRANSFORMATION AND CODE GENERATION, PI: Mary Hall (Utah), co-PIs Sam Williams (LBNL) and Paul Hovland (ANL), DOE/ASCR, $280K/yr, 1 summer month, 03/17-02/20.

SHF: MEDIUM: COLLABORATIVE RESEARCH: AN INSPECTOR/EXECUTOR COMPILATION FRAMEWORK FOR IRREGULAR APPLICATIONS, PI: Mary Hall (Utah), co-PIs Michelle Strout (Arizona) and Cathie Olschanowsky (Boise), NSF, $100K/yr, 1 summer month, 09/16—02/20.

EAGER: APPLICATION-DRIVEN DATA PRECISION SELECTION METHODS, PI: Ganesh Gopalakrishnan (Utah), NSF, $25K/year, 0.25 summer month, 09/16 - 8/18.

USING ACTIVE HARMONY AND CHILL TO AUTOTUNE CHAPEL, PI: Jeff Hollingsworth (UMD), DoD, $70K/yr, 0.25 summer month, 01/17—09/17.

Pending Support

SciDAC 4, co-PIs: Rob Ross (ANL) and Leonid Oliker (LBNL), DOE, $110K/yr, 1 summer month, 09/17-08/21.

SHF: MEDIUM: HIERARCHICAL TUNING OF FLOATING-POINT COMPUTATIONS, PI: Ganesh Gopalakrishnan (Utah), co-PIs Zvonomir Rakamaric and Hari Sundar, NSF, $100K/year, 0.5 summer month, 08/17 -- 07/20.

Current and Pending Support – Neil Fortner

Current Support

Source of Funding:	Department of Energy (DOE)
Award Number:	DE-SC0016474
Project Title:	A Software Defined Approach to Exascale Storage Services
Total Award Amount:	$674,972.48
Starting Date:	05/15/2016
Ending Date:	05/14/2019
Person Months (Calendar):	2.8

Remaining support is currently being provided by HDF Group operational funds.

Pending Support

Source of Funding:	Department of Energy (DOE)
Proposal Title:	Alliance for Application Performance at Extreme Scale
Award Amount (Requested):	$1,927,144
Starting Date (Anticipated):	01/01/2018
Ending Date (Anticipated):	09/30/2020
Person Months (Calendar):	9.7

Overlap: If pending projects are fully funded, then effort will be adjusted accordingly.

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.

Investigator: **Kevin Harms**	Other agencies (including NSF) to which this proposal has been/will be submitted

Support: ☒ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title: Argonne Leadership Computing Facility (Papka-PI)

Source of Support: Department of Energy - ASCR

Total Award Amount: $80,000,000 Total Award Period Covered: 10/01/2016—09/30/2017

Location of Project: Argonne National Laboratory

Person-Months Per Year Committed to the Project. Cal: 7.8 Acad: Sumr:

Support: ☒ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title: Novel Storage Architectures (PI: Robert B. Ross)

Source of Support: WFO/Los Alamos National Laboratory

Total Award Amount: $1,000K Total Award Period Covered: 06/01/2013 – 09/30/2016

Location of Project: Argonne National Laboratory

Person-Months Per Year Committed to the Project. Cal: 2.4 Acad: Sumr:

Support: ☐ Current ☒ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title: Application Based Analysis to Characterize systems with User defined performance metricS (ABACUS).

Source of Support: National Science Foundation (NSF)

Total Award Amount: $500,000 Total Award Period Covered: 08/01/2015 – 07/31/2017

Location of Project: Argonne National Laboratory

Person-Months Per Year Committed to the Project. Cal: 1.8 Acad: Sumr:

Support: ☐ Current ☒ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title: A SciDAC Institute for Computer Science and Data – (PI: Robert B. Ross)

Source of Support: DOE-ASCR

Total Award Amount: $18,272,065 Total Award Period Covered: 09/30/2017 – 09/29/2020

Location of Project: Argonne National Laboratory

Person-Months Per Year Committed to the Project. Cal: 1.2 Acad: Sumr:

Support: ☐ Current ☐ Pending ☐ Submission Planned in Near Future ☐ *Transfer of Support
Project/Proposal Title:

Source of Support:

Total Award Amount: Total Award Period Covered:

Location of Project:

Person-Months Per Year Committed to the Project. Cal: Acad: Sumr:

The Principal Investigator has multiple funding sources. Should there be an overlap, effort levels will be adjusted Accordingly.

Current & Pending
Paul Hovland

Current

Organization Providing Support	DOE-ASCR
Award Number	56703/PRJ1001288
Title of Award	Automatic Differentiation for Extreme-scale Analysis and Machine Learning
Project Principal Investigator	Paul Hovland
Award Period	10/01/2015 – 09/30/2018
Total Amount for Entire Award Period	$1,315,000
Number of Person-months per Year	1.8

Organization Providing Support	DOE-ASCR
Award Number	57L18/PRJ1001325
Title of Award	Program Verification for Extreme-scale Applications
Project Principal Investigator	Paul Hovland
Award Period	09/01/2014 – 08/31/2017
Total Amount for Entire Award Period	$1,216,000
Number of Person-months per Year	2.1

Organization Providing Support	DOE-ASCR
Award Number	57K30/PRJ1001313
Title of Award	Institute for Sustained Performance, Energy, and Resilience (SUPER)
Project Principal Investigator	Paul Hovland
Award Period	08/01/2011 – 03/31/2017
Total Amount for Entire Award Period	$2,885,000
Number of Person-months per Year	0.3

Organization Providing Support	DOE-ASCR
Award Number	82725/PRJ1003781
Title of Award	Autotuning Compilier Technology for Cross-Architecture Transformation and Code Generation (Y-Tune)
Project Principal Investigator	Paul Hovland
Award Period	01/25/2017 – 09/30/2019
Total Amount for Entire Award Period	$1,483,632
Number of Person-months per Year	0.6

Note: Effort allocated to programmatic overhead / Laboratory LDRD projects is not included.
Overlap: If projects are fully funded, effort will be adjusted accordingly.

Pending

Organization Providing Support	DOE-ASCR
Award Number	57L98
Title of Award	MACSER: Multifaceted Mathematics for Rare, Extreme Events in Complex Energy and Environment Systems
Project Principal Investigator	Mihai Anitescu
Award Period	10/01/2017 – 09/30/2022
Total Amount for Entire Award Period	$18,957,131
Number of Person-months per Year	4.0

Organization Providing Support	DOE-ASCR
Award Number	57L94
Title of Award	HPC framework for event generation at colliders
Project Principal Investigator	Hoche (SLAC); Childers (ANL)
Award Period	07/19/2017 – 07/18/2020
Total Amount for Entire Award Period	$312,088
Number of Person-months per Year	0.5

Organization Providing Support	DOE-ASCR/BES
Award Number	TBD
Title of Award	High-Accuracy Quantum Approaches for Predictions of Catalysis on Solid Surfaces (HAQ-CAT)
Project Principal Investigator	Maria Chan
Award Period	09/30/2017 – 09/29/2021
Total Amount for Entire Award Period	$7,432,681
Number of Person-months per Year	1.0

Organization Providing Support	DOE-ASCR
Award Number	TBD
Title of Award	A SciDAC Institute for Computer Science and Data
Project Principal Investigator	Rob Ross
Award Period	09/01/2017 – 08/31/2020
Total Amount for Entire Award Period	$18,545,400
Number of Person-months per Year	1.2

Organization Providing Support	DOE-ASCR
Award Number	57L97
Title of Award	Mathematics of Multiscale Control and Design (MMuSCaD)
Project Principal Investigator	Carol S. Woodward
Award Period	10/01/2017 – 09/30/2022
Total Amount for Entire Award Period	$17,500,000
Number of Person-months per Year	2.6

Note: Effort allocated to programmatic overhead / Laboratory LDRD projects is not included.
Overlap: If projects are fully funded, effort will be adjusted accordingly.

Current and Pending Support for Kevin Huck

Current Support

Sponsor	Proposal/Project Title	Project dates		Total Award	Annual Commitment
		start	end		
ORNL	PROTEAS: Programming Toolchain for Emerging Architectures and Systems	5/1/17	8/31/19	$834,309	3.0 cal months

Pending Support

Sponsor	Proposal/Project Title	Project dates		Total Award	Annual Commitment
		start	end		
NSF	Collaborative Research: Phylanx: Python-based Array Processing in HPX	7/1/17	6/30/18	$93,300	2.6 cal months
ORNL	Adaptive near Real Time Infrastructure for Science aT Scale (ARTISTS)	9/1/17	8/31/20	$375,000	3.0 cal months
ORNL	A SciDAC Institute for Computer Science and	9/30/17	9/30/20	$6,000,000	3.0 cal months

Costin Iancu – LBNL

Current Support

- **Project Title:** Binoculars: Berkeley Institute for Compilers, Languages and Runtimes,
  Principal Investigator: Katherine Yelick (LBNL)
  Funding Agency: DoD
  Total Award: $1.5M/year
  Funding level: 8 months
  Dates: 10/16 – 10/17
- **Project Title:** THOR: Throughput Oriented Runtime Design
  Principal Investigator: Costin Iancu
  Funding Agency: DoD
  Total Award: $500K/year
  Funding level: 4 months
  Dates: 10/16 – 9/17

Pending Support

- **Project Title:** Quantum Algorithms, Mathematics and Compilation Tools for Chemical Sciences
  Principal Investigator: Wibe de Jong (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $1491K/year
  Funding level: 2 months
  Dates: 10/17 – 9/20
- **Project Title:** (proposed) A SciDAC Institute for Computer Science and Data (SciDAC4)
  Principal Investigator: Robert Ross (ANL)
  Funding Agency: DOE/ASCR
  Total Award: $1766K/year
  Funding level: 2 months
  Dates: 10/17 – 9/20
- **Project Title:** Binoculars: Berkeley Institute for Compilers, Languages and Runtimes,
  Principal Investigator: Katherine Yelick (LBNL)
  Funding Agency: DoD
  Total Award: $1.5M/year
  Funding level: 4 months
  Dates: 10/17 – 09/18
- **Project Title:** THOR: Throughput Oriented Runtime Design
  Principal Investigator: Costin Iancu
  Funding Agency: DoD
  Total Award: $500K/year
  Funding level: 4 months
  Dates: 10/17 – 9/18
- **Project Title:** (proposed) A SciDAC Institute for Computer Science and Data (SciDAC4)
  Principal Investigator: Robert Ross (ANL)
  Funding Agency: DOE/ASCR
  Total Award: $1766K/year
  Funding level: 2.4 months
  Dates: 10/17 – 9/20

Current and Pending Support for Robert Latham

Current

Data Libraries and Services Enabling Exascale Science (FWP #82733/PRJ1003785)
Source: DOE-ASCR Award Period: 01/26/2017-09/30/2019
Award Amount (to ANL): $2,695,180 PI: Robert B. Ross
Location: Argonne National Laboratory Effort: 10.8 person-months/year (90%)
Research Description: Development and productization of storage and I/O system software for exascale computing.

Exascale MPI (FWP #82729/PRJ1003780)
Source: DOE-ASCR Award Period: 01/27/2017-09/30/2019
Award Amount (to ANL): $5,130,900 PI: Pavan Balaji
Location: Argonne National Laboratory Effort: 1.2 person-months/year (10%)
Research Description: Development of MPI software and specification for Exascale.

Pending

A SciDAC Institute for Computer Science and Data (Pending)
Source: DOE-ASCR Award Period: 09/01/2017-08/31/2020
Award Amount (to ANL): $1.5 M/year PI: Robert B. Ross
Location: Argonne National Laboratory Effort: 3.0 person-months/year (25%)
Research Description: Development of I/O tools and solutions for SciDAC applications

Overlap: If projects are fully funded, then effort will be adjusted accordingly.

Chunhua Liao:

is currently funded as follows:

- 35% DOE ECP
- 25% LLNL/LDRD Detecting Data-Races in High-Performance Computing
- 25% ASC ATDM Next Generation Computing Technology
- 15% LLNL/ROSE Techbase

If this proposal is funded then he would be funded as follows:

- 22.5% DOE ECP
- 25% LLNL/LDRD Detecting Data-Races in High-Performance Computing
- 25% ASC ATDM Next Generation Computing Technology
- 15% LLNL/ROSE Techbase
- 12.5% ASCR SciDAC IV CS (this proposal)

Current & Pending Support
Scott Klasky
Oak Ridge National Laboratory

Current: (Amount is per year)

Title: Science-driven Data Management for multi-tiered storage (SIRIUS)
Project Number: ERKJ311
Role: lead PI
Amount: $1,650,000
Location: Oak Ridge National Laboratory
Sponsor: Department of Energy ASCR
Award Period: 9/1/15 – 9/30/18
Major goals of this project: Understand the nature of storage across a multi-tier storage architecture, and develop new techniques to refactor data.
Person Months: 1.0

Title: A Co-Design Center for Online Data Analysis and Reduction at the Exascale
Project Number:
Role: Co PI
Amount: $4,000,000
Location: Oak Ridge National Laboratory
Sponsor: ECP: Department of Energy ASCR
Award Period: 9/1/16 – 9/30/20
Major goals of this project: To co-design scientific software for the exascale project focusing on the tradeoffs between accuracy and performance for on-line analysis.
Person Months: 2.0

Title: The ADIOS framework for Scientific Data on exascale systems
Project Number:
Role: Lead PI
Amount: $1,500,000
Location: Oak Ridge National Laboratory
Sponsor: ECP: Department of Energy ASCR
Award Period: 9/1/16 – 9/30/20
Major goals of this project: Develop the next-generation version of ADIOS and deliver this to the exascale community
Person Months: 2.0

Title: High-Fidelity Whole Device Modeling of Magnetically Confined Fusion Plasmas
Project Number:
Role: Co-PI
Amount: $2,500,000
PI: Amitava Bhattacharje (PPPL)
Location: Oak Ridge National Laboratory
Sponsor: ECP: Department of Energy ASCR
Award Period: 9/1/16 – 9/30/20
Major goals of this project: To develop a framework to couple multiple physics codes together to understand instabilities in a fusion reactor.
Person Months: 2.0

Title: RSVP: Runtime System for Voluminous Data Processing
Project Number: ERKJU60

Role: Lead PI
Amount: $700,000
Location: Oak Ridge National Laboratory
Sponsor: Department of Energy FES/ASCR
Award Period: BASE Funding, - 2017
Major goals of this project: To create a runtime and plug-ins for in situ processing of data.
Person Months: 2.4

Pending

Title: Partnership Center for High-Fidelity Boundary Plasma Simulation
ORNL Amount: $470,000
P.I.: C. S. Chang (PPPL)
Location: Oak Ridge National Laboratory
Sponsor: Department of Energy FES/ASCR
Award Period: Pending
Person Months: 1.0

Title: ISEP: Integrated Simulation of Energetic Particles in Burning Plasma
Amount: $2,100,000
P.I.: Zhihong Lin
Location: Oak Ridge National Laboratory
Sponsor: Department of Energy FES/ASCR
Award Period: Pending
Person Months: 0.6

Title: ARTISTS: Adaptive near Real Time Infrastructure for Science aT Scale
Amount: $700,000
Location: Oak Ridge National Laboratory
Sponsor: Department of Energy FES/ASCR
Award Period: Pending
Person Months: 2.4

Title: RAPIDS
Amount: $6,000,000
Location: Oak Ridge National Laboratory
Lead PI: R. Ross (ANL)
Sponsor: Department of Energy ASCR
Award Period: Pending
Person Months: 1.25

Seyong Lee — ORNL

Current Support

Title	*Source*	*Dates*	*Current Annual Funding*	*Effort*
ARES: Abstract Representations for the Extreme-Scale Software Stack	DOE	7/14–6/17	$1.0M	30%
ECP PROTEAS: Programming Systems for Emerging Architectures	DOE	9/16–8/19	$1.5M	50%
ECP ExaSMR: Coupled Monte Carlo Neutronics and Fluid Flow Simulation of Small Modular Reactors	DOE	9/16–8/19		20%

Pending Support

Title	*Source*	*Dates*	*Requested Annual Funding*	*Effort*
SciDAC Institute for Computer Science and Data (this proposal)	DOE	10/17–9/20	$18M	30%

Note: Effort level will be adjusted if pending proposals are funded.

Current and Pending Support
Investigator: Wei-keng Liao

CURRENT

Support: Current (Co-PI)
Proposal Title: Dynamical Modeling of Dense Star Clusters with a Parallel Monte Carlo Code
Project Dates: 9/1/14-8/31/17
Source of Support: NASA Goddard Space Flight Center
Award Amount: $455,666
Location of Project: Northwestern University
Person-Months Per Year Committed to the Project: 0.25 calendar months

Support: Current (Co-PI)
Project Title: CDS&E: Black Holes in Dense Star Clusters
Project Dates: 9/1/13-8/31/17
Source of Support: National Science Foundation
Award Amount: $448,377
Location of Project: Northwestern University
Person-Months Per Year Committed to the Project: 0.48 calendar months

Support: Current (Co-PI)
Project Title: SHF:Medium:Collaborative Research: Scalable Algorithms for Spatio-temporal Data Analysis
Project Dates: 6/1/14-5/31/18
Source of Support: National Science Foundation
Award Amount: $709,342
Location of Project: Northwestern University
Person-Months Per Year Committed to the Project: 0.96 calendar months

Support: Current (Co-PI)
Project Title: Scalable, In-situ Data Clustering Data Analysis for Extreme Scale Scientific Computing
Project Dates: 8/1/15-7/31/18
Source of Support: Department of Energy
Award Amount: $1,219,900
Location of Project: Northwestern University
Person-Months Per Year Committed to the Project: 3.00 calendar months

PENDING

Support: Pending (Co-PI)
Project Title: MRI: Acquisition of a High-Performance Computing Cluster to Unveil the Sources of Gravitational Waves
Project Dates: 8/1/17-7/31/18
Source of Support: National Science Foundation
Award Amount: $512,704
Location of Project: Northwestern University
Person-Months Per Year Committed to the Project: 0.00 calendar months

Support: Pending (PI)
Project Title: SI2-SSE: Parallel Machine Learning Algorithms for High-performance Computing Systems
Project Dates: 9/1/17-8/31/20
Source of Support: National Science Foundation
Award Amount: $500,000
Location of Project: Northwestern University
Person-Months Per Year Committed to the Project: 1.00 calendar months

Support: Pending
Project Title: A SciDAC Institute for Computer Science and Data (this proposal)
Project Dates: 9/30/17-9/29/20
Source of Support: Department of Energy
Award Amount: $360,000
Location of Project: Northwestern University
Person-Months Per Year Committed to the Project: 1.50 calendar months

Current and Pending Support for Allen Malony

Current Support

Sponsor	Proposal/Project Title	Project dates		Total Award	Annual Commitment (months)
		start	end		
DOE	Collaborative Research: Performance Retargeting of Instrumentation, Measurement, and Analysis Technologies for Exascale Computing: the PRIMA-X Project (with S Shende, F Wolf & B Mohr)	9/1/13	12/31/17	$450,000	0.9 Acad
NSF	CNH-RCN: Forest Governance and Climate Change in Driving Native Insect Outbreaks (w/Bone, Bartlein,Gavin, Moseley&Davis)	8/1/14	7/31/18	$1,360,000	0.1 Acad
DOE	Performance Understanding and Analysis for Exascale Data Management Workflows (with Eisenhauer & Abassi)	9/1/14	8/31/18	$382,500	0.1 Acad
NSF	SI2-SSI: Collaborative Research: A Software Infrastructure for MPI Performance Engineering: Integrating MVAPICH and TAU via the MPI Tools Interface (w/Shende & D Panda)	9/1/15	8/31/19	$1,200,000	0.5 Sum
NSF	University of Oregon Planning Proposal: I/UCRC for Big Learning (w/Dou, Fickas & Lowd)	2/1/17	1/31/18	$15,000	0.01 Acad
ORNL	PROTEAS: Programming Toolchain for Emerging Architectures and Systems	5/1/17	9/30/19	$834,309	1.0 Sum

Pending Support

Sponsor	Proposal/Project Title	Project dates		Total Award	Annual Commitment (months)
		start	end		
NSF	Collaborative Research: Phylanx: Python-based Array Processing in HPX	7/1/17	6/30/18	$93,300	0.3 Sum
NSF	Phase 1 IUCRC University of Oregon Center for Big Learning (w/Dou & Sventek)	1/1/18	12/31/22	$749,998	.44 Sum
ORNL	A SciDAC Institute for Computer Science and	9/1/17	8/31/20	$6,000,000	.5 Sum

Current and Pending Support
Kshitij Mehta
Oak Ridge National Laboratory

Current:

1. **Title: A Co-Design Center for Online Data Analysis and Reduction at the Exascale**
   Amount: $4,000,000 **Location:** Oak Ridge National Laboratory
   Sponsor: ECP: Department of Energy ASCR
   Award Period: 9/1/16 – 9/30/20
   Person Months: 6.0

2. **Title: Project Name: Oak Ridge Leadership Class Facility (PI/ Project Manager: James J. Hack)**
   Amount: > $100,000,000 **Location:** Oak Ridge National Laboratory
   Sponsor: Department of Energy, ASCR
   Award Period: 10/1/16 - 9/30/17
   Person Months: 6.0

Pending:

1. **Project Name: A SciDAC Institute for Computer Science and Data (This proposal)**
   Amount: $6,000,000
   Location: Oak Ridge National Laboratory
   Funding Source: Department of Energy FES/ASCR
   Lead PI: R. Ross (ANL)
   Period covered: 9/30/17 – 9/30/20
   Person Months: 6.0

2. **Title: ARTISTS**
   Amount: $2,100,000 + $675,000 **Location**: Oak Ridge National Laboratory
   Sponsor: DOE ASCR
   Award Period: 09/01/2017 – 08/31/2020
   Person Months: 3.0

Kathryn Mohror:

is currently funded as follows:

- 20% ASC Common Computing Environment
- 25% ASC ATDM Next Generation Computing Enablement
- 5% LLNL LDRD on Programming Models for MPI Fault Tolerance
- 10% ASCR Veracity
- 15% ECP UnifyCR
- 15% ECP VeloC
- 10% ASCR SciDAC SUPER (ending at end of FY17)

If this proposal is funded then she would be funded as follows:

- 20% ASC Common Computing Environment
- 25% ASC ATDM Next Generation Computing Enablement
- 5% LLNL LDRD on Programming Models for MPI Fault Tolerance
- 10% ASCR Veracity
- 15% ECP UnifyCR
- 15% ECP VeloC
- 10% ASCR SciDAC IV CS (this proposal)

Shirley V. Moore (ORNL)

Current
(1) Sponsor: DOE ASCR
Award Number:
Title: ECP Application Assessment
Principal Investigator: Kenneth Roche (PNNL)
Total award amount (including direct and indirect costs): $1.3M/year
Number of Person-months per year effort: 4.8
Dates: 7/17 – 9/19

(2) Sponsor: DOE ASCR
Award Number: FWP # 3ECP4210
Title: ECP Design Space Evaluation
Principal Investigator: John Shalf (LBNL)
Total award amount (including direct and indirect costs): $4,022,595
Number of Person-months per year effort: 2.4
Dates: 4/17 – 3/20

(3) Sponsor: ORNL General Hire
Number of Person-months per year effort: 4.8
Dates: 6/16-

(4) **Sponsor**: **DOE ASCR**
Award Number:
Title: Enabling GAMESS for Exascale Computing in Chemistry & Materials
Principal Investigator: Mark Gordon (Ames Lab)
Total award amount: $3.8M over 3 years
Number of Person-months per year effort: 2.4
Dates: (ORNL funding starts 10/17)

Pending (Person-months of effort will be adjusted with new hires as needed)
(1) Sponsor: DOE ASCR
Title: SciDAC Institute for Computer Science and Data
Total amount for entire award period (including indirect costs): $18M over 3 years
Number of Person-months per year effort: 2.4

(2) Sponsor: ORNL LDRD
Title: Application Characterization and Mapping for Abstract Exascale Machine Models
Principal Investigator: Shirley V. Moore
Total amount for entire award period (including indirect costs): $800K over 2.5 years
Number of Person-months per year effort: 4.2

(3): Sponsor: DOE ASCR
Title: Methods and Interfaces for Quantum Acceleration of Scientific Applications (MIQASA)
Principal Investigator: Raphael Pooser
Total award amount: $17.5M over five years.
Number of Person-months per year effort: 5
Dates: 9/17-9/22

(4) Sponsor: AFOSR (subward to ORNL through Strategic Partnerships)
Title: Advancing Electronic Structure Theory Codes to the Exascale
Principal Investigator: Mark Gordon (Iowa State University/Ames Lab)
Total award amount for entire award period (including direct and indirect costs): $3M over 3 years
Number of Person-months per year effort: 2.4
Dates: 12/17-11/20

Current and Pending Funding

Dmitriy Morozov
Lawrence Berkeley National Laboratory

July 2017

Current Projects

1. Funding Organization: DOE/SC/ASCR
   Title: Towards Exascale: High Performance Visualization and Analytics.
   Award Amount: $750K/yr
   Award Period: August 2014 – July 2017
   Location: Lawrence Berkeley National Laboratory
   PI: E. Wes Bethel
   Effort: 20%

2. Funding Organization: DOE/SC/ASCR
   Title: Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery.
   Award Amount: $1.5M/yr
   Award Period: August 2014 – July 2017
   Location: Lawrence Berkeley National Laboratory
   PI: E. Wes Bethel
   Effort: 30%

3. Funding Organization: UC Berkeley.
   Title: Berkeley Institute for Data Science Fellowship.
   Award Period: September 2016 – August 2018
   Location: Lawrence Berkeley National Laboratory
   PI: Dmitriy Morozov
   Effort: 50%

Pending Proposals

1. Funding Organization: DOE/SC/ASCR and DOE/SC/FES
   Title: Whole Device Modeling at the Center for Effective Research in Fusion.
   Award Amount: $3M/yr
   Award Period: August 2017 – July 2022
   Location: Lawrence Berkeley National Laboratory
   PI: Alice Koniges
   Effort: 40%

2. Funding Organization: DOE/SC/ASCR and DOE/SC/HEP
   Title: Accelerating HEP Science: Inference and Machine Learning at Extreme Scales.
   Award Amount: $6.6M/yr
   Award Period: August 2017 – July 2022
   Location: Argonne National Laboratory
   PI: Salman Habib
   Effort: 10%

3. Funding Organization: DOE/SC/ASCR.
   Project Title: Scalable Analysis Methods and *In Situ* Infrastructure for Extreme Scale Knowledge Discovery
   Location: Lawrence Berkeley National Laboratory
   PI: E. Wes Bethel

Dates of award: August 2017—July 2020.
Award amount: $2.25M.
Level of effort: 30%

4. Funding Organization: DOE/SC/ASCR.
   Project Title: Scalable Data-Computing Convergence and Scientific Knowledge Discovery
   Location: Lawrence Berkeley National Laboratory
   PI: E. Wes Bethel
   Dates of award: August 2017—July 2020.
   Award amount: $750K/yr.
   Level of effort: 20%

5. Funding Organization: DOE/SC/ASCR.
   Project Title: Center for Integrated Multiscale Model Reduction Computational Capabilities
   PI: Panayot S. Vassilevski
   Location: Lawrence Livermore National Laboratory
   Dates of award: October 2017—September 2020.
   Award amount: $3.5M/yr.
   Level of effort: 15%

6. Funding Organization: DOE/SC/ASCR.
   Project Title: SciDAC CDI
   PI: Robert Ross
   Location: Argonne National Laboratory
   Dates of award: August 2017—July 2020.
   Award amount: $6M/yr.
   Level of effort: 33%

Current and Pending Support for Boyana Norris
(Effort will be adjusted if the pending proposals are funded.)

Current Support

Sponsor	Number	Title	Months Effort	PI	Period	Total Award
ANL Sub.	DOE ASCR	Performance analysis, modeling, and optimization of DOE scientific applications	2.16	Norris (UO)	10/13–09/17	$219k
RNET Technologies	DOE SB163-007	Automated Solver Selection for Nuclear Engineering Simulations	1.8	Norris (UO)	08/16–07/18	$250K
NSF	CISE	EAGER: Collaborative Research: Lighthouse: A User-Centered Web System for High-Performance Software Development	0.76	Norris (UO)	10/15–09/17	$150K
DOE ECP	ECP AD	Advancing Software Productivity for Exascale Applications: IDEAS-ECP	3	Heroux (SNL), McInnes (ANL)	1/17–12/20	$10M
DOE ASCR	LAB17-1698	Nuclear Low Energy Initiative (NUCLEI)	0.1	Carlson (LANL)	09/17–08/22	$10.0M
DOE ASCR	LAB17-1697	HEP Event Reconstruction with Cutting Edge Computing Architectures	0.1	Cerati (Fermilab)	09/17–08/20	$950k

Pending Support

Sponsor	Number	Title	Months Effort	PI	Period	Total Award
NSF SHF	NSF 16-578	Title: SHF: Small: Collaborative Research: Automated Numerical Solver EnviRonment (ANSER)	0.75	Norris (UO)	09/17–08/20	$250K
NSF SPX	NSF 16-605	Title: SPX: Collaborative Research: SANDY: Sparsification-based Approach for Analyzing Network Dynamics	0.2	Das (Missouri S&T)	09/17–08/20	$763K
DOE OFES-ASCR	LAB17-1670	Integrated Simulation Tools for Whole Device Modeling during Disruptions and Transient Events in Tokamaks	0.13	Hassanein (Purdue)	08/17–07/22	$7M
DOE ASCR	LAB17-1787	A SciDAC Institute for Computer Science and Data *(this project)*	0.1	Ross (ANL)	09/17–08/20	$6.0M

Leonid Oliker – LBNL

Current Support

- **Project Title:** Science Driven Computer Architecture)
  Principal Investigator: Sudip Dosanjh (LBNL)
  Funding Agency: DOE/ASCR
  Total Award:
  Funding level: 2.4 months
  Dates: 8/08 – ongoing
  Collaboration with NERSC to integrate research activities into facilities environment. No overlap

- **Project Title:** ECP Application Assessment Solutions for Microbiome Analysis (ECP)
  Principal Investigator: Kenneth Roche (PNNL)
  Funding Agency: DOE/ASCR
  Total Award: $1.3M/year
  Funding level: 2.4 months
  Dates: 7/17 – 9/19

- **Project Title:** Exascale Solutions for Microbiome Analysis (ECP)
  Principal Investigator: Katherine Yelick (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $650K/year
  Funding level: 1.2 months
  Dates: 10/16 – 9/19

- **Project Title:** Design Space Evaluation (ECP)
  Principal Investigator: David Donofrio (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $300K/year (#DE-AC02-05CH11231)
  Funding level: 2.4 months
  Dates: 4/17 – 3/20

Pending Support

- **Project Title:** (proposed) A SciDAC Institute for Computer Science and Data (This proposal)
  Principal Investigator: Rob Ross (ANL)
  Funding Agency: DOE/ASCR
  Total Award: $6M/year
  Funding level: 2.4 months
  Dates: 9/17 – 9/20

- **Project Title:** (proposed) A Calibrated Global Cloud Resolving Model designed for Extreme Climate Simulation (SciDAC4)
  Principal Investigator: William Collins (LBNL)
  Funding Agency: DOE/ASCR,BER

Total Award: $818K/year
Funding level: 3.6 months
Dates: 9/17 – 2/20

- **Project Title:** (proposed) Boundary Plasma Advanced Computation (BPAC) Center (SciDAC4)
  Principal Investigator: Xueqiao Xu (LLNL)
  Funding Agency: DOE/ASCR,FES
  Total Award: $300K/year
  Funding level: 1.2 months
  Dates: 9/17 – 2/20

- **Project Title:** (proposed) High Performance Monte Carlo for Physics (SciDAC4)
  Principal Investigator: Craig Tull (LBNL)
  Funding Agency: DOE/ASCR,HEP
  Total Award: $750K/year
  Funding level: 1.2 months
  Dates: 9/17 – 2/20

John Patchett
Los Alamos National Laboratory

Current

Project title	LANL ASC CSSE
Role	Production Visualization Lead
Source of support	Department of Energy
Total award amount	$9,000,000
Total award period covered	15 SEPT 2014-14 SEP 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	100%

Pending: Under review

Project title	LANL ASC CSSE
Role	Investigator
Source of support	Department of Energy
Total award amount	$9,000,000
Total award period covered	15 SEPT 2014-14 SEP 2019
Location of project	Los Alamos National Laboratory
Time commitment to project	75%
Project title	This proposal
Role	Investigator
Source of support	DOE SCIDAC
Total award amount	$900,000
Total award period covered	September 30, 2017September 30, 2020
Location of project	Argonne National Laboratory
Time commitment to project	25%

Investigator: Dr. Manish Parashar									
Proposal Title	**Funding Agent**	**Award Start**	**Award End**	**Award Total**	**Support**	**Sumr**	**CY**	**AY**	**Location of Project**
BIG DATA: Collaborative Research: IA: F: Fractured Subsurface Characterization using High Performance Computing & Guided by Big Data **Award Number: IIS 1546145**	National Science Foundation	1/1/16	12/31/18	$314,574	Current	0.25	0	0	Rutgers University
The Rutgers component of this project explores the use of data and machine learning to drive subsurface characterization, and provide some of the conceptual and software elements that the proposed project will build on.									
Partnership for Physics Simulation **Award Number: DOE -SC0008455**	Department of Energy	8/1/12	7/31/17	$325,000	Current	0	0	0.01	Rutgers University
The Rutgers component of this project develops a scalable coupling framework for coupled fusion simulations and provides some of the conceptual and software elements that the proposed project will build on.									
Development and Dissemination of MuscleMiner: An Imaging Tool for Muscle **Award Number: UFOER00010411**	National Health Institute	9/1/14	8/31/18	$150,000	Current	0	0	0.01	Rutgers University
The Rutgers component of this project explores the use of clouds for image processing. There is no overlap with the proposed project.									
OOI Cyberinfrastructure Operations and Maintenance **Award Number: SA# 11-09**	National Science Foundation	3/1/11	12/31/17	$22,387,157	Current	0	0	0.01	Rutgers University
The Rutgers component of this project operates and manages the cyberinfrastructure for the Ocean Observatories Initiative. There is no overlap with the proposed project.									
BD Hubs: Collaborative Proposal: NORTHEAST: The North East Big Data Innovation Hub **Award Number: 1546145**	National Science Foundation	10/1/15	9/30/18	$1,250,000	Current	0	0	0.01	Rutgers University
This project coordinates BigData related research in the Northeast. There is no Rutgers funding in this award, and does not overlap with the proposed project.									
SIRIUS: Science-Driven Data Management for Multi-tiered Storage **Award Number: DE-SC0015160**	Department of Energy	2/15/16	2/14/19	$472,461	Current	0.5	0	0	Rutgers University
The Rutgers component of this project explores data place as well as data movement across deep memory hierarchies for in-situ workflows, and provides some of the conceptual and software elements that the proposed project will build on.									
CIF21 DIBBs: EI: Virtual Data Collaboratory – A Regional Cyberinfrastructure for Collaborative Data Intensive Science **Award Number: 1640834**	National Science Foundation	9/1/16	8/31/20	$4,000,000	Current	0	0	0.01	Rutgers University
This project funds the design and deployment of a virtual, data sharing collaboratory. There is no overlap with the proposed project.									
ECP: High-Fidelity Whole Device Modeling of Magnetically Confined Plasmas **Award Number: DE-S015384-F**	Department of Energy ASCR PPPL	10/26/16	9/30/20	$498,420	Current	0.25	0	0	Rutgers University
The Rutgers component of this project uses an extreme scale data staging service to enabled coupled fusion simulations, and provides some of the conceptual and software elements that the proposed project will build on.									
ECP: Co-Design Center for Online Data Analysis and Reduction at the Exascale **Award Number: 325137**	Department of Energy ASCR BNL	12/30/16	12/31/18	$195,504	Current	0	0	0.01	Rutgers University
The Rutgers component of this project develops in-staging data services using an extreme scale data staging service, and provides some of the conceptual and software elements that the proposed project will build on.									
ECP: The ADIOS framework for Scientific Data on exascale systems **Award Number: 400152987**	Department of Energy ASCR ORNL	12/30/16	9/30/20	$515,850	Current	0.25	0	0	Rutgers University
The Rutgers component of this project explores develops an extreme scale staging service as part of the ADIOS 2.0 framework, and provides some of the conceptual and software elements that the proposed project will build on.									
Student Support: 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid'17) **Award Number: 1724679**	National Science Foundation	5/1/17	4/30/19	$20,000	Current	0	0	0.01	Rutgers University
This project funds US-based students to attend the CCGrid 2017 conference. There is no overlap with the proposed project.									
NSF Large Facilities Cyberinfrastructure	National Science	5/15/17	4/30/18	$65,118	Current	0	0	0.01	Rutgers University

Workshop **Award Number: 1742969**	Foundation								
This project funds the organization of the NSF large facilities cyberinfrastructure workshop. There is no overlap with the proposed project.									
Exploration of Resilient Distributed Data Warehouse to enable Scalable Local Recovery of AMT and SPMD Programming Models **Award Number 1774372**	Department of Energy Sandia National Laboratory	4/3/17	9/15/17	$50,000	Current	0	0	0.01	Rutgers University
II-EN: Collaborative Research: Enhancing the Parasol Experimental Testbed for Sustainable Computing **Award Number 1730043**	National Science Foundation	7/1/17	6/30/20	$691,713	Current	0.25	0	0	Rutgers University
EAGER: ACI: Online Processing of Data in Large Facilities using National ACI	National Science Foundation	7/1/17	6/30/19	$292,445	Pending	0	0	0.01	Rutgers University
SPX: Collaborative Research: Cross-layer Application-Aware Resilience at Extreme Scale (CAARES)	National Science Foundation	8/1/17	7/31/20	$267,247	Pending	0.5	0	0	Rutgers University
Management and Operation of the Ocean Observatories Initiative (OOI)	National Science Foundation	1/1/18	12/31/22	$9,961,865	Pending	0	0	0.01	Rutgers University
SI2-S2I2 Conceptualization: Collaborative Research: Software Institute for Robust Science: Reproducibility, Automation, and Trust in Software for Science	National Science Foundation	1/1/18	12/31/18	$50,000	Pending	0.5	0	0	Rutgers University
RCN: Reproducibility in Computational and Data-Enabled Science-Paradigms, Practices and Infrastructure	National Science Foundation	9/1/17	8/31/20	$215,667	Pending	0.5	0	0	Rutgers University
RCN: Coordinating Platforms for Science	National Science Foundation	9/16/17	9/15/18	$52,533	Pending	1	0	0	Rutgers University
Scientific Discovery through Advanced Computing: Partnerships in Basic Energy Sciences	Department of Energy ASCR BNL	9/1/17	8/31/21	$600,000	Pending	0.5	0	0	Rutgers University
A SciDAC Institute for Computer Science and Data THIS PROPOSAL	Department of Energy ORNL	9/30/17	9/30/20	$297,201	Pending	0.25	0	0	Rutgers University

ROBERT M. PATTON CURRENT & PENDING SUPPORT

Current

Project (PI): Scalable Deep Learning Algorithms for Exascale Data Analytics
Organization: ORNL Laboratory Directed Research and Development (LDRD)
Dates: Oct 2015–Sept 2017
Award amount: $780,000 total

Project (PI): Publication Mining for Better Materials
Organization: ORNL Laboratory Directed Research and Development (LDRD)
Dates: October 2016–September 2017
Award amount: $173,000

Pending

Project: Securing Heterogeneous SCADA Networks with Unknown Protocols Using Multi-Stage Machine Learning
Organization: ORNL Laboratory Directed Research and Development (LDRD)
Dates: October 2017–September 2019
Award amount: $800,000 total

Swann Perarnau – ANL

Current Support

- **Project Title:** Argo: Resource Management for Exascale (ECP)
  Principal Investigator: Pete Beckman (ANL)
  Funding Agency: DOE/ASCR
  Total Award: To ANL: $5.1M
  Funding level: 12 months
  Dates: 01/17 – 9/19

Pending Support

- **Project Title:** A SciDAC Institute for Computer Science and Data
  Principal Investigator: Robert B. Ross (ANL)
  Funding Agency: DOE/ASCR
  Total Award: To ANL: $18M
  Funding level: 4 months
  Dates: 09/17 – 08/20

Overlap: If projects are fully funded, then effort will be adjusted accordingly.

Tom Peterka

Current and Pending Support for Tom Peterka

- (Current) Title: High Performance Decoupling of Tightly-Coupled Data Flows
  Source: DOE Office of Science
  Award Amount: $637.5K/yr
  Award Period: 09/02/2014 - 09/01/2017
  Location: Argonne National Laboratory
  PI: Tom Peterka
  Effort: 50%

- (Current) Title: Extreme-Scale Distribution-Based Data Analysis
  Source: DOE Office of Science
  Award Amount: $425K/yr
  Award Period: 09/15/2014 - 09/14/2017
  Location: The Ohio State University
  PI: Han-Wei Shen
  Effort: 20%

- (Current) Title: CODAR: A Co-design Center for Online Data Analysis and Reduction at the Exascale
  Source: DOE Office of Science
  Award Amount: $4M/yr
  Award Period: 10/1/2016 - 09/30/2020
  Location: Argonne National Laboratory
  PI: Ian Foster
  Effort: 20%

- (Current) Title: ALETHEIA: A Framework for Automatic Detection of Corruption in Extreme-Scale Scientific Applications
  Source: NSF
  Award Amount: $450K
  Award Period: 10/01/2016 - 09/30/2019
  Location: University of Illinois
  PI: Marc Snir
  Effort: 10%

- (Pending) Title: A Continuous Model of Discrete Scientific Data
  Source: DOE Office of Science
  Award Amount: $500K/yr
  Award Period: 9/1/2017 - 08/31/2022
  Location: Argonne National Laboratory
  PI: Tom Peterka
  Effort: 50%

- (Pending) Title: Improved Predictability Through Efficient Ensemble Coupled Data Assimilation System in ACME
  Source: DOE Office of Science
  Award Amount: $2M/yr
  Award Period: 8/1/2017 - 07/31/2022
  Location: Los Alamos National Laboratory
  PI: Luke Van Roekel
  Effort: 50%

- (Pending) Title: Coupling Approaches for Next-Generation Architectures (CANGA)
  Source: DOE Office of Science
  Award Amount: $3M/yr
  Award Period: 8/1/2017 - 07/31/2022
  Location: Los Alamos National Laboratory
  PI: Philip Jones
  Effort: 40%

- (Pending) Title: HEP Data Analytics in HPC
  Award Amount: $2M/yr
  Award Period: 8/1/2017 - 07/31/2022
  Location: Fermi National Accelerator Laboratory
  PI: Jim Kowalkowski
  Effort: 20%

- (Pending) Title: Nested Task-Parallel Workflows for Scientific Applications
  Source: DOE Office of Science
  Award Amount: $300K/yr
  Award Period: 10/01/2017 - 09/30/2020
  Location: Argonne National Laboratory
  PI: Tom Peterka
  Effort: 35%

- (Pending) Title: Visual Analytics for Large Scale Scientific Ensemble Datasets
  Source: DOE Office of Science
  Award Amount: $400K/yr
  Award Period: 10/01/2017 - 09/30/2020
  Location: Los Alamos National Laboratory

PI: Jonathan Woodring
Effort: 15%

- (Pending) Title: SciDAC4 CS Institute
  Source: DOE Office of Science
  Award Amount: $6M/yr
  Award Period: 10/01/2017 - 09/30/2020
  Location: Argonne National Laboratory
  PI: Rob Ross
  Effort: 15%

Overlap: If projects are fully funded, then effort will be adjusted accordingly.

Current and Pending Support

Norbert Podhorszki

Oak Ridge National Laboratory

Current

Project Name: ORNL National Leadership Class Computing in the National Center for Computational Sciences
Amount: > $100,000,000
Location: Oak Ridge National Laboratory
Funding Source: Department of Energy ASCR
Lead P.I.: James J. Hack
Period covered: 10/01/2016 – 09/30/2017
Person Months: 6

Project Name: The ADIOS framework for Scientific Data on exascale systems
Amount: $1,500,000
Location: Oak Ridge National Laboratory
Funding Source: ECP, Department of Energy ASCR
Lead P.I.: Scott Klasky, ORNL
Period covered: 9/15/2016 – 9/15/2019
Person Months: 5.64 (47%)

Pending

Project Name: A SciDAC Institute for Computer Science and Data (This proposal)
Amount: $6,000,000
Location: Oak Ridge National Laboratory
Funding Source: Department of Energy FES/ASCR
Lead PI: R. Ross (ANL)
Period covered: 9/30/2017 – 9/30/2020
Person Months: 6

Current and Pending Funding

David Pugmire
Oak Ridge National Laboratory

July 2017

Current Support

1. XVis
   Funding Agency: DOE/ASCR
   Award Number: FWP 14-017566
   Total Funding: $4M
   Effort: 0.20FTE
   Award Period: 2014-2017

2. Oak Ridge Leadership Computing Facility
   Funding Agency: DOE
   Effort: 0.05FTE
   Total Funding: > $100M
   Award Period: 2013-ongoing

3. VTK-m ECP
   Funding Agency: DOE/ECP
   Effort: 0.20 FTE Total Funding: $4.5M
   Award Period: 2017-2019

Pending Projects

1. Scalable Analysis Methods for In Situ Infrastructure for Extreme Scale Knowledge
   Funding Agency: DOE/ASCR
   Total Funding: $2.5M
   Total Effort: 0.10FTE

2. XVis2
   Funding Agency: DOE/ASCR
   Award Number: FWP 14-017566
   Total Funding: $4M
   Effort: 0.20FTE

3. This Proposal
   Funding Agency: DOE SciDAC
   Total Funding: $6M
   Effort: 0.25FTE

Current and Pending Support for Robert Ross

Co-Design of Exascale Storage and Science Data Facilities (Current), FWP#57796
Source: DOE Office of Science Award Period: 10/01/13-09/14/16
Award Amount: $2,325K PI: Robert B. Ross
Location: Argonne National Laboratory Effort: 0.0 person-months/year (0%)
Research Description: Researching and developing scalable simulation capabilities targeting understanding of HPC storage systems.

Optimizing High Energy Physics Data Management and Analysis Capabilities: The Exaflop/Exabyte Nexus (Current), FWP#57L47
Source: DOE Offices of Science & HEP Award Period: 9/1/2015-8/31/2017
Award Amount (to ANL): $400K PI: Adam Lyons
Location: Argonne National Laboratory Effort: 1.2 person-months/year (10%)
Research Description: Building simulation capabilities and virtualized compute environments to better enable HEP data analytics.

Supporting Co-Design of Extreme Scale Systems with In Situ Visual Analysis of Event Driven Simulations (Current), FWP #57L39
Source: DOE Office of Science Award Period: 10/1/14-9/30/17
Award Amount (to ANL): $450K PI: Kwan-Liu Ma
Location: Argonne National Laboratory Effort: 1.2 person-months/year (10%)
Research Description: Researching and developing novel methods for understanding complex and highly parallel discrete event simulations.

Northwestern Joint Appointment (Current), FWP #05J65
Source: NSF Award Period: 5/1/2015-4/30/2017
Award Amount (to NU): $191K PI: Marianne Winslett
Location: Northwestern University Effort: 0.6 person-months/year (5%)
Research Description: Automated development of application- and system-specific benchmarks.

A Software Defined Storage Approach to Exascale Storage Services (Current), FWP#57L65
Source: DOE Office of Science Award Period: 9/30/15-09/29/18
Award Amount (to ANL): $1,280K PI: Robert B. Ross
Location: Argonne National Laboratory Effort: 0.6 person-months/year (5%)
Research Description: R&D to understand the composition of software defined I/O services in support of large scale scientific applications.

Data Libraries and Services Enabling Exascale Science (Current)
Source: DOE ECP Project Award Period: 9/15/2016-9/14/2019
Award Amount (to ANL): $2,695K PI: Robert B. Ross
Location: Argonne National Laboratory Effort: 1.8 person-months/year (15%)
Research Description: Development and productization of storage and I/O system software for exascale computing.

Exascale Computing Project CAM-ST Data Management & Workflows (Current), FWP#82691
Source: DOE ECP Project Award Period: 10/05/2016-09/30/2017
Award Amount (to ANL): $100K PI: Robert B. Ross
Location: Argonne National Laboratory Effort: 1.2 person-months/year (10%)
Research Description: Assist development and oversee Exascale Computing Project data management and workflows.

HEP Data Analytics on HPC (Pending)

Source: DOE	Award Period: 08/01/2017-07/31/2022
Award Amount (to ANL): $3,387.5K	PI: Robert B. Ross
Location: Argonne National Laboratory	Effort: 0.9 person-months/year (7.5%)

A SciDAC Institute for Computer Science and Data (Pending)

Source: DOE	Award Period: 09/30/2017 – 09/29/2020
Award Amount (to ANL): $3,543.0K	PI: Robert B. Ross
Location: Argonne National Laboratory	Effort: 2.4 person-months/year (20.0%)

Note: 45% of effort is covered on programmatic overhead.
Overlap: If projects are fully funded, then effort will be adjusted accordingly.

Philip C. Roth (ORNL)

Current

(1) Sponsor: DOE FES & ASCR
Award Number: DOE FWP ERAT336
Title: Plasma Surface Interactions: Bridging from the Surface to the Micron Frontier through Leadership Class Computing
Total award amount for entire award period (including direct and indirect costs): $5.25M over 5 years
Number of Person-months per year effort: 3 (ending in FY2017)

(2) Sponsor: DOE NE & ASCR
Award Number: DOE FWP ERKJ317
Title: Advancing Understanding of Fission Gas Behavior and Release in Nuclear Fuel through Leadership Class Computing
Total award amount for entire award period (including direct and indirect costs): $1.08M over 2.5 years
Number of Person-months per year effort: 2.4

(3) Sponsor: DOE ASCR
Award Number: DOE FWP ERKJ280
Title: An Evaluation of the Performance of Functional Programming for Extreme Scale Computing
Total award amount for entire award period (including direct and indirect costs): $500K over 3 years
Number of Person-months per year effort: 4.8

(4) Sponsor: DOE ASCR
Award Number: DOE FWP ERKJN12
Title: SciDAC Institute for Sustained Performance, Energy, and Resilience
Total award amount for entire award period (including direct and indirect costs): $15M over 5 years
Number of Person-months per year effort: 1.8 (ending in FY2017)

Pending

(1) Sponsor: DOE ASCR
Title: A SciDAC Institute for Computer Science and Data
Total amount for entire award period (including indirect costs): $18M over 3 years
Number of Person-months per year effort: 3

(2) Sponsor: DOE BES & ASCR
Title: Atomic- To Continuum-Scale Transport and Reaction in Porous Media
Total amount for entire award period (including indirect costs): $8M over 4 years
Number of Person-months per year effort: 3.6

(3) Sponsor: DOE FES & ASCR
Title: Plasma Surface Interactions: Predicting the Performance and Impact of Evolving PFC Surfaces
Total award amount for entire award period (including direct and indirect costs): $22M over 5 years
Number of Person-months per year effort: 3

(4) Sponsor: ORNL LDRD
Title: Application Characterization and Mapping for Abstract Exascale Machine Models
Total award amount for entire award period (including direct and indirect costs): $800K over 3 years
Number of Person-months per year effort: 2.5

Current and Pending Funding

OLIVER RÜBEL
LAWRENCE BERKELEY NATIONAL LABORATORY

April, 2017

Current Projects

1. Funding Organization: DOE/SC/ASCR (Exascale Computing Project (ECP))
   Project Title: ECP ALPINE: Algorithms and Infrastructure for In Situ Visualization and Analysis
   Program Lead PI: James Ahrens (LANL)
   Dates of award: January 2017 – November 2019
   Annual funding: $2M ($400K/yr for LBNL Visualization)
   Level of effort: 0.5FTE

2. Funding Organization: KAVLI Foundation
   Project Title: Infrastructure and Algorithms to Enable Data Driven Discovery and Dissemination on HPC for the B.R.A.I.N. Initiative
   Program Lead PI: Kristofer Bouchard (LBNL)
   Dates of award: January 2017 – October 2017
   Annual funding: $350K
   Level of effort: 0.5FTE

Pending Proposals

1. Funding Organization: DOE/SC/ASCR and DOE/SC/FES
   Project Title: Whole Device Modeling at the Center for Effective Research in Fusion
   PI: A. Koniges (LBNL)
   Dates of award: August 2017 – July 2022
   Total Annual Funding: $3M/yr
   Level of effort: 0.35 FTE

2. Funding Organization: LDRD (LBNL)
   Project Title: The chemical universe through the eyes of generative adversarial neural networks
   PI: Wibe Albert de Jong (LBNL)
   Dates of award: October 2017 – October 2018
   Total Annual Funding: $328K/yr
   Level of effort: 0.05 FTE

3. Funding Organization: DOE/SC/ASCR.
   Project Title: Scalable Data-Computing Convergence and Scientific Knowledge Discovery
   PI: E. Wes Bethel.
   Dates of award: August 2017 – July 2020.
   Annual funding: $1000K.
   Level of effort: 0.20 FTE.

4. Funding Organization: DOE/SC/ASCR.
   Project Title: SciDAC RAPIDS (this project)
   PI: Robert Ross (ANL)
   Dates of award: August 2017 – July 2020.
   Annual funding: $6M/yr
   Level of effort: 0.33 FTE.

Current and Pending Support for Han-Wei Shen

Current

Title: BIGDATA: SMALL: DA: Data Summarization, Analysis, and Triage for Very Large Scale Flow Fields
Award Amount: $727,258
Award Period Covered: 09/15/13 - 08/31/17
Source of Support: NSF
PI: Han-Wei Shen

Title: Extreme-Scale Distribution-Based Data Analysis
Award Amount: $637,500
Award Period: 09/01/14 - 08/31/17
Source of Support: DOE
PI: Han-Wei Shen

Pending

Title: BIGDATA:F:Visual Analytics for Large Scale Scientific Ensemble Datasets
Award Amount: $785,231
Award Period Covered: 09/01/17 - 08/31/20
Source of Support: NSF
PI: Han-Wei Shen

Stephen F. Siegel: Current and Pending Support

1 Current Support

1. Title: *SHF: Small: Contracts for Message-Passing Parallel Programs*
   - Sponsor: National Science Foundation
   - Award number: CCF-1319571
   - Total Amount: $449,999
   - Dates: September 1, 2013 – August 31, 2018
   - Effort: 1 summer month per year
2. Title: *Program Verification for Extreme-Scale Applications*
   - Sponsor: Department of Energy
   - Award number: DE-SC0012566
   - Total Amount: $510,000
   - Dates: September 1, 2014 – August 31, 2018
   - Effort: 1 summer month per year

2 Pending Support

1. Title: *A SciDAC Institute for Computer Science and Data* (this proposal)
   - Sponsor: Department of Energy
   - Dates: September 30, 2017 – September 30, 2020
   - Total Amount: $246,600 (for UD)
   - Effort: 0.5 summer months per year

Current and Pending Support

Alex Sim

Current Support
Funding Organization: Dept. of Energy
Project Title: Scientific Data Management Research in the Exascale Era
Entire Award Period: 07/01/15 – 06/30/18
Award Amount: 800K per year
PI: K. John Wu
Person-months per Year Devoted to the Project: 7.2 months

Funding Organization: Dept. of Energy
Project Title: ExaHDF5: Advancing HPC I/O to Enable Scientific Discovery
Entire Award Period: 09/01/13 – 09/30/17
Award Amount: 535K per year at LBNL
PI: Surendra Byna
Person-months per Year Devoted to the Project: 4.8 months

CHAD A. STEED CURRENT & PENDING SUPPORT

Current

Project (PI): New Multi-modal Interactive Data Visualization Techniques for Scientific Data Analysis
Organization: ORNL Laboratory Directed Research and Development (LDRD)
Dates: May 2017–May 2018
Award amount: $190,000 total
Person months per year: 2

Project: Cyber Analytic Tools and Techniques
Organization: Department of Energy, Office of Intelligence and Counterintelligence
Dates: October 2016–September 2017
Award amount: $600,000
Person months per year: 5

Pending

Project: Physics-informed Machine Learning Based Closure Laws for Nuclear Systems Design and Analysis Codes
Organization: ORNL Laboratory Directed Research and Development (LDRD)
Dates: October 2017–September 2019
Award amount: $900,000 total
Person months per year: 2

Project: Multimodal Imaging of Shoot Apical Meristem to Model Plant System Growth and Response
Organization: ORNL Laboratory Directed Research and Development (LDRD)
Dates: October 2017–September 2019
Award amount: $900,000
Person months per year: 2

Samuel Williams – LBNL

Current Support

- **Project Title:** AMReX: Block-Structured AMR Co-Design Center (ECP)
  Principal Investigator: John Bell (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $3M/year (#DE-AC02-05CH11231)
  Funding level: 6 months
  Dates: 10/16 – 9/20
- **Project Title:** Autotuning Compiler Technology for Cross-Architecture Transformation and Code Generation (ECP)
  Principal Investigator: Mary Hall (Utah)
  Funding Agency: DOE/ASCR
  Total Award: $300K/year (#DE-AC02-05CH11231)
  Funding level: 2.4 months
  Dates: 10/16 – 9/19
- **Project Title:** Design Space Evaluation (ECP)
  Principal Investigator: David Donofrio (LBNL)
  Funding Agency: DOE/ASCR
  Total Award: $300K/year (#DE-AC02-05CH11231)
  Funding level: 0.6 months
  Dates: 4/17 – 3/20

Pending Support

- **Project Title:** (proposed) ISEP: Integrated Simulation of Energetic Particles in Burning Plasmas (SciDAC4)
  Principal Investigator: Zhihong Lin (UC Irvine)
  Funding Agency: DOE/ASCR,FES
  Total Award: $200K/year
  Funding level: 2.4 months
  Dates: 6/17 – 5/22
- **Project Title:** (proposed) CTTS: Center for Tokamak Transient Simulations (SciDAC4)
  Principal Investigator: Stephen Jardin (PPPL)
  Funding Agency: DOE/ASCR,FES
  Total Award: $300K/year
  Funding level: 1.2 months
  Dates: 6/17 – 5/22
- **Project Title:** (proposed) A MultiScale Approach to Nuclear Structure and Reactions: Applications to Fundamental Symmetries (SciDAC4)
  Principal Investigator: Wick Haxton (LBNL)
  Funding Agency: DOE/ASCR,NP
  Total Award: $800K/year
  Funding level: 0.6 months
  Dates: 6/17 – 5/22
- **Project Title:** (proposed) Particle Accelerator Machine Learning with Multi-Physics Parallel Simulators (SciDAC4)
  Principal Investigator: Ji Qiang (LBNL)
  Funding Agency: DOE/ASCR,HEP

Total Award: $150K/year
Funding level: 0.6 months
Dates: 6/17 – 5/22

- **Project Title:** (proposed) A Calibrated Global Cloud Resolving Model designed for Extreme Climate Simulation (SciDAC4)
  Principal Investigator: Williams Collins (LBNL)
  Funding Agency: DOE/ASCR,BER
  Total Award: $818K/year
  Funding level: 3.6 months
  Dates: 6/17 – 11/19
- **Project Title:** (proposed) Transforming Climate Modeling with Solution-Adaptive Computational Techniques (SciDAC4)
  Principal Investigator: Christiane Jablonowski (U. Michigan)
  Funding Agency: DOE/ASCR,BER
  Total Award: $800K (total)
  Funding level: 1.2 months
  Dates: 6/17 – 11/19
- **Project Title:** (proposed) Advancing Catalysis Modeling: From Atomistic Chemistry to Whole System Simulation (SciDAC4)
  Principal Investigator: Martin Head Gordon (LBNL)
  Funding Agency: DOE/ASCR,BES
  Total Award: $250K/year
  Funding level: 0.6 months
  Dates: 10/17 – 9/20
- **Project Title:** (proposed) A SciDAC Institute for Computer Science and Data (SciDAC4)
  Principal Investigator: Robert Ross (ANL)
  Funding Agency: DOE/ASCR
  Total Award: $1766K/year
  Funding level: 2.4 months
  Dates: 10/17 – 9/20

Current and Pending Funding

JEFFREY S. VETTER

July 24, 2017

Dr. Vetter has multiple appointments. Vetter serves as the Future Technologies Group Leader at Oak Ridge National Laboratory (ORNL), and as a Joint Professor in the Electrical and Computer Engineering Department at the University of Tennessee, Knoxville. In this arrangement, Dr. Vetter leads projects across these organizations; the majority of funding from these projects provides support for staff, students, resources, and travel at their respective locations. The following list outlines active projects where Dr. Vetter is PI, co-PI, or senior personnel.

Status	Title	Sponsor	Location	Partners	Total Award ($K)	Start Date	Duration (yr)	Commitment
Current	Tahoe: Designing and programming exascale memory hierarchies	LDRD	ORNL	GT	720	10/2016	2	10
Current	Exascale Computing Project	DOE	ORNL	(many)	n/a	10/2016	3	40
Current	Panorama: Predictive Modeling and Diagnostic Monitoring of Extreme Science Workflows	DOE	ORNL	ISI, RPI, RENCI, LBNL	4500	09/2014	3	10
Current	ARES: Abstract Representations for Extreme-Scale Software Stack	DOE	ORNL	LANL	3000	06/2014	3	10
Current	Oxbow: Synthesizing architectural features for extreme-scale co-design applications	DOE	ORNL		2250	10/2013	3	10
Current	Uncertainty Quantification Methodologies for Enabling Extreme-Scale Science	DOE	ORNL	Texas, MIT	2250	08/2013	3	10
Pending	Collaborative: SI2-S2I2 Conceptualization: Software Institute for Robust Science: Reproducibility, Automation and Trust in Software for Science	NSF	UTK	ISI, Rutgers, UIUC, Delaware, Indiana		01/2018	1	5
Pending	Sawtooth: Investigating solutions to memory and storage challenges	DOE	ORNL	Micron Technologies	12000	09/2017	3	40
Pending	RAPIDS: SciDAC4 Computer Science and Data Institute	DOE	ORNL	ANL, LBL, LANL, LLNL, Oregon	6000	09/2017	3	20
Pending	Methods and Interfaces for Quantum Acceleration of Scientific Applications	DOE	ORNL	IBM, Google, IonQ, UCSB, Georgia Tech, Virginia Tech, Dalhousie University, Vencore	4500	09/2017	3	10
Pending	Heterogeneous Quantum Dynamics Simulator	DOE	ORNL	University of Basque, Michigan State, University of Washington	4500	09/2017	3	10
Pending	Quantum Sampling Algorithms for Scientific Applications of Machine Learning	DOE	ORNL	USC, LANL, FNAL, CalTech	4500	09/2017	3	10
Pending	Panorama360: Performance data capture and analysis for end-to-end scientific workflows	DOE	ORNL	ISI, RENCI, LBNL	3000	09/2017	3	10

Shinjae Yoo

Current and Pending Support
Investigator: Adam Akin (Lead, LBL), Rick Stevens (ANL), Bob Cottingham (ORNL), **Shinjae Yoo** (BNL)
Support: ☒ Funded ☐ Pending
Sponsor or Source of Funding: DOE BER
Award Number or Identifying Number: FWP#ELS165
Title of the Funded Research Project: KBase: An Integrated Knowledgebase for Predictive Biology and Environmental Research
The total award amount, including direct and indirect costs: $780k /yr
The number of person-months per year to be devoted to the project: 2.4
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Building integrated knowledgebase for predictive biology for synthetic biology and system biology. Research on extreme scale knowledge inference using machine learning.

Current and Pending Support
Investigator: **Shinjae Yoo**, Deyu Lu, Dmitri Zakharov, Eric Stach
Support: ☒ Funded ☐ Pending
Sponsor or Source of Funding: BNL
Award Number or Identifying Number: LDRD 16-039
Title of the Funded Research Project: Machine Learning Assisted Material Discovery
The total award amount, including direct and indirect costs: $1,342,557
The number of person-months per year to be devoted to the project: 1.2
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Automated near real-time image understanding and core level spectroscopy genome with multi-modal material characterization using machine learning.

Current and Pending Support
Investigator: Ian Foster (ANL), Scott Klasky (ORNL), Kerstin Kleese Van Dam, Todd Munson (ANL), **Shinjae Yoo** (BNL)
Support: ☒ Funded ☐ Pending
Sponsor or Source of Funding: DOE ECP
Title of the Funded Research Project: CODAR: Co-Design Center for Online Data Analysis and Reduction at the Exascale
The total award amount, including direct and indirect costs: $16M
The number of person-months per year to be devoted to the project: 3.6

Current and Pending Support
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Porting and adapting existing machine learning algorithms to exascale application, hardware and software for online analysis.

Current and Pending Support
Investigator: **Shinjae Yoo**, Yiming Yang, Yangang Liu, Nicholas D'Imperio
Support: ☒ Funded ☐ Pending
Sponsor or Source of Funding: DOE ASCR
Award Number or Identifying Number: FWP#CC064CCAA
Title of the Funded Research Project: Robust Extreme-scale Multimodal Structured Learning from Spatio-temporal Data
The total award amount, including direct and indirect costs: $1.95M
The number of person-months per year to be devoted to the project: 1.2
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Exascale algorithm developments for extreme scale spatio-temporal data analysis and advancing ML to the next level in HPC.

Current and Pending Support
Investigator: Ian Blaby, **Shinjae Yoo**
Support: ☒ Funded ☐ Pending
Sponsor or Source of Funding: BNL
Award Number or Identifying Number: LDRD 17-018
Title of the Funded Research Project: Genomes to predictive biology: machine learning for the integration of inter-species functional genomics data
The total award amount, including direct and indirect costs: $600K
The number of person-months per year to be devoted to the project: 3.6
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Multimodal analysis of various evidences from literature to phylogenetic relationships to have better functional understanding.

Current and Pending Support
Investigator: Salman Habib, Katrin Heitmann, Anze Slosar, Earl Lawrence, Zarija Lukic, **Shinjae Yoo**
Support: ☒ Funded ☐ Pending
Sponsor or Source of Funding: DOE HEP/ASCR
Award Number or Identifying Number: FWP#CC100

Current and Pending Support
Title of the Funded Research Project: Accelerating HEP Science: Inference and Machine Learning at Extreme Scales
The total award amount, including direct and indirect costs: $0.71M
The number of person-months per year to be devoted to the project: 1.2
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Exascale algorithm developments for extreme scale spatio-temporal data analysis and advancing ML to the next level in HPC.

Current and Pending Support
Investigator: Yangang Liu, Yu Xie, Qilong Min, **Shinjae Yoo**
Support: ☐ Funded ☒ Pending
Sponsor or Source of Funding: DOE EERE
Title of the Funded Research Project: Advancing the WRF-Solar Model to Improve Solar Irradiance Forecast in Cloudy Environments
The total award amount, including direct and indirect costs: $1.8M
The number of person-months per year to be devoted to the project: 2.1
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Build extreme scale statistical WRF-Solar emulator using machine learning and online streaming analysis.

Current and Pending Support
Investigator: Dimitri Katramas, **Shinjae Yoo**, Meng Yue, Michael DePhillips
Support: ☐ Funded ☒ Pending
Sponsor or Source of Funding: LDRD
Title of the Funded Research Project: Analysis on the Wire
The total award amount, including direct and indirect costs: $1.35M
The number of person-months per year to be devoted to the project: 1.2
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Build approximated streaming analysis algorithms over analysis on the wire framework.

Current and Pending Support
Investigator: Robert Konik, Gabi Kotliar, **Shinjae Yoo**, Manish Parashar, Kipton Barros, Ivan Bozovic
Support: ☐ Funded ☒ Pending
Sponsor or Source of Funding: DOE BES/ASCR
Title of the Funded Research Project: Designing and Discovering New Oxide Heterostructure through Dynamic Mean Field Theory Informed by Machine Learning

Current and Pending Support
The total award amount, including direct and indirect costs: $5.3M
The number of person-months per year to be devoted to the project: 1.2
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Build extreme scale statistical dynamic mean field emulator using machine learning.

Current and Pending Support
Investigator: Robert Harrison, Hubertus van Dam, **Shinjae Yoo**, Deyu Lu, Ping Liu, Anatoly Frenkel, David Prendergast
Support: ☐ Funded ☒ Pending
Sponsor or Source of Funding: DOE BES
Title of the Funded Research Project: Machine learning and computational core-level absorption spectroscopy assisted discovery of structure-activity relations of catalytic complexes in operando experiments
The total award amount, including direct and indirect costs: $7.6M
The number of person-months per year to be devoted to the project: 1
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Build tools for X-ray Absorption Spectroscopy analysis among structures and experiments

Current and Pending Support
Investigator: Adam Akin (Lead, LBL), Christopher Henry (ANL), Bob Cottingham (ORNL), **Shinjae Yoo** (BNL)
Support: ☐ Funded ☒ Pending
Sponsor or Source of Funding: DOE BER
Title of the Funded Research Project: KBase: An Integrated Knowledgebase for Predictive Biology and Environmental Research
The total award amount, including direct and indirect costs: $1.3M
The number of person-months per year to be devoted to the project: 1.2
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: Building integrated knowledgebase for predictive biology for synthetic biology and system biology. Research on extreme scale knowledge inference using machine learning.

Current and Pending Support
Investigator: Rober Ross (Lead, LBL), **Shinjae Yoo** (BNL)
Support: ☐ Funded ☒ Pending
Sponsor or Source of Funding: DOE ASCR
Title of the Funded Research Project: A SciDAC Institute for Computer Science and Data
The total award amount, including direct and indirect costs: $450k
The number of person-months per year to be devoted to the project: 0.6

Current and Pending Support
Briefly describe the research being performed and explicitly identify any overlaps with the proposed research: This proposal

Current and Pending Funding

Gunther H. Weber
Lawrence Berkeley National Laboratory

July 2017

Current Projects

1. Funding Organization: DOE/SC/ASCR.
   Scalable Analysis Methods and *In Situ* Infrastructure for Extreme Scale Knowledge Discovery
   Lead PI: E. Wes Bethel (LBNL).
   Dates of award: August 2014–July 2017.
   Annual funding: $2.25M.
   Level of effort: 0.3 FTE.

2. Funding Organization: DOE/SC/ASCR.
   Project Title: Towards Exascale: High Performance Visualization and Analytics Program
   Lead PI: E. Wes Bethel (LBNL).
   Dates of award: August 2014–July 2017.
   Annual funding: $750K.
   Level of effort: 0.2 FTE.

3. Funding Organization: DOE
   ECP ALpINe: Algorithms and Infrastructure for In Situ Visualization and Analysis
   Lead PI: James Ahrens (LANL)
   Total Funding: US$4,946,600
   Award Period: December 2016–November 2019
   Level of effort: 0.3 FTE

4. ScienceSearch: Automated Metadata using Machine Learning
   Funding Organization: DOE ASCR
   Project Lead PI: Katie Antypas (LBNL)
   Dates of award: October 2016–September 2019
   Annual funding: US$ 1,950,000
   Level of effort: 0.2 FTE

Pending Proposals

1. Funding Organization: DOE/SC/ASCR
   Project Title: A SciDAC Institute for Computer Science and Data (this project)
   Lead PI: Robert Ross (ANL)
   Dates of award: September 2017–September 2020
   Annual funding: $6M
   Level of effort: 0.33FTE

2. Funding Organization: DOE FES/ASCR
   High-fidelity Calculation of Tokamak Plasmas with Magnetic Islands in Whole-Device Modeling
   Lead PI: Lynda L. LoDestro
   Dates of award: October 2017–September 2022
   Total funding: US$ 5,625,000
   Level of effort: 0.25 FTE

3. Funding Organization: DOE/SC/ASCR
   Scalable Data-Computing Convergence and Scientific Knowledge Discovery
   PI: E. Wes Bethel (LBNL)
   Dates of award: August 2017–July 2020.
   Annual funding: $1000.
   Level of effort: 0.15 FTE

4. Funding Organization: DOE/SC/ASCR.
   Scalable Analysis Methods and *In Situ* Infrastructure for Extreme Scale Knowledge Discovery
   Lead PI: E. Wes Bethel
   Dates of award: August 2017–July 2020
   Annual funding: $1735K
   Level of effort: 0.3 FTE

5. Funding Organization: DOE/SC/ASCR.
   Project Title: Center for Integrated Multiscale Model Reduction Computational Capabilities
   Program Lead PI: S. Vassilevski (Lawrence Livermore National Laboratory)
   Dates of award: October 2017–September 2022
   Annual funding: $3.5M
   Level of effort: 0.25 FTE (Year 1); 0.125 FTE (Year 2-5)

Current and Pending Support

K. John Wu

Project Name:	Scientific Data Services -- Autonomous Data Management on Exascale Infrastructure
Funding Source:	Department of Energy
Award number:	DE-AC02-05CH11231
Funding Level:	2400K total over three years
P.I.:	K. John Wu
Period covered:	07/01/15 – 06/30/18
Man-months per year:	6.0
Work performed:	Manage the project, design SDS system, implement indexing techniques, and interact with users

Overlaps with the proposed research: None

Project Name:	SciDAC Scalable Data management, Analysis and Visualization Institute
Funding Source:	Department of Energy
Award number:	DE-AC02-05CH11231
Funding Level:	2750K total over five years
P.I.:	Arie Shoshani
Period covered:	03/01/12 – 02/28/17
Man-months per year:	4.8
Work performed:	Apply indexing techniques to SciDAC applications

Overlaps with the proposed research: None

Funding Organization:	Dept. of Energy
Project Title:	The ADIOS framework for Scientific Data on exascale systems
Entire Award Period:	09/15/16 – 09/14/19
Lead P.I.:	Scott Klasky
Award Amount:	4500K total over three years

Person-months per Year Devoted to the Project: 1.2

Work performed:	Manage the project at LBNL and design interoperability functions

Overlaps with the proposed research: None

Appendix 3: References

[1] K. Asanovic, R. Bodik, J. Demmel, T. Keaveny, K. Keutzer, J. Kubiatowicz, N. Morgan, D. Patterson, K. Sen, J. Wawrzynek, D. Wessel, and K. Yelick, "A view of the parallel computing landscape," *Commun. ACM*, vol. 52, no. 10, pp. 56–67, Oct. 2009. [Online]. Available: http://doi.acm.org/10.1145/1562764.1562783

[2] A. Dubey and L. McInnes, "Idea paper: Software lifecycle for scientific simulation software," 2016, wSSSPE-4.

[3] M. A. Heroux and J. M. Willenbring, "Barely sufficient software engineering: 10 practices to improve your CSE software," in *Proceedings of the 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering*, ser. SECSE '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 15–21. [Online]. Available: http://dx.doi.org/10.1109/SECSE.2009.5069157

[4] G. Wilson, D. Aruliah, C. T. Brown, N. P. C. Hong, M. Davis, R. T. Guy, S. H. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumblet, B. Waugh, E. P. White, and P. Wilson, "Best practices for scientific computing," *CoRR*, vol. abs/1210.0530, 2013.

[5] A. Dubey, K. Antypas, E. Coon, and K.M. Riley, "Software process for multicomponent multiphysics codes," in *Software Engineering for Science*, J. Carver, N.C. Hong, and G. Thiruvathukal, Eds. Taylor and Francis, 2016, ch. 1.

[6] R. A. Bartlett, A. Dubey, X. Sherry Li, J.D. Moulton, J.M. Willenbring, and U.M. Yang, "Testing of scientific software: Impacts on research credibility, development productivity, maturation, and sustainability," in *Software Engineering for Science*, J. Carver, N.C. Hong, and G. Thiruvathukal, Eds. Taylor and Francis, 2016, ch. 4.

[7] A. Dubey and K.M. Riley, "Experience paper: Software engineering and community codes track in ATPESC," 2016, wSSSPE-4.

[8] F. Zhang, C. Docan, M. Parashar, S. Klasky, N. Podhorszki, and H. Abbasi, "Enabling in-situ execution of coupled scientific workflows," in *Proceedings for the 26th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2012), Shanghai, China*, 2012.

[9] S. R. Slattery, P. P. H. Wilson, and R. P. Pawlowski, "The data transfer kit: a geometric rendezvous-based tool for multiphysics data transfer," in *International conference on mathematics & computational methods applied to nuclear science & engineering (M&C 2013)*, 2013.

[10] C. Docan, M. Parashar, and S. Klasky, "Dataspaces: an interaction and coordination framework for coupled simulation workflows," *Cluster Computing*, vol. 15, no. 2, pp. 163–181, 2012.

[11] Exascale Computing Project, "Training events," https://exascaleproject.org/, last accessed 2017-07-18.

[12] Testing Scientific HPC Software, https://cse-software.org/sc16-tutorial-testing-of-hpc-scientific-software/, last accessed 2017-07-24.

[13] CSE Collaboration through Software: Improving Productivity and Sustainability, http://meetings.siam.org/sess/dsp_programsess.cfm?SESSIONCODE=61488, last accessed 2017-07-24.

[14] R. Latham, R. Ross, and B. Welch, "Parallel I/O in practice," SC 2016, Salt Lake City, UT, November 2016, (to appear).

[15] Argonne Training Program on Extreme-Scale Computing, https://extremecomputingtraining.anl.gov/, last accessed 2017-07-18.

[16] R. Latham, R. Ross, and Q. Koziol, "HPC I/O for computational scientists," Argonne Training Program on Extreme-Scale Computing (ATPESC), St. Charles, IL, Aug. 2014.

[17] Interoperable Design of Extreme-scale Application Software (IDEAS), https://ideas-productivity.org/, last accessed 2017-07-18.

[18] Center for Scalable Application Development Software (CScADS), http://cscads.rice.edu/, last accessed 2017-07-18.

[19] G. Haller, "A variational theory of hyperbolic Lagrangian coherent structures," *Physica D: Nonlinear Phenomena*, vol. 240, no. 7, pp. 547–598, 2011.

[20] H. Guo, W. He, T. Peterka, H.-W. Shen, S. M. Collis, and J. J. Helmus, "Finite-time Lyapunov exponents and Lagrangian coherent structures in uncertain unsteady flows," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 6, pp. 1672–1682, 2016.

[21] M. Otto, T. Germer, and H. Theisel, "Uncertain topology of 3D vector fields," in *Proceedings of IEEE Pacific Visualization Symposium 2011*, 2011, pp. 67–74.

[22] C. M. Wittenbrink, A. Pang, and S. K. Lodha, "Glyphs for visualizing uncertainty in vector fields," *IEEE Trans. Vis. Comput. Graph.*, vol. 2, no. 3, pp. 266–279, 1996.

[23] R. P. Botchen, D. Weiskopf, and T. Ertl, "Texture-based visualization of uncertainty in flow fields," in *Proceedings of IEEE Visualization 2005*, 2005, pp. 647–654.

[24] C. A. Steed, D. M. Ricciuto, G. Shipman, B. Smith, P. E. Thornton, D. Wang, X. Shi, and D. N. Williams, "Big data visual analytics for exploratory earth system simulation analysis," *Computers & Geosciences*, vol. 61, pp. 71–82, 2013.

[25] C. A. Steed, W. Halsey, R. Dehoff, S. L. Yoder, V. Paquit, and S. Powers, "Falcon: Visual analysis of large, irregularly sampled, and multivariate time series data in additive manufacturing," *Computers & Graphics*, vol. 63, pp. 50–64, 2017.

[26] D. Smirnov and D. Morozov, "Triplet merge trees," in *Proceedings of Topology-Based Methods in Visualization (TopoInVis)*, 2017.

[27] D. Morozov and G. H. Weber, "Distributed merge trees," in *Proceedings of the Annual Symposium on Principles and Practice of Parallel Programming (PPOPP)*, 2013, pp. 93–102.

[28] D. Morozov and T. Peterka, "Block-parallel data analysis with diy2," 2016.

[29] J. Zhang, H. Guo, F. Hong, and X. Y. T. Peterka, "Dynamic load balancing based on constrained k-d tree decomposition for parallel particle tracing," *IEEE Trans. Vis. Comput. Graph. (Proc. IEEE SciVis 17)*, 2017 (to appear).

[30] H. Childs, E. Brugger, B. Whitlock, J. Meredith, S. Ahern, D. Pugmire, K. Biagas, M. Miller, C. Harrison, G. H. Weber, H. Krishnan, T. Fogal, A. Sanderson, C. Garth, E. W. Bethel, D. Camp, O. Rübel, M. Durant, J. M. Favre, and P. Navrátil, "VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data," in *High Performance Visualization–Enabling Extreme-Scale Scientific Insight*, Oct 2012, pp. 357–372.

[31] U. Ayachit, *The ParaView Guide: A Parallel Visualization Application*. USA: Kitware, Inc., 2015.

[32] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit (4th Ed.)*. Kitware, 2006.

[33] K. Moreland, C. Sewell, W. Usher, L. t. Lo, J. Meredith, D. Pugmire, J. Kress, H. Schroots, K. L. Ma, H. Childs, M. Larsen, C. M. Chen, R. Maynard, and B. Geveci, "VTK-m: Accelerating the visualization toolkit for massively threaded architectures," *IEEE Computer Graphics and Applications*, vol. 36, no. 3, pp. 48–58, May 2016.

[34] Q. Liu, J. Logan, Y. Tian, H. Abbasi, N. Podhorszki, J. Y. Choi, S. Klasky, R. Tchoua, J. Lofstead, R. Oldfield, M. Parashar, N. Samatova, K. Schwan, A. Shoshani, M. Wolf, K. Wu, and W. Yu, "Hello ADIOS: the challenges and lessons of developing leadership class I/O frameworks," *Concurrency and Computation: Practice and Experience*, vol. 26, pp. 1453–1473, 2014, software available at https://www.olcf.ornl.gov/center-projects/adios/.

[35] S. S. Shende and A. D. Malony, "The TAU parallel performance system," *Int. J. High Perform. Comput. Appl.*, vol. 20, no. 2, pp. 287–311, May 2006. [Online]. Available: http://dx.doi.org/10.1177/1094342006064482

[36] L. Adhianto, S. Banerjee, M. Fagan, M. Krentel, G. Marin, J. Mellor-Crummey, and N. R. Tallent, "HPCTOOLKIT: Tools for performance analysis of optimized parallel programs," *Concurrency Computation Practice and Experience*, vol. 22, no. 6, pp. 685–701, 2010.

[37] D. Lecomber and P. Wohlschlegel, *Debugging at Scale with Allinea DDT*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 3–12. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-37349-7_1

[38] B. Whitlock, J. M. Favre, and J. S. Meredith, "Parallel in situ coupling of simulation with a fully featured visualization system," in *Proceedings of the 11th Eurographics Conference on Parallel Graphics and Visualization*, ser. EGPGV '11. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2011, pp. 101–109. [Online]. Available: http://dx.doi.org/10.2312/EGPGV/EGPGV11/101-109

[39] U. Ayachit, A. Bauer, B. Geveci, P. O'Leary, K. Moreland, N. Fabian, and J. Mauldin, "Paraview catalyst: Enabling in situ data analysis and visualization," in *Proceedings of the First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization*, ser. ISAV2015. New York, NY, USA: ACM, 2015, pp. 25–29. [Online]. Available: http://doi.acm.org/10.1145/2828612.2828624

[40] U. Ayachit, B. Whitlock, M. Wolf, B. Loring, B. Geveci, D. Lonie, and E. W. Bethel, "The SENSEI generic in situ interface," in *2016 Second Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (ISAV)*, Nov 2016, pp. 40–44.

[41] S. Habib and R. Roser, "High energy physics exascale requirements review," DOE ASCR Workshop Report, June 2015.

[42] T. Windus, T. Devereaux, and M. Banda, "Basic energy sciences exascale requirements review," DOE ASCR Workshop Report, November 2015.

[43] M. Berry, T. E. Potok, P. Balaprakash, H. Hoffmann, R. Vatsavai, and Prabhat, "Machine learning and understanding for intelligent extreme scale scientific computing and discovery," DOE ASCR Workshop Report, Tech. Rep., 2015. [Online]. Available: http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Machine_Learning_DOE_Workshop_Report.pdf

[44] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[45] S. R. Young, D. C. Rose, T. P. Karnowski, S.-H. Lim, and R. M. Patton, "Optimizing deep learning hyper-parameters through an evolutionary algorithm," in *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments*, ser. MLHPC '15. New York, NY, USA: ACM, 2015, pp. 4:1–4:5. [Online]. Available: http://doi.acm.org/10.1145/2834892.2834896

[46] A. M. Terwilliger, G. N. Perdue, D. Isele, R. M. Patton, and S. R. Young, "Vertex reconstruction of neutrino interactions using deep learning," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 2275–2281.

[47] P. Balaprakash, A. Tiwari, S. M. Wild, and P. D. Hovland, "AutoMOMML: Automatic Multi-objective Modeling with Machine Learning," in *High Performance Computing: 31st International Conference, ISC High Performance 2016, Frankfurt, Germany, June 19-23, 2016, Proceedings*, M. J. Kunkel, P. Balaji, and J. Dongarra, Eds. Springer International Publishing, 2016, pp. 219–239.

[48] P. Balaprakash, Y. Alexeev, S. A. Mickelson, S. Leyffer, R. Jacob, and A. Craig, "Machine-learning-based load balancing for community ice code component in CESM," in *High Performance Computing for Computational Science – VECPAR 2014, Revised Selected Papers*, ser. Lecture Notes in Computer Science, M. Daydé, O. Marques, and K. Nakajima, Eds., vol. 8969. Springer International Publishing, 2015, pp. 79–91.

[49] F. Isaila, P. Balaprakash, S. M. Wild, D. Kimpe, R. Latham, R. Ross, and P. D. Hovland, "Collective I/O tuning using analytical and machine learning models," in *IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, 2015, pp. 128–137.

[50] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, 2006, vol. 1.

[51] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[52] B. Settles, "Active learning literature survey," *University of Wisconsin, Madison*, vol. 52, no. 55-66, p. 11, 2010.

[53] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[54] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.

[55] S. Yoo, H. Huang, and S. P. Kasiviswanathan, "Streaming spectral clustering," in *Data Engineering (ICDE), 2016 IEEE 32nd International Conference on*. IEEE, 2016, pp. 637–648.

[56] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. V. Le *et al.*, "Large scale distributed deep networks," in *Advances in neural information processing systems*, 2012, pp. 1223–1231.

[57] B. Recht, C. Re, S. Wright, and F. Niu, "Hogwild: A lock-free approach to parallelizing stochastic gradient descent," in *Advances in neural information processing systems*, 2011, pp. 693–701.

[58] M. Li, D. G. Andersen, J. W. Park, A. J. Smola, A. Ahmed, V. Josifovski, J. Long, E. J. Shekita, and B.-Y. Su, "Scaling distributed machine learning with the parameter server." in *OSDI*, vol. 1, no. 10.4, 2014, p. 3.

[59] M. Zinkevich, M. Weimer, L. Li, and A. J. Smola, "Parallelized stochastic gradient descent," in *Advances in neural information processing systems*, 2010, pp. 2595–2603.

[60] T. Kraska, A. Talwalkar, J. C. Duchi, R. Griffith, M. J. Franklin, and M. I. Jordan, "MLbase: A distributed machine-learning system." in *CIDR*, vol. 1, 2013, pp. 2–1.

[61] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, "cuDNN: Efficient primitives for deep learning," *arXiv preprint arXiv:1410.0759*, 2014.

[62] (2017) Intel data analytics acceleration library. [Online]. Available: https://software.intel.com/en-us/intel-daal

[63] J. Kepner, D. Bader, A. Buluç, J. Gilbert, T. Mattson, and H. Meyerhenke, "Graphs, matrices, and the GraphBLAS: Seven good reasons," *Procedia Computer Science*, vol. 51, pp. 2453–2462, 2015.

[64] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[65] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary algorithms made easy," *Journal of Machine Learning Research*, vol. 13, pp. 2171–2175, jul 2012.

[66] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.

[67] B. Baker, O. Gupta, N. Naik, and R. Raskar, "Designing neural network architectures using reinforcement learning," *arXiv preprint arXiv:1611.02167*, 2016.

[68] J. Bergstra, D. Yamins, and D. D. Cox, "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms," in *Proceedings of the 12th Python in Science Conference*, 2013, pp. 13–20.

[69] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *arXiv preprint arXiv:1603.06560*, 2016.

[70] B. Bischl, J. Richter, J. Bossek, D. Horn, J. Thomas, and M. Lang, "mlrMBO: A modular framework for model-based optimization of expensive black-box functions," *arXiv preprint arXiv:1703.03373*, 2017.

[71] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[72] A. Azad, G. Ballard, A. Buluç, J. Demmel, L. Grigori, O. Schwartz, S. Toledo, and S. Williams, "Exploiting multiple levels of parallelism in sparse matrix-matrix multiplication," *SIAM Journal on Scientific Computing (SISC)*, vol. 38, no. 6, pp. C624–C651, 2016.

[73] P. Koanantakool, A. Azad, A. Buluç, D. Morozov, S.-Y. Oh, L. Oliker, and K. Yelick, "Communication-avoiding parallel sparse-dense matrix-matrix multiplication," in *Proceedings of the IPDPS*, 2016.

[74] A. Buluç and J. R. Gilbert, "The Combinatorial BLAS: Design, implementation, and applications," *The International Journal of High Performance Computing Applications*, vol. 25, no. 4, pp. 496 – 509, 2011.

[75] A. Buluç, T. Mattson, S. McMillan, J. Moreira, and C. Yang, "Design of the GraphBLAS API for C," in *IEEE Workshop on Graph Algorithm Building Blocks, IPDPSW*, 2017.

[76] B. Dally, "GPU computing to exascale and beyond," 2010.

[77] J. S. Vetter and S. Mittal, "Opportunities for nonvolatile memory systems in extreme-scale high performance computing," *Computing in Science and Engineering*, vol. 17, no. 2, pp. 73–82, 2015.

[78] S. Mittal and J. S. Vetter, "A survey of software techniques for using non-volatile memories for storage and main memory systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. PP, no. 99, pp. 1–1, 2015.

[79] NVIDIA, "CUDA," [Online]. Available: https://developer.nvidia.com/cuda-zone, 2007, (Accessed July 17, 2017).

[80] OpenACC, "OpenACC: Directives for accelerators," 2015.

[81] OpenMP, "OpenMP reference," 1999.

[82] M. Bauer, S. Treichler, E. Slaughter, and A. Aiken, "Legion: Expressing locality and independence with logical regions," *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*, 2012.

[83] Kokkos, "The C++ Performance Portability Programming Model," [Online]. Available: https://github.com/kokkos/kokkos, 2014, (Accessed July 17, 2017).

[84] J. E. Stone, D. Gohara, and G. Shi, "OpenCL: A parallel programming standard for heterogeneous computing systems," *Computing in Science and Engineering*, vol. 12, no. 3, pp. 66–73, 2010.

[85] A. Sabne, P. Sakdhnagool, S. Lee, and J. S. Vetter, "Understanding portability of a high-level programming model on contemporary heterogeneous architectures," *Micro, IEEE*, vol. 35, no. 4, pp. 48–58, 2015.

[86] S. Lee and J. S. Vetter, "Early evaluation of directive-based GPU programming models for productive exascale computing," in *SC12: ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis*. Salt Lake City, Utah, USA: IEEE press, 2012.

[87] S. Ramos and T. Hoefler, "Capability models for manycore memory systems: A case-study with xeon phi knl," in *Parallel and Distributed Processing Symposium (IPDPS), 2017 IEEE International*. IEEE, 2017, pp. 297–306.

[88] A. Danalis, G. Marin, C. McCurdy, J. S. Meredith, P. C. Roth, K. Spafford, V. Tipparaju, and J. S. Vetter, "The scalable heterogeneous computing (SHOC) benchmark suite," in *ACM Workshop on General-Purpose Computation on Graphics Processing Units (GPGPU)*. Pittsburgh, Pennsylvania: ACM, 2010, pp. 63–74.

[89] S. Lee and J. S. Vetter, "OpenARC: Open accelerator research compiler for directive-based, efficient heterogeneous computing," in *ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC)*. Vancouver: ACM, 2014.

[90] S. Lee, J. Kim, and J. S. Vetter, "OpenACC to FPGA: A framework for directive-based high-performance reconfigurable computing," in *IEEE International Parallel & Distributed Processing Symposium (IPDPS)*. Chicago: IEEE, 2016.

[91] S. Lee, S.-J. Min, and R. Eigenmann, "OpenMP to GPGPU: a compiler framework for automatic translation and optimization," in *14th ACM SIGPLAN symposium on Principles and practice of parallel programming*. Raleigh, NC, USA: ACM, 2009.

[92] S. Lee and R. Eigenmann, "OpenMPC: Extended openMP programming and tuning for GPUs," in *2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE Computer Society, 2010, pp. 1–11.

[93] M. Martineau, J. Price, S. McIntosh-Smith, and W. Gaudin, *Pragmatic Performance Portability with OpenMP 4.x*. Cham: Springer International Publishing, October 2016, pp. 253–267. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-45550-1_18

[94] cuBLAS, "Dense Linear Algebra on GPUs," [Online]. Available: https://developer.nvidia.com/cublas, 2012, (Accessed July 17, 2017).

[95] Thrust, "A Parallel Algorithm Library," [Online]. Available: https://developer.nvidia.com/thrust, 2012, (Accessed July 17, 2017).

[96] LLVM, "The LLVM Compiler Infrastructure," [Online]. Available: http://llvm.org, 2007, (Accessed July 17, 2017).

[97] J. Kim, S. Lee, and J. S. Vetter, "An openACC-based unified programming model for multi-accelerator systems (poster)," in *Proceedings of the 20th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*. New York, NY, USA: ACM, 2015, p. 257258.

[98] N. Liu, J. Cope, P. Carns, C. Carothers, R. Ross, G. Grider, A. Crume, and C. Maltzahn, "On the role of burst buffers in leadership-class storage systems," in *IEEE 28th Symposium on Mass Storage Systems and Technologies (MSST)*. IEEE, 2012, pp. 1–11.

[99] D. Li, J. S. Vetter, G. Marin, C. McCurdy, C. Cira, Z. Liu, and W. Yu, "Identifying opportunities for byte-addressable non-volatile memory in extreme-scale scientific applications," in *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. Shanghai: IEEEE, 2012.

[100] S. Sreepathi, M. Grodowitz, R. Lim, P. Taffet, P. Roth, J. Meredith, S. Lee, D. Li, and J. S. Vetter, "Application characterization using oxbow toolkit and PADS infrastructure," in *First International Workshop on Hardware-Software Co-Design for High Performance Computing (colocated with SC14)*. New Orleans: ACM, 2014.

[101] I. Peng, R. Gioiosa, G. Kestor, P. Cicotti, E. Laure, and S. Markidis, "RTHMS: a tool for data placement on hybrid memory system," in *Proceedings of the 2017 ACM SIGPLAN International Symposium on Memory Management*. Barcelona, Spain: ACM, 2017, pp. 82–91.

[102] U. Bondhugula, A. Hartono, J. Ramanujam, and P. Sadayappan, "A practical automatic polyhedral parallelizer and locality optimizer," in *Proceedings of the ACM SIGPLAN 2008 Conference on Programming Language Design and Implementation (PLDI 08), Tucson, AZ (June 2008)*, 2008, pp. 101–113.

[103] T. Grosser, H. Zheng, R. Aloor, A. Simbürger, A. Größlinger, and L.-N. Pouchet, "Polly-polyhedral optimization in LLVM," in *Proceedings of the First International Workshop on Polyhedral Compilation Techniques (IMPACT)*, vol. 2011, 2011.

[104] P. Markthub, M. Belviranli, and J. Vetter, "DRAGON: Direct resource access for GPUs on NVM," in *submitted*, 2017.

[105] S. Perarnau, J. A. Zounmevo, B. Gerofi, K. Iskra, and P. Beckman, "Exploring data migration for future deep-memory many-core systems," in *IEEE Cluster Conference*, ser. Cluster'16, 2016.

[106] J. Kim, K. Sajjapongse, S. Lee, and J. S. Vetter, "Design and Implementation of Papyrus: Parallel Aggregate Persistent Storage," in *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, ser. IPDPS '17, 2017.

[107] J. Kim, S. Lee, and J. S. Vetter, "PapyrusKV: A High-Performance Parallel Key-Value Store for Distributed NVM Architectures," in *The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, ser. SC '17, 2017.

[108] M. Luo, D. K. Panda, K. Z. Ibrahim, and C. Iancu, "Congestion avoidance on manycore high performance computing systems," in *Proceedings of the 26th ACM International Conference on Supercomputing*, ser. ICS '12. New York, NY, USA: ACM, 2012, pp. 121–132. [Online]. Available: http://doi.acm.org/10.1145/2304576.2304594

[109] W. Lavrijsen and C. Iancu, "Application level reordering of remote direct memory access operations," in *Proceedings of IEEE International Parallel & Distributed Processing Symposium*, ser. IPDPS '17, 2017.

[110] S. Williams, "HPGMG-FV." [Online]. Available: http://crd.lbl.gov/departments/computer-science/PAR/research/hpgmg

[111] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, "NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations," *Computer Physics Communications*, 2010.

[112] A. Venkat, M. Shantharam, M. Hall, and M. M. Strout, "Non-affine extensions to polyhedral code generation," in *Proceedings of Annual IEEE/ACM International Symposium on Code Generation and Optimization*, ser. CGO '14. New York, NY, USA: ACM, 2014, pp. 185:185–185:194. [Online]. Available: http://doi.acm.org/10.1145/2544137.2544141

[113] A. Venkat, M. Hall, and M. Strout, "Loop and data transformations for sparse matrix codes," in *Proceedings of the 36th ACM SIGPLAN conference on Programming Language Design and Implementation*, ser. PLDI '15, Jun. 2015.

[114] R. Kaleem, A. Venkat, S. Pai, M. Hall, and K. Pingali, "Synchronization trade-offs in GPU implementations of graph algorithms," in *2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, May 2016, pp. 514–523.

[115] H. Zhang, A. Venkat, and M. Hall, "Compiler transformation to generate hybrid sparse computations," *2016 6th Workshop on Irregular Applications: Architecture and Algorithms (IA3)*, vol. 00, pp. 34–41, 2016.

[116] K. Ahmad, A. Venkat, and M. Hall, "Optimizing LOBPCG: Sparse matrix loop and data transformations in action," in *Languages and Compilers for Parallel Computing: 29th International Workshop*, ser. Editors: Chen Ding, John Criswell and Peng Wu. Springer-Verlag, 2016, pp. 218–231.

[117] A. Venkat, M. S. Mohammadi, J. Park, H. Rong, R. Barik, M. M. Strout, and M. Hall, "Automating wavefront parallelization for sparse matrix computations," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '16. Piscataway, NJ, USA: IEEE Press, 2016, pp. 41:1–41:12. [Online]. Available: http://dl.acm.org/citation.cfm?id=3014904.3014959

[118] C. Choudary, J. Godwin, J. Holewinski, D. Karthik, D. Lowell, A. Mametjanov, B. Norris, G. Sabin, and P. Sadayappan, "Stencil-aware GPU optimization of iterative solvers," *SIAM Journal on Scientific Computing*, vol. 35, no. 5, pp. S209–S228, October 2013. [Online]. Available: http://www.mcs.anl.gov/uploads/cels/papers/P3008-0712.pdf

[119] A. Mametjanov, D. Lowell, C.-C. Ma, and B. Norris, "Autotuning stencil-based computations on GPUs," in *Proceedings of IEEE Cluster 2012*, 2012, (Acceptance rate: 58/201=29%) Also available as Preprint ANL/MCS-P2094-0512. [Online]. Available: http://www.mcs.anl.gov/uploads/cels/papers/P2094-0512.pdf

[120] A. Hartono, B. Norris, and P. Sadayappan, "Annotation-based empirical performance tuning using Orio," in *Proceedings of the 23rd IEEE International Parallel & Distributed Processing Symposium*, Rome, Italy, 2009, also available as Preprint ANL/MCS-P1556-1008. [Online]. Available: http://www.mcs.anl.gov/uploads/cels/papers/P1556.pdf

[121] S. Williams, A. Watterman, and D. Patterson, "Roofline: An insightful visual performance model for floating-point programs and multicore architectures," *Communications of the ACM*, April 2009.

[122] T. Ligocki, "Roofline toolkit." [Online]. Available: https://bitbucket.org/berkeleylab/cs-roofline-toolkit

[123] NERSC, "Measuring arithmetic intensity." [Online]. Available: https://www.nersc.gov/users/application-performance/measuring-arithmetic-intensity

[124] D. Doerfer, J. Deslippe, S. Williams, L. Oliker, B. Cook, T. Kurth, M. Lobet, T. Malas, J.-L. Vay, and H. Vincenti, "Applying the roofline performance model to the Intel Xeon Phi Knights Landing processor," in *Intel Xeon Phi User Group Workshop (IXPUG)*, 2016.

[125] T. Barnes, B. Cook, J. Deslippe, D. Doerfler, B. Friesen, Y. H. He, T. Kurth, T. Koskela, M. Lobet, T. M. Malas, L. Oliker, A. Ovsyannikov, A. Sarje, J.-L. Vay, H. Vincenti, S. Williams, P. Carrier, N. Wichmann, M. Wagner, P. R. C. Kent, C. Kerr, and J. M. Dennis, "Evaluating and optimizing the NERSC workload on Knights Landing." in *PMBS*, 2016.

[126] Intel, "Intel advisor roofline." [Online]. Available: https://software.intel.com/en-us/articles/intel-advisor-roofline

[127] S. F. Siegel, M. Zheng, Z. Luo, T. K. Zirkel, A. V. Marianiello, J. G. Edenhofner, M. B. Dwyer, and M. S. Rogers, "CIVL: The Concurrency Intermediate Verification Language," in *SC15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. New York, NY, USA: ACM, Nov 2015, article no. 61, pages 1–12. [Online]. Available: http://doi.acm.org/10.1145/2807591.2807635

[128] M. Zheng, M. S. Rogers, Z. Luo, M. B. Dwyer, and S. F. Siegel, "CIVL: Formal verification of parallel programs," in *Proceedings of the 30th ACM/IEEE International Conference on Automated Software Engineering (ASE'15)*. New York, NY, USA: ACM, 2015, pp. 830–835.

[129] S. F. Siegel, "CIVL solutions to VerifyThis 2016 challenges," *ACM SIGLOG News*, vol. 4, no. 2, pp. 55–75, May 2017. [Online]. Available: http://doi.acm.org/10.1145/3090064.3090070

[130] S. Lee, D. Li, and J. S. Vetter, "Interactive program debugging and optimization for directive-based, efficient GPU computing," in *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. Phoenix: IEEE, 2014.

[131] D. Lee, A. Sim, J. Choi, and K. Wu, "Novel data reduction based on statistical similarity," in *SSDBM '16*. New York, NY, USA: ACM, 2016, pp. 21:1–21:12. [Online]. Available: http://doi.acm.org/10.1145/2949689.2949708

[132] M. Ainsworth, O. Tugluk, B. Whitney, and S. Klasky, "Multilevel techniques for compression and reduction of scientific data–the univariate case," *Computer Visualisation and Science (Submitted for publication)*, 2017.

[133] T. Lu, E. Suchyta, J. Choi, N. Podhorszki, S. Klasky, Q. Liu, D. Pugmire, M. Wolf, and M. Ainsworth, "Canopus: Enabling extreme-scale data analytics on big hpc storage via progressive refactoring," in *9th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 17)*. USENIX Association, 2017.

[134] P. Carns, R. Latham, R. Ross, K. Iskra, S. Lang, and K. Riley, "24/7 characterization of petascale I/O workloads," in *Proceedings of the First Workshop on Interfaces and Abstractions for Scientific Data Storage (IASDS)*, New Orleans, LA, Sep. 2009.

[135] R. Thakur, W. Gropp, and E. Lusk, "On implementing MPI-IO portably and with high performance," in *Proceedings of the 6th Workshop on I/O in Parallel and Distributed Systems*. ACM Press, May 1999, pp. 23–32.

[136] J. Li, W. keng Liao, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, and M. Zingale, "Parallel netCDF: A high-performance scientific I/O interface," in *Proceedings of SC2003*, Nov. 2003.

[137] "HDF5," `http://www.hdfgroup.org/HDF5/`.

[138] Department of Energy, "FY 2018 congressional budget request," 2017. [Online]. Available: https://energy.gov/sites/prod/files/2017/05/f34/FY2018BudgetVolume4_1.pdf

[139] M. Folk, G. Heber, Q. Koziol, E. Pourmal, and D. Robinson, "An overview of the HDF5 technology suite and its applications," in *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases*. ACM, 2011, pp. 36–47, software at http://www.hdfgroup.org/HDF5/.

[140] B. Behzad, J. Huchette, H. V. T. Luu, R. Aydt, Q. Koziol, S. Byna, M. Chaarawi, and Y. Yao, "Auto-tuning of parallel I/O parameters for HDF5 applications," in *SC'12*, 2012.

[141] J. Y. Choi, H. Abbasi, D. Pugmire, N. Podhorszki, S. Klasky, C. Capdevila, M. Parashar, M. Wolf, J. Qiu, and G. Fox, "Mining hidden mixture context with ADIOS-P to improve predictive pre-fetcher accuracy," in *E-Science (e-Science), 2012 IEEE 8th International Conference on*. IEEE, 2012, pp. 1–8.

[142] J. Dayal, D. Bratcher, G. Eisenhauer, K. Schwan, M. Wolf, X. Zhang, H. Abbasi, S. Klasky, and N. Podhorszki, "Flexpath: Type-based publish/subscribe system for large-scale science analytics," in *14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing(CCGRID '14)*, 2014.

[143] F. Zheng, H. Zou, G. Eisenhauer, K. Schwan, M. Wolf, J. Dayal, T.-A. Nguyen, J. Cao, H. Abbasi, S. Klasky *et al.*, "FlexIO: I/O middleware for location-flexible scientific data analytics," in *Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on*. IEEE, 2013, pp. 320–331.

[144] C. Docan, M. Parashar, and S. Klasky, "DataSpaces: an interaction and coordination framework for coupled simulation workflows," *Cluster Computing*, vol. 15, no. 2, pp. 163–181, Feb. 2011.

[145] E. DAzevedo, S. Abbott, T. K. P. Worley, S.-H. Ku, S. Ethier, E. Yoon, M. Shephard, R. Hager, J. Lang, J. Choi, N. Podhorszki, S. Klasky, M. Parashar, and C.-S. Changh, *Enabling Kinetic Study of Multiscale Edge Turbulent Transport in ITER*. Will appear in *Exascale Scientific Applications*, eds: T. P. Straatsma, K. B. Antypas and T. J. Williams: Chapman & Hall/CRC, 2017, pp. 525–547.

[146] E. Deelman, K. Vahi, G. Juve, M. Rynge, S. Callaghan, P. J. Maechling, R. Mayani, W. Chen, R. F. da Silva, M. Livny, and K. Wenger, "Pegasus, a workflow management system for science automation," *Future Generation Computer Systems*, vol. 46, pp. 17–35, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X14002015

[147] J. Freire, D. Koop, F. S. Chirigati, and C. T. Silva, "Reproducibility using VisTrails," *Implementing Reproducible Research*, vol. 33, 2014.

[148] K. Wu, E. N. Coviello, S. Flanagan, M. Greenwald, X. Lee, A. Romosan, D. P. Schissel, A. Shoshani, J. Stillerman, and J. Wright, "MPO: A system to document and analyze distributed heterogeneous workflows," in *International Provenance and Annotation Workshop*. Springer, 2016, pp. 166–170.

[149] I. Altintas, C. Berkley, E. Jaeger, M. Jones, B. Ludascher, and S. Mock, "Kepler: an extensible system for design and execution of scientific workflows," in *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on*. IEEE, 2004, pp. 423–424.

[150] P. Mouallem, R. Barreto, S. Klasky, N. Podhorszki, and M. Vouk, "Tracking files in the Kepler provenance framework," in *International Conference on Scientific and Statistical Database Management*. Springer, 2009, pp. 273–282.

[151] J. Dayal, J. Lofstead, G. Eisenhauer, K. Schwan, M. Wolf, H. Abbasi, and S. Klasky, "SODA: Science-driven orchestration of data analytics," in *e-Science (e-Science), 2015 IEEE 11th International Conference on*. IEEE, 2015, pp. 475–484.

[152] C. Docan, F. Zhang, T. Jin, H. Bui, Q. Sun, J. Cummings, N. Podhorszki, S. Klasky, and M. Parashar, "ActiveSpaces: Exploring dynamic code deployment for extreme scale data processing," *Concurrency and Computation: practice and Experience*, vol. 27, no. 14, pp. 3724–3745, 2015.

[153] Y. Tian, S. Klasky, H. Abbasi, J. Lofstead, R. Grout, N. Podhorszki, Q. Liu, Y. Wang, and W. Yu, "EDO: improving read performance for scientific applications through elastic data organization," in *Cluster Computing (CLUSTER), 2011 IEEE International Conference on*. IEEE, 2011, pp. 93–102.

[154] K. Wu, S. Ahern, E. W. Bethel, J. Chen, H. Childs, E. Cormier-Michel, C. Geddes, J. Gu, H. Hagen, B. Hamann, W. Koegler, J. Lauret, J. Meredith, P. Messmer, E. Otoo, V. Perevoztchikov,

A. Poskanzer, Prabhat, O. Rübel, A. Shoshani, A. Sim, K. Stockinger, G. Weber, and W.-M. Zhang, "FastBit: Interactively searching massive data," in *SciDAC 2009*, 2009, lBNL-2164E. [Online]. Available: http://www.iop.org/EJ/abstract/1742-6596/180/1/012053

[155] K. Stockinger, J. Cieslewicz, K. Wu, D. Rotem, and A. Shoshani, *Using Bitmap Indexing Technology for Combined Numerical and Text Queries*, ser. Annals of Information Systems. Springer, 2008, vol. 3, pp. 1–23, preprint appeared as LBNL Tech Report LBNL-61768.

[156] K. Madduri and K. Wu, "Massive-scale RDF processing using compressed bitmap indexes," in *SS-DBM*. Springer, 2011, pp. 470–479.

[157] M. Alhamad, T. Dillon, and E. Chang, *A Survey on SLA and Performance Measurement in Cloud Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 469–477. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-25106-1_4

[158] A. Goodman, A. Pepe, A. W. Blocker, C. L. Borgman, K. Cranmer, M. Crosas, R. Di Stefano, Y. Gil, P. Groth, M. Hedstrom *et al.*, "Ten simple rules for the care and feeding of scientific data," *PLoS computational biology*, vol. 10, no. 4, p. e1003542, 2014.

[159] C. Wood, S. Sane, D. Ellsworth, A. Gimenez, K. Huck, T. Gamblin, and A. Malony, "A scalable observation system for introspection and in situ analytics," in *Extreme-Scale Programming Tools (ESPT), Workshop on*. IEEE, 2016, pp. 42–49.

[160] B. Dong, S. Byna, K. Wu, Prabhat, H. Johansen, J. N. Johnson, and N. Keen, "Data elevator: Low-contention data movement in hierarchical storage system," in *2016 IEEE 23rd International Conference on High Performance Computing (HiPC)*, Dec 2016, pp. 152–161.

[161] T. Wang, A. Moody, Y. Zhu, K. Mohror, K. Sato, T. Islam, and W. Yu, "MetaKV: A key-value store for metadata management of distributed burst buffers," in *Parallel and Distributed Processing Symposium (IPDPS), 2017 IEEE International*. IEEE, 2017, pp. 1174–1183.

[162] D. Lee, A. Sim, J. Choi, and K. Wu, "Expanding statistical similarity based data reduction to capture diverse patterns," in *2017 Data Compression Conference (DCC)*, April 2017, pp. 445–445.

Appendix 4: Facilities and Other Resources

LBNL Compute Facilities and Resources

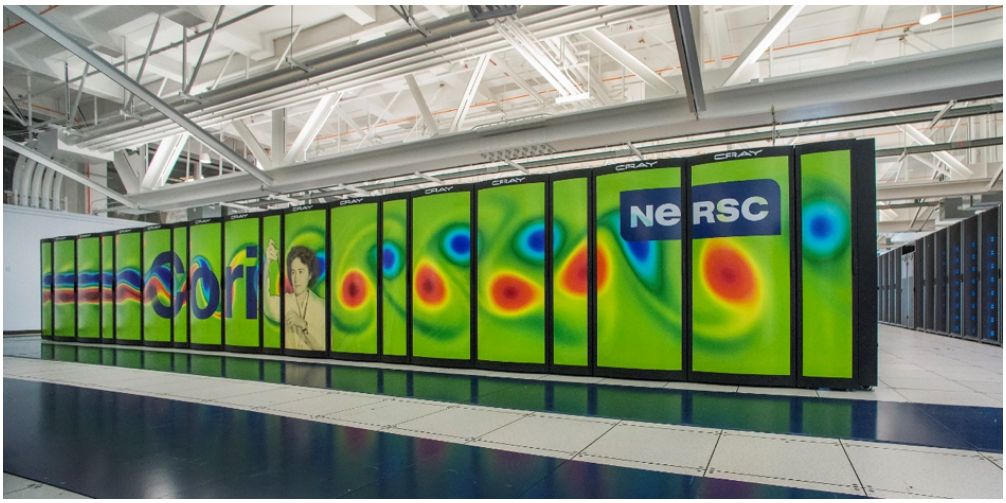
Lawrence Berkeley National Laboratory is the leading provider of computing and networking resources supporting the DOE Office of Science's research mission. Berkeley Lab researchers have access to leading-edge computing platforms and services at the National Energy Research Scientific Computing Center (NERSC) and have 100 Gbps connectivity to other national labs and institutions via ESnet, DOE's Energy Sciences Network, both of which are managed by Berkeley Lab. The lab also manages several departmental clusters.

At the end of 2015, NERSC moved from its facility in Oakland to the newly built Shyh Wang Hall on the main Berkeley Lab campus. The move re-unites NERSC with the staff of ESnet and the Computational Research Division and is expected to lead to more collaboration.

NERSC

In addition to providing world-class supercomputers, NERSC staff provide expert support to ensure users make the most efficient and effective use of the resources. This support has made NERSC one of the most scientifically productive centers in the world— in 2016, NERSC users reported more than 2,000 refereed papers based on work performed at the center and NERSC staff contributed some 120 papers to scientific journals and conferences.

Coinciding with the move was the installation of NERSC's newest supercomputer, an Intel-based Cray XC40 with a theoretical peak performance of 30 petaflop/s and will deliver 10x the sustained computing capability of NERSC's recently retired Hopper system, a Cray XE6 supercomputer. Named "Cori" in honor of biochemist Gerty Cori, the first American woman to receive a Nobel Prize in science, the system has a number of new features that will benefit data-intensive science. Cori was delivered in two phases. The first phase—also known as the Data Partition—was installed in late 2015 and comprises 12 cabinets and more than 1,600 Intel Xeon "Haswell" compute nodes. It was customized to support data-intensive science and the analysis of large datasets through a combination of hardware and software configurations and queue policies. The second phase of Cori, installed in summer 2016, added another 52 cabinets and more than 9,300 nodes with second-generation Intel Xeon Phi processors (code-named Knights Landing, or KNL for short), making Cori the largest supercomputing system for open science based on KNL processors. With 68 active physical cores on each KNL and 32 on each Haswell processor, Cori has almost 700,000 processor cores. The two phases of Cori are integrated via the Cray Aries interconnect, which has a dragonfly network topology that provides scalable bandwidth.



Cori features a Burst Buffer based on the Cray DataWarp technology. The Burst Buffer, a 1.5 PB layer of NVRAM storage, sits between compute node memory and Cori's 30-petabyte Lustre parallel scratch file

system. The burst buffer provides about 1.5 TB/sec of I/O bandwidth, more than twice that of the scratch file system. NERSC has also added software defined networking features to Cori to more efficiently move data in and out of the system, giving users end-to-end connectivity and bandwidth for real-time data analysis, and a real-time queue for time-sensitive analyses of data.

NERSC's other large system is Edison, a Cray XC30 supercomputer that is the first Cray supercomputer with Intel processors, a new Aries interconnect and a dragonfly topology. Installed in 2013, Edison was designed to optimize data motion—the primary bottleneck for many of our applications—as opposed to peak speed. It has very high memory bandwidth, interconnect speed and bisection bandwidth. In addition, each node has twice the memory of many leading systems. This combination of fast data motion and large memory per node make it well suited for both our traditional HPC workload and newly emerging data intensive applications. Edison features 124,608 compute cores for running scientic applications, 332 terabytes of memory, and 7.6 petabytes of online disk storage with a peak I/O bandwidth of 144 gigabytes (GB) per second. Edison has a theoretical peak performance of 2.39 petaflops/second.

NERSC's research in data-intensive computing is grounded in their operation of a major production facility, the PDSF (Parallel Distributed Systems Facility). PDSF is a networked distributed computing environment used to meet the detector simulation and data analysis requirements of large scale Physics, High Energy Physics and Astrophysics and Nuclear Science investigations.

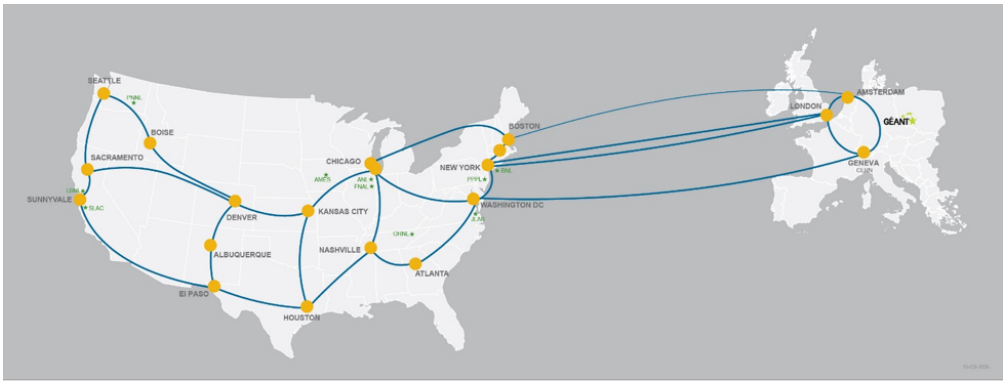
All NERSC systems are connected to the NERSC Global Filesystem (NGF), a collection of center-wide file systems, based on IBM's GPFS, available on nearly all systems at the facility. The several different file systems comprising NGF, including one providing a common login user environment for all our systems, one for sharing data among collaborators on a science project or team, and one for high bandwidth short term storage across systems at the facility. The main focus of NGF is data sharing, ease of workflow management (i.e., not moving data around or maintaining unnecessary copies of data), and data analysis.

Finally, these systems are connected to a High Performance Storage System (HPSS) for archival storage. NERSC's HPSS system currently contains more than 42 petabytes, making it one of the largest unclassified archival storage systems in the world.

ESnet

Access to Berkeley Lab's computational and experimental facilities from anywhere in the U.S. or the world is available through ESnet, which provides a 100-gigabit Ethernet backbone connection between NERSC, LBNL and other DOE national laboratories. In December 2014, ESnet extended its connectivity to Europe

with three 100 Gbps links and one 40 Gbps connection. ESnet also provides major backbone links including peering with hundreds of domestic and international research and education networks.



Argonne National Laboratory Facilities and Resources

Argonne National Laboratory provides facilities for computing resources; data and networking; and data analytics and visualization. Computing Resources The Joint Laboratory for System Evaluation (JLSE) is a collaboration between the Mathematics and Computer Science (MCS) Division and the Argonne Leadership Computing Facility (ALCF) with the aim of evaluating future high-performance computing platforms, developing system software and measuring power. JLSE hosts more than 20 different cutting-edge hardware platforms, including IBM Power8, Intel Haswell-EX E7-8867v3.

The Argonne Leadership Computing Facility (ALCF) features a 10-petaflop IBM Blue Gene/Q system dedicated to breakthrough science and engineering. Mira has 48 racks, with 1,024 nodes per rack and16 cores per node for a total of 786,432 cores; 16 gigabytes of RAM per node for a total of 768 terabytes; 35 PB of storage; and 384 I/O nodes. As of November 2014, Mira ranked number 5 in the Linpack Top500 list. In late 2016, the Argonne Leadership Computing Facility (ALCF) deployed Theta, a next-generation Cray/Intel system based on the 2nd Generation Intel Xeon Phi processor. The system has over 2500 nodes, each with a KNL 60-or-more-core processor having up to 16 GB of high-bandwidth in-package memory (IPM) and 192 GB of DDR4 RAM. Aurora will have 3rd Generation Intel Xeon Phi processors (code name: Knights Hill (KNH)). It will have over 50,000 nodes and over 7 PB total combination of memory.

ALCF also hosts a 504-node cloud computing system, Magellan, with a theoretical peak performance of 40 TF/s. Each node has two quad-core Nehalem 2.66 GHz CPUs, 24 GB RAM, and 500 GB of disk storage. The system also features four login nodes and eight file servers with a total storage capacity of 160 TB. The nodes are interconnected by using QDR InfiniBand, and the whole system is connected to the outside world at 10 Gbps. A part of the Magellan system is also equipped with dual Fermi 2070 GPUs. Apart from the ALCF facilities, Argonne also hosts several other resources, including Blues in the Laboratory Computing Resource Center. Blues features 310 state-of-the-art nodes and 4,900 processors and offers a peak performance of 103 TF.

Through the **University of Chicago/Argonne Computation Institute (CI)**, Argonne has access to the XSEDE system. XSEDE is composed of multiple partner institutions known as Service Providers, each of which contributes one or more allocatable services. Resources include various supercomputers and high-end visualization and data analysis resources. The CI also has a 151-teraflop, 18,000-core Cray XE6 system Beagle to support computation, simulation, and data analysis for the biomedical research community. Data Storage The ALCF's data storage system is used to retain the data generated by simulations and visualizations. Disk storage provides intermediate-term storage for active projects, offering a means to access, analyze, and share simulation results. Tape storage is used to archive data from completed projects.

BNL Compute Facilities and Resources

Brookhaven National Laboratory has a strong history in the successful operation of large-scale computational science, data management, and analysis infrastructure. This expertise and tools developed at the Lab have been key factors in the success of the scientific programs at RHIC, NSLS-II, and CFNall DOE Office of Science User Facilitiesand also in biological, atmospheric, and energy systems science. These capabilities have also been a crucial part of the Lab's participation in national and international research collaborations like the ATLAS experiment at the LHC, US QCD, and the Atmospheric Radiation Measurement Program etc.

Computing Resources

RHIC ATLAS Computing Facility - RHIC ushered in high throughput, data intensive computing as a BNL capability 15 years ago, and ATLAS computing at BNL has built on that. The RHIC and ATLAS Computing Facility (RACF) is a unified facility delivering leveraged, cost-effective computing to both RHIC (NP) and ATLAS (HEP). RCF provides 90% of the computing capacity for PHENIX and STAR data analysis. The ATLAS Tier-1 facility is the largest of 11 Tier-1 centers worldwide and contributes a share of 23% of the worldwide computing capacity to ATLAS data analysis at 99% service availability (12 month average). Besides CERN the BNL Atlas Center is the most important ATLAS data repository: delivers 200 TB/day to $>$100 data centers around the world. Together with the Physics Application Software group these resources and capabilities make BNL one of the largest Big Data / High Throughput Computing (HTC) resources and expertise pools in US science.

- RACF occupies 15,000 sq. ft.
- 250 Computer Racks + 9 Robotic Linear Magnetic Tape Libraries (85k tape cartridge cells, 200 PB Capacity)
- HPSS based tape management system, one of the largest instances worldwide
- Scalable parallel filesystem with GPFS (currently deployed instance scales to 100 GB/s)
- 60,000 X86 based compute cores, 125 TB memory, 35 PB Disk
- Bisection bandwidth of 3 Terabits/sec between storage and compute resources
- 800 kW Power Utilization (7 MWh per year)

Institutional Clusters - BNL currently operates two institutional cluster (IC):

- A 108-node cluster (3888 computing cores in total), each node with a dual Intel Xeon E5-2695v4 (Broadwell) CPU, dual NVidia K80 GPUs, SAS-based local storage, and 256 GB of memory. This cluster is inter-connected with a non-blocking Infiniband EDR fabric.
- A 72-node cluster (4608 computing cores in total), each node with a single Intel Xeon Phi 7230 (Knight's Landing) CPU, SSD-based local storage, and 192 GB of memory. This sub-cluster is inter-connected with an Intel Omni-Path dual-rail, non-blocking fabric, capable of up to 22 Gb/s aggregate bi-directional bandwidth.

A GPFS-based storage system with a bandwidth of up to 24 Gb/s is connected to the IC.

IBM Blue Gene Systems - In 2007 BNL acquired the 18 rack New York Blue/L and in 2009 the 2 rack New York Blue/P. New York Blue/L debuted at number 5 on the June 2007 Top 500 list of the world's fastest computers. New York Blue/P debuted on the June 2009 Top 500 list at number 250. New York Blue/L was decommissioned in January, 2014 and New York Blue/P was decommissioned in October, 2015. With a peak performance of 103.22 teraflops (trillion floating-point calculations per second) and 27.9 teraflops respectively, New York Blue/L and New York Blue/P allowed computations critical for research in biology, medicine, material science, nanoscience, renewable energy, climate science, finance and technology. Today BNL operates a Blue Gene Q as part of three facilities/collaborations, NYCCS, RIKEN, and LQCD. 3.5 Racks, 1PB of NFS Storage 3584 IBM Blue GeneQ nodes, 64512 cores, 57344 of which are compute cores,

57344 GB of RAM. Total performance capability is 700 TFlops. These systems were acquired in the fall of 2011. One rack of the Blue Gene Q system was benchmarked for data intensive applications and debuted at number 6 on June 2012 Graph 500 List. The Blue GeneQ is connected to 1PB of NFS storage.

Storage

BNL currently provides one of the world's largest data archive and analysis capabilities for scientific data. With now over 100 PB of archived data, BNL hosts the 2nd largest archive in the US (just behind NOAA) and the 4th largest in the world. BNL annually ingests 30 PB, distributes 32 PB and processes 150 PB of data for a worldwide user community based in national laboratories, universities and other research institutes. In addition BNL provides around 50PB of fast disk storage for its users, as well as various storage technologies including object store are deployed providing resilience and remote access through the most popular protocols (incl. Grid FTP and its Globus Connect Server packaging). The BNL tape system is optimized for high bandwidth input and output allowing for fast retrieving of data while archiving at the same time. The tape system has proven to be very reliable; it is designed to meet the requirements of data preservation for the RHIC (Relativistic Heavy Ion Collider at BNL) program; where data access has to be guaranteed for at least 30 years. It is common practice to retrieve data archived more than a decade ago. File size on the tape system ranges from a few MB to several GB. Storage are organized in several layers in order get rapid access to most popular data.

Networking

Fast external networking (200GB/s set to be upgraded to 300 GB/s) and ESNET support on site enables BNL computer service users to effectively ingest and distribute their research data. BNL builds hereby on its experience of serving the worldwide LHC Atlas experimental community, as its largest data center outside CERN, it delivers currently over 200TB/day peak to over 100 data centers in the US including LCFs like ANL, NERSC and around the world. Internally all BNL computing capabilities are part of a high speed network (terabit scale) that ensures fast data access for data processing and analysis in preparation for archival or retrieval.

Los Alamos National Laboratory: Facilities and Resources

Los Alamos National Laboratory has a number of state-of-the-art computing resources that can be accessed via the Turquoise Network; a network designed to enhance collaboration between Los Alamos and external scientific institutions. Major computational systems on the Turquoise network at Los Alamos National Laboratory include:

Grizzly is an Intel Xeon Broadwell cluster running Linux, and is available to Institutional Computing and Advanced Simulation and Computing users. Its consists of 1490 nodes, each with 36 cores, for a total of 53,640 cores. It has 128GB per compute node, for a total of 191 TB memory. It has Intel OmniPath interconnect, and peak of 1,806 TFlops with 15.2 PB Lustre storage.

Lightshow is an Intel Xeon X5650/2x nVidia Quadro GPU cluster running Linux, and is available to Institutional Computing and Advanced Simulation and Computing users. It consists of 16 nodes, with 12 cores per node for a total of 192 total cores. It has 96 GB of memory per compute node, for a total of 9.86 TB of memory. It has Mellanox Inifiniband Fat-Tree interconnect, and 4.0 TFlops peak performance, with 15.2 PB Lustre storage.

Moonlight is an HPC Linux Cluster in the Turquoise network available to Institutional Computing and Advanced Simulation and Computing users. It consists of 308 compute nodes, each with two eight-core Intel Xeon E5-2670 2.6 GHz processors. Each compute node also has two NVIDIA Tesla M2090 cards connected to PCIe-2.0 x16 slots. Its theoretical peak performance is 488 TFlops.

Pinto is an Intel Xeon E5-2670 cluster running Linux, and is available to Institutional Computing and Advanced Simulation and Computing users. It consists of 154 nodes, with 16 cores per node for a total of 2,464 total cores. It has 32 GB of memory per compute node, for a total of 4.9 TB of memory. It has Qlogic InfiniBand Fat-Tree interconnect, and 51.3 TFlops peak performance, with 15.2 PB Lustre storage.

Wolf is an Intel Xeon E5-2670 cluster running Linux, and is available to Institutional Computing and Advanced Simulation and Computing users. It consists of 616 nodes, with 16 cores per node for a total of 9,856 total cores. It has 64 GB of memory per compute node, for a total of 39.4 TB of memory. It has Qlogic InfiniBand Fat-Tree interconnect, and 205 TFlops peak performance, with 15.2 PB Lustre storage.

In addition, members of this proposal will have access to resources of the Applied Computer Science group at Los Alamos, including the **Darwin cluster**. This heterogeneous 300-node cluster includes a wide variety of node configurations, including some nodes with NVIDIA GPUs, large main memory (3TB), KNL CPUs, POWER8+ CPUs with NVLINK 1.0 and Pascal GPUs, ARM64 CPUs, local NVMe SSD storage, and numerous other server configurations.

Description of Facilities and Resources

Livermore Computing | Lawrence Livermore National Laboratory (LLNL)

LLNL computational scientists are supported by Livermore Computing (LC), which delivers a balanced High Performance Computing (HPC) environment with constantly evolving hardware resources and a wealth of HPC expertise in porting, running, and tuning real-world, large-scale applications. Currently LC delivers multiple petaflops of compute power, massive shared parallel file systems, powerful data analysis platforms, and archival storage capable of storing hundreds of petabytes of data. This balanced hardware environment supports key collaborations between LLNL applications developers and LC experts on the creation, production use, and performance monitoring and analysis of results of HPC parallel applications in a wide variety of scientific disciplines.

In addition to general access for development and debugging, researchers access allocations on these resources through several competitive internal programs. With these allocations, science and technology directorates sustain and enhance institutional capabilities to help ensure mission successes. Such efforts benefit the Laboratory as a whole by developing the expertise of the next generation of scientists and engineers who will, in turn, use these capabilities for continued growth of Laboratory programs. Furthermore, researchers with "work-for-others" (WFO) projects and external collaborators needing HPC access can purchase cycles on the M&IC systems.

Major systems include the 20 PFLOP Sequoia system with a 55 PB file system, the 5 PFLOP Vulcan system, the Jade and Quartz systems at 3 PFLOPs each, the 970 TFLOP Zin system, the 431 TFLOP Cab system, and additional large multi-core, multi-socket Linux clusters with a variety of processor types, ranging from IBM PowerPC to Intel Broadwell processors, and nVidia Graphical Processing Units (GPUs) for some platforms. In total, more than 135,000 nodes with more than 2,300,000 cores and more than 3.2PB of memory are available across the production LC systems on two networks.

Computational scientists may also take advantage of several testbeds for evaluating next generation hardware and software. Researchers use these testbeds to investigate hardware advances in areas such as multi-core processors, neuromorphic computing, networking technologies, I/O, GPUs, memory, and power-aware HPC (via a dedicated power lab), as well as investigations of software technologies. In addition, LC hosts production collaboration environments that facilitate the sharing of scientific data among international research groups, including the Green Data Oasis and the Green Collaboration Environment with a Linux cluster and storage resources available to collaborators worldwide.

Several facilities house the simulation infrastructure at LLNL. The largest (LEED-certified) simulation facility offers 48,000ft^2 and 37.5MW of power for systems and peripherals, and additional power for the associated cooling system. Engineering and facilities staff maintains it in a physically secure environment.

The balanced LLNL simulation environment includes Lustre multi-cluster file systems, HPSS-based archival resources, a Federated Ethernet networking infrastructure, advanced visualization resources, and a rich tool environment, as described on the reverse.

Lustre: LLNL contributes to the development of and supports the open source Lustre parallel file systems, which are mounted across multiple compute clusters and delivers high-performance, global access to data.

High Performance Archival Storage: LC provides high-performance archival storage services via High Performance Storage System (HPSS). A world-class array of hardware integrated beneath HPSS includes disk arrays, tape subsystems, mover nodes, networks, robotics, and petabytes of media.

LC Networking: LC's simulation environment includes a Federated Ethernet networking infrastructure as well as Infiniband SANs. LC employs InfiniBand and Omni-Path fabrics for high-speed interconnects. Testbed work includes evaluation of next generation networking equipment.

Visualization facilities: Data analysis resources include the 324-node Max machine with more than 82 TB of memory and the 162-node Surface machine with 41 TB of memory and 316 Kepler K40 GPUs. These interactive data analysis machines are equipped with high-speed access to Lustre and local NFS storage. LC operates several visualization theaters, ranging from auditoriums with PowerWalls to smaller collaboration spaces.

HPC Tool Environment: LC provides a stable, usable, leading-edge parallel application development environment that significantly increases the productivity of applications developers by enabling better scalable performance and enhancing application reliability. The tool environment includes high-performance compilers, debuggers, analyzers, editors, and locally developed custom libraries and application packages for software development. Through collaborations with vendors and other third party software developers, LC ensures a complete environment that meets the needs of today's code developers while steering their code development to exploit emerging technologies.

Primary Production Computing Platforms

Sequoia and **Vulcan**: Sequoia is a 20 PFLOP system based on IBM BlueGene Q (BG/Q) technology. Rated the most powerful supercomputer in the world in June, 2012 shortly after its arrival, Sequoia has 1.6 petabytes of memory, 96 racks, 98,304 compute nodes, and 1.6 million cores. Vulcan is a BG/Q system one-quarter the size of Sequoia, running on the unclassified network. The BG/Q machines are based on the POWER architecture.

Jade and Quartz: Identically configured Penguin Computing systems, Jade and Quartz each have 2,688 nodes and 343 TB of memory and a peak speed of 3.2 PFLOPs. Each system is a 14 Scalable Unit multiprocessor cluster with Intel Broadwell processors. **RZTopaz** is a smaller 4SU machine that is liquid cooled.

Zin: A Cray system with 2,916 nodes and 93.3 TB of memory, Zin has a peak speed of 970 TFLOPs and uses Intel Sandy Bridge processors.

Cab: A Cray system with 1,296 Intel Sandy Bridge nodes and 41.5 TB of memory, Cab has a peak speed of 431 TFLOPs. The second-most powerful machine on LC's unclassified network, next to Vulcan, Cab has Intel Sandy Bridge processors.

Syrah: This Cray system is a multi-core, multi-socket Linux cluster. Syrah has 324 nodes and 20 TB of memory, clocking in at a peak speed of 108 TFLOPs.

Ray, Shark and **RZManta:** IBM hybrid Power8/GPU based machines. The 598 TFLOP Shark and RZ Manta machines each have 36 nodes with every node containing 20 cores, 4 GPUs and 256GB of memory. The 896 TFLOP Ray machine has 54 of these hybrid nodes combined with 1.6TB of NVMe flash storage per node.

Catalyst: A Cray cluster with 304 nodes, 7,776 cores, 41.5 TB of memory and 243 TB of NVRAM this system is part of a partnership between LLNL, Intel and Cray focusing on HPC big data technologies, architectures and applications.

For a complete list of more than 20 production compute platforms supported by LC, see the **Livermore Computing Systems Summary on hpc.llnl.gov.**

Institutional Resources and Facilities

LLNL researchers benefit from an institutional IT infrastructure that provides desktop support and experts in server technologies. The latter includes virtualization expertise that has been applied to provide multiple operating systems on shared resources and to create a wide variety of virtual machines to leverage resources servers across LLNL. Networking service is also provided by an enterprise team. Connections into LLNL include ESnet and a wide variety of programmatic networks.

Prepared by LLNL under Contract DE-AC52-07NA27344.

Description of Facilities and Resources

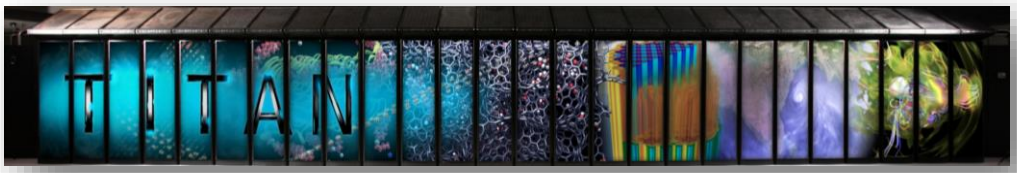
Oak Ridge National Laboratory
and the
UT-ORNL Joint Institute for Computational Sciences

1. Oak Ridge National Laboratory

Computer Facilities. The Oak Ridge National Laboratory (ORNL) hosts three petascale computing facilities: the Oak Ridge Leadership Computing Facility (OLCF), managed for DOE; the National Institute for Computational Sciences (NICS), operated for the National Science Foundation (NSF); and the National Climate-Computing Research Center (NCRC), formed as collaboration between ORNL and the National Oceanographic and Atmospheric Administration (NOAA) to explore a variety of research topics in climate sciences. Each of these facilities has an experienced operational and engineering staff comprising groups in high-performance computing (HPC) operations, technology integration, user services, scientific computing, and application performance tools. The ORNL computer facility staff provides continuous operation of the centers and immediate problem resolution. On evenings and weekends, computer room operators provide front-line problem resolution for users with additional user support and system administrators on-call for problems that require escalation.

Other Facilities. The Oak Ridge Science and Technology Park at ORNL was the nation's first technology park on the campus of a national laboratory. The technology park is available for private sector companies that are collaborating with research scientists. This facility is used to establish new companies from emerging technologies developed at ORNL.

1.1 Primary Systems



Titan is a Cray XK7 system consisting of 18,688 AMD sixteen-core Opteron™ processors providing a peak performance of more than 3.3 petaflops (PF) and 600 terabytes (TB) of memory. A total of 512 service input/output (I/O) nodes provide access to the 32 petabytes (PB) "Spider II" Lustre parallel file system at more than 1 terabyte (TB/s). External login nodes provide a powerful compilation and interactive environment using dual-socket, twelve-core AMD Opteron processors and 256 GB of memory. Each of the 18,688 Titan compute nodes is paired with an NVIDIA Kepler graphics processing unit (GPU) designed to accelerate calculations. With a peak performance per Kepler accelerator of more than 1TF, the aggregate performance of Titan exceeds 27PF. Titan is the Department of Energy's most powerful open science computer system and is available to the international science community through the INCITE program, jointly managed by DOE's Leadership Computing Facilities at Argonne and Oak Ridge National Laboratories and through the DOE Office of Science's ALCC program.

Gaea is a Cray XC40 system consisting of 2,992 compute

socket R3 Intel 16-core Haswell processors, providing 47,872 (95,744 logical, when using Intel Hyper-Threading) compute cores, 93.5 TB of double data rate 4 (DDR4) memory, and a peak performance of 1.76 petaflops (PF) all within a small fraction of the footprint of the previously decommissioned XE6 systems. The XC40 uses the Cray Aries Interconnect with the Dragonfly network topology. This provides a higher bandwidth and lower latency interconnect than that of previous systems.

Gaea is supported by a series of external login nodes and a single Lustre file system. The F1 file system is based on more than 2,000 Nearline-SAS drives and provides just under 6 PB (formatted) space to Gaea as well as data transfer capability to the NOAA archive. Gaea is one of NOAA's most powerful computer systems and is available to the climate research community through the Department of Commerce/NOAA allocation mechanisms.

The ORNL Institutional Cluster (OIC) consists of two phases. The original OIC consists of a bladed architecture from Ciara Technologies called VXRACK. Each VXRACK contains two login nodes, three storage nodes, and 80 compute nodes. Each compute node has dual Intel 3.4 GHz Xeon EM64T processors, 4 GB of memory, and dual gigabit Ethernet interconnects. Each VXRACK and its associated login and storage nodes are called a block. There are a total of nine blocks of this type. Phase 2 blocks were acquired and brought online in 2008. They are SGI Altix machines. There are two types of blocks in this family.

- Thin nodes (3 blocks). Each Altix contains 1 login node, 1 storage node, and 28 compute nodes within 14 chassis. Each node has eight cores and 16 GB of memory. The login and storage nodes are XE240 boxes from SGI. The compute nodes are XE310 boxes from SGI.
- Fat nodes (2 blocks). Each Altix contains 1 login node, 1 storage node, and 20 compute nodes within 20 separate chassis. Each node has eight cores and 16 GB of memory. These XE240 nodes from SGI contain larger node-local scratch space and a much higher I/O to this scratch space because the space is a volume from four disks.

EOS is a 744-node Cray XC30 cluster with a total of 47.6 TB of memory. The processor is the Intel® Xeon® E5-2670. Eos uses Cray's Aries interconnect in a network topology called Dragonfly. Aries provides a higher bandwidth and lower latency interconnect than Gemini. Support for I/O on Eos is provided by (16) I/O service nodes. The system has (2) external login nodes.

The compute nodes are organized in blades. Each blade contains (4) nodes connected to a single Aries interconnect. Every node has (64) GB of DDR3 SDRAM and (2) sockets with (8) physical cores each. Intel's Hyper-threading (HT) technology allows each physical core to work as two logical cores so each node can functions as if it has (32) cores. Each of the two logical cores can store a program state, but they share most of their execution resources. Each application should be tested to see how HT impacts performance before HT is used. The best candidates for a performance boost with HT are codes that are heavily memory-bound. The default setting on Eos is to execute without HT, so users must invoke HT with the -j2 option to aprun.

In total, the Eos compute partition contains 11,904 traditional processor cores (23,808 logical cores with Intel Hyper-Threading enabled), and 47.6 TB of memory.

Rhea is a (196)-node commodity-type Linux cluster. The primary purpose of Rhea is to provide a conduit for large-scale scientific discovery via pre- and post-processing of simulation data generated on Titan. Users with accounts on INCITE- or ALCC-supported projects are automatically given an account on Rhea. Director's Discretion (DD) projects may request access to Rhea.

Each of Rhea's nodes contains two 8-core 2.0 GHz Intel Xeon processors with Hyper-Threading and 128GB of main memory. Nine GPU nodes complement the processing power each adding 1TB of main memory and two NVIDIA K80 GPU's, similar to what is in production on Titan. Rhea is connected to the OLCF's 32PB high performance Lustre filesystem "Atlas".

CADES (Compute and Data Environment for Science) facility has been developed as a capability that builds upon ORNL's key strengths in data system infrastructure and delivery of new capabilities through data intensive science to meet the mission needs of R&D projects at ORNL and beyond to address big data analytics and science needs. The technical objective of the CADES facility is to provide a data intensive infrastructure that support the mission needs of key projects at ORNL and external users. The hardware infrastructure will comprise a multi-petabyte data storage environment coupled with a multi-teraop data intensive HPC compute environment and a multi-node cloud compute infrastructure. This system will include the necessary software to apply the system to important data intensive problems at ORNL.

1.2 The Joint Institute for Computational Sciences

The University of Tennessee Knoxville (UTK) and Oak Ridge National Laboratory (ORNL) established the Joint Institute for Computational Sciences (JICS) in 1991 to encourage and facilitate the use of high-performance computing in the state of Tennessee and later expanding to encompass a world-class center for research, education, and training in computational science and engineering. JICS contains a national supercomputing user facility, the National Institute for Computational Sciences (NICS) that has delivered more than 4 billion core hours to the National Science Foundations' (NSF's) open science and engineering community as a founding partner in NSF's national cyberinfrastructure, the eXtreme Science and Engineering Discovery Environment (XSEDE). NICS has supported over 6,500 users in 54 fields of science and engineering. The center's high-bandwidth connectivity up to 100 Gbps ensures seamless access and use of data from multiple sources including UTK and remote sites via Internet2.

JICS advances scientific discovery and state-of-the-art engineering by enhancing knowledge of computational modeling and simulation through educating a new generation of scientists and engineers that are well versed in the application of computational modeling and simulation to solve the world's most challenging scientific and engineering problems. To fulfil these roles in the advancement of computational science and education JICS has a professional and experienced operational and engineering staff comprising group in HPC operations, technology integration, user services, and scientific computing. JICS also employs professional research staff, joint faculty, postdoctoral fellows and students, and administrative staff. UTK/ORNL joint faculty hold dual appointments as faculty members in departments at UTK and as staff members in ORNL research groups. The following items describe the relevant resources JICS and NICS currently operate or have available.

The JICS facility, Figure 1, represents a large investment by the state of Tennessee and features a state-of-the-art interactive distance learning center with seating for 66 people, conference rooms, informal and open meeting space, executive offices for distinguished scientists and directors, and incubator suites for students and visiting staff.

Figure 1 Joint Institute for Computational Sciences building

The JICS facility is a hub of computational and engineering interactions. Joint faculty, postdocs, students, and research staff shares the building, which is designed specifically to provide intellectual and practical stimulation. The auditorium serves as the venue for invited lectures and seminars by representatives from academia, industry, and other laboratories The open lobby doubles as casual meeting space and the site for informal presentations and poster sessions, including an annual student poster session with over 200 presenters.

JICS employs professional research staff, joint faculty, postdoctoral fellows and students, and administrative staff. The joint faculty holds dual appointments as faculty members in departments at UT and as staff members in ORNL research groups.

One of JICS' main projects is the National Institute for Computational Sciences (NICS), originally founded in 2007. The mission of NICS is to enable the scientific discoveries of researchers nationwide by providing leading-edge computational resources and education, outreach, and training. NICS has a professional, experienced operational and engineering staff comprising groups in HPC operations, technology integration, user services, scientific computing, and application performance tools.

1.2.1. JICS Resources

JICS has various resources that are the result of National Science Foundation (NSF) awards, UTK investments, and/or strategic partnerships. These resources are available, based on funding and award status, to researchers at the University of Tennessee, Oak Ridge National Laboratory, JICS partners, and the national science and engineering community through director's discretionary allocations, open calls for proposals, and through the NSF funded Extreme Science and Engineering Discovery Environment (XSEDE) program. Resource allocations are made to researchers based on the discretion of resource or institute directors, internal review, or peer-review. The following sections describe the resources that JICS currently operates.

Darter

Darter is a Cray XC30 system funded from UT investments and NSF award ACI:0711134. This system has 1,448 compute sockets each with an Intel Sandy Bridge processor (8 cores/socket.) In total, the machine provides 240.9 TFlops of compute, 11,584 compute cores, and 23.2 TB of memory. This compute resource is a highly integrated computational platform with a high performance interconnect and Dragonfly network topology. This machine was purchased to provide computational cycles to UTK and other academic institutions. The machine is located in the OLCF computer facility.

Beacon

Beacon is an existing JICS cluster, funded from UT investments and NSF awards OCI:1137907 and ACI:0711134. Beacon is an energy efficient Cray cluster that utilizes Intel Xeon Phi coprocessors. Beacon consists of 48 compute nodes and 6 I/O nodes connected with FDR InfiniBand. Beacon has 768 conventional cores and 11,520 accelerator cores that provide more than 210 TFLOP/s of combined computational performance, 12 TB of system memory, 1.5 TB of coprocessor memory, and over 73 TB of SSD storage. The configuration of Beacon's compute nodes makes it an ideal and versatile platform with 256 GB of RAM, 1 TB of SSD storage, and 16 conventional processor cores per node. Using 36 of these compute nodes, a Green500 run was performed that demonstrated a new world record for power efficiency delivering just under 2.5 GFlops per Watt in November of 2012.

High-Performance Storage

JICS currently supports two Lustre file systems located in the OLCF computer facility. A direct attached scratch file system (connected to Darter) is available and comprised of two couplets of Cray Sonexion 1600 storage controllers and back end disk, accessed through an FDR InfiniBand storage area network (SAN). The Sonexion scratch file system provides approximately 350 TB of short-term, high-performance storage to users, with a peak I/O rate of 11 GB/s. The Medusa file system is a multi-cluster file system implemented as a site-wide file system at NICS. Medusa is currently running as three couplets of DDN 10K controllers and their back-end disks, and is accessed through a QDR InfiniBand storage area network (SAN). The Medusa file system provides approximately a 1.3 PB of capacity with a peak I/O rate of 30 GB/s.

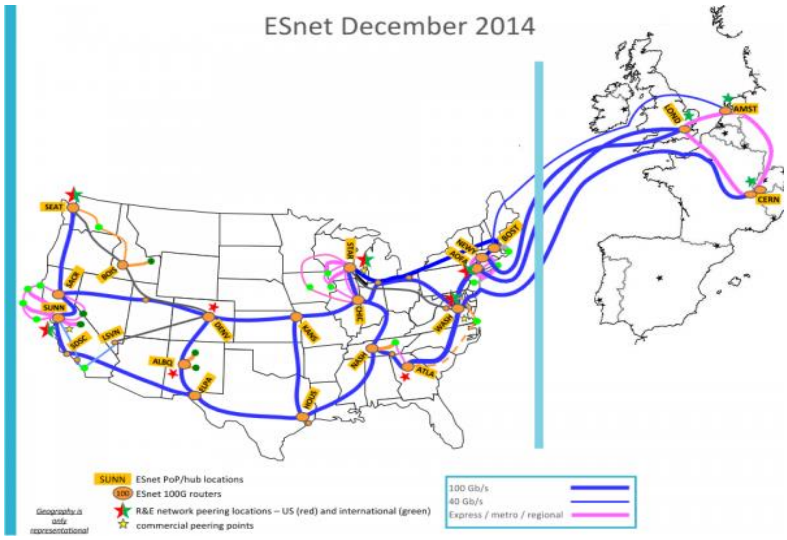
1.3 Computer Facilities

In June 2004, JICS moved into the then brand new 52,000 ft^2 building next door to the ORNL OLCF computer facility. The JICS building has a 1,500 ft^2 computer room, which is home to Beacon. The OLCF computer facility, located on the ORNL campus, is among the nation's most modern facilities for scientific computing and currently is home to Darter and the JICS high-performance file systems. The OLCF facility has 40,000 ft^2 divided equally into two rooms with 9 x 2.5MVA transformers, and another 27,000 ft^2 divided over two rooms in a recently added expansion building with 1 x 2.5 MVA transformer and the ability to expand; and finally 6,600 tons of chilled water – all of which is designed specifically for high-end computing systems. Finally, the UTK campus has a 2,116 ft^2 computer room. ORNL and JICS utilize staff who provide continuous monitoring and immediate problem resolution in all of the computing facilities except UTK, where the UTK Office of Information Technology is responsible. On-site Operators and Tier-1 System Administrators provide front-line problem resolution for users 24x7x265. Additional user support and system administration services are available for issues that require further escalation.

2. Infrastructure

Physical and Cyber Security. ORNL has a comprehensive physical security strategy including fenced perimeters, patrolled facilities, and authorization checks for physical access. An integrated cyber security plan encompasses all aspects of computing. Cyber security plans are risk-based. Separate systems of differing security requirements allow the appropriate level of protection for each system, while not hindering the science needs of the projects.

Network Connectivity. ORNL has access to all major research and education (R&E) networks around the world. Connectivity to these networks is provided via the Department of Energy (DOE) Energy Sciences Network (ESnet), the Southern Crossroads (SoX) southeastern regional R&E exchange, and the ORNL optical network operated by UT-Battelle. ORNL connects to diverse ESnet hubs at 100Gbps and SoX at 10Gbps. The ORNL optical network uses leased fiber-optic cable and connects the OLCF to major networking hubs in Atlanta, Nashville, and Chattanooga.



ORNL network connectivity to university, national laboratory, and industry partners.

The EVEREST laboratory has been upgraded with dual power walls and 3-D capability.

Visualization and Collaboration. ORNL has state-of-the-art visualization facilities that can be used on site or accessed remotely.

ORNL's **E**xploratory **V**isualization **E**nvironment for **RE**search in **S**cience and **T**echnology (EVEREST) facility is a scientific laboratory deployed and managed by the Oak Ridge Leadership Computing Facility (OLCF). The primary mission of this laboratory is to provide tools to be leveraged by scientists for analysis and visualization of simulation data generated on the OLCF supercomputers.

Three computing systems are currently provided in the laboratory. These consist of a distributed memory Linux cluster, a shared memory Linux node, and a shared memory Windows node. Access to the Linux computing resources requires an EVEREST account and an RSA Secure ID. Access to the Windows computing resources requires a standard ORNL UCAMS account and does not require a specific EVEREST account.

Two tiled display walls are provided. The primary display wall spans 30.5' x 8.5' and consists of eighteen 1920×1080 stereoscopic Barco projection displays arranged in a 6 x 3 configuration. The secondary display wall consists of sixteen 1920×1080 Planar displays arrange in a 4 x 4 configuration providing a standard 16:9 aspect ratio.

There are four additional peripheral video inputs located on pop-out boxes in the conference table. Each input supports both digital DVI and analog VGA. Users of the laboratory are welcome to control either wall using personal hardware that is brought into the laboratory. Power outlets are provided at the conference table.

The laboratory instruments are controlled using a touch panel interface located at the control desk. All computing resources can be routed to any available display wall. User hardware using the video input ports on the conference table can also be routed via the touch panel.

OLCF tape archive.

High Performance Storage and Archival Systems. To meet the needs of ORNL's diverse computational platforms, a shared parallel file system capable of meeting the performance and scalability requirements of these platforms has been successfully deployed. This shared file system, based on Lustre, Data Direct Networks (DDN), and InfiniBand technologies, is known as Spider and provides centralized access to petascale datasets from all major on-site computational platforms. Delivering more than 1 TB/s of aggregate performance, scalability to more than 20,000 file system clients, and 30-petabyte (PB) storage ca-pacity, Spider is one of the world's largest scale Lustre file system. Spider consists of 36 DDN SFA12KX storage arrays managing 20,160 2-TB Nearline-SAS drives served by 288 Dell dual-socket, quad-core I/O servers Metadata are stored on a NetApp EF560 storage array and are served by eight Dell single-socket, eight-core systems with an aggregate of over 2 Terabytes of memory. ORNL systems are interconnected to Spider II via an InfiniBand system area network, which consists of 3 Mellanox SX6506 Director Class IB switches, and 36 Mellanox SX6036 IB switches; with more than 3 miles of optical cables. Archival data are stored on the center's High Performance Storage System (HPSS), developed and operated by ORNL. HPSS is capable of archiving hundreds of petabytes of data and can be accessed by all major leadership computing platforms. Incoming data are written to disk and later migrated to tape for long term archiving. This hierarchical infrastructure provides high-performance data transfers while leveraging cost effective tape technologies. Robotic tape libraries provide tape storage. The center has six SL8500 tape libraries each holding up to 10,000 cartridges. The libraries house 72 T-10K-D tape drives (8 terabyte cartridges, uncompressed). Several generations of media exist, but all new data is being written to T10K-D. Each drive delivers a bandwidth in excess of 200 MB/s. ORNL's HPSS disk storage is provided by DDN and NetApp storage arrays with nearly 20 petabytes of capacity and over 200 GB/s of bandwidth. This infrastructure has allowed the archival system to scale to meet increasingly demanding capacity and bandwidth requirements with more than 53 PB of data stored as of January 2016.

Appendix 5: Equipment

Relevant equipment is included in the discussion of Facilities and Other Resources (Appendix 4).

Appendix 6: Project Data Management Plan

We recognize the value of validation of results as part of the scientific process and the associated need for the availability of data behind our work. The PI will take responsibility for the collection, management, and sharing of the research data. The project members will conform to best practices and standards to facilitate the reproduction and verification of our results.

Data Types and Sources

This project will produce two general classes of data: (1) new and expanded software and (2) experimental results as a function of problem specification, input parameters, hardware specifications, software optimizations, and so on. Validation of results from our work will rely on two components: (1) availability of source code that implements our designs, along with associated documentation; and (2) data resulting from experiments, particularly data forming the basis of charts and figures in published work. Access to relevant compute resources by third parties in order to repeat experiments is outside the context of this document. Software will be accompanied by technical documentation. Any experimental results this project publishes will be documented with the hardware and software information, to facilitate reproduction/validation where possible.

All software produced in this work will comply with the terms of the ASCR policy on open source software. Augmentations to existing software will be released under the existing license and integrated into the existing repository. New software will be made available under a compliant license after appropriate laboratory software processes have been followed. Care will be taken to select a compliant license that is amenable to expected uses (e.g., integration into the Linux kernel, or reuse by a vendor). Documentation will be provided to facilitate build and test execution, as appropriate for a research activity.

Data resulting from experiments will be provided in the form used to generate charts and figures in published work (e.g., R or GNUplot scripts, Excel spreadsheets, raw log files). A link to this data will be provided in publications to facilitate access by interested parties, and data will be made available by the time of publication. The data will be annotated as appropriate to facilitate interpretation.

Availability

A web site will be created as a single location for finding information on this project. Data will be hosted at that site. Our expectation is that the nature of the activities proposed here will lead to only a modest amount of data to be made available in this way; thus, our expectation is that our needs will be satisfied by the standard capabilities provided by our facility. All data produced by this project will be retained for the duration of the project and a minimum of 12 months following the project's completion. The web site will provide guidance on how to reference software and data as appropriate.

Because our work is in part derived from existing and ongoing software activities, it does not make sense to duplicate software into a single repository. Rather, existing software repositories will be referenced from this web site as appropriate, and new web sites similarly referenced for software specific to this project. For such repositories, participating institutions will use best practices for service providers, including regular backups, disaster recovery plans, security plans, and general availability guarantees. Documentation of how to build any composites will be provided from the web site to facilitate building our software artifacts.

Data Protection

Under no circumstances will proprietary, business confidential, or personally identifiable information be exposed as part of this project. Laboratory processes will be followed for code, document, and data release to ensure that this is the case. Research will be conducted on open systems and applications for which public information is already available. We will ensure any data from outside sources has gone through the respective release processes of those sources before being made available by the project team.