



Portland,
OR | 2009

Conference Program

Oregon Convention Center

Portland, Oregon

November 14-20, 2009

<http://sc09.supercomputing.org>

*The premier international conference on
high performance computing, networking,
storage and analysis*

Computing for a Changing World

Sponsors of Tomorrow.™ 

We take speed beyond fast.

Visit Intel booth #1935

www.intel.com/technology/business/hpc



Where data becomes knowledge.

The latest technology from Intel helps you accelerate your application today and scale for tomorrow. Address real-time, real-world challenges with intelligent performance from Intel.



Computing for a Changing World

Contents

| | |
|--|--|
| 1 Governor and Mayor's Welcome | 49 Posters |
| 7 Chair's Welcome | 69 Masterworks |
| 8 Thrust Areas/Conference Overview | 70 Future Energy Enabled by HPC |
| 11 Registration | 71 Data Challenges in Genome Analysis |
| 11 Registration/Conference Store | 71 Finite Elements and Your Body |
| 11 Registration Levels | 72 HPC in Modern Medicine |
| 11 Exhibit Hours | 73 Multi-Scale Simulations in Bioscience |
| 11 Media Room Registration | 74 Toward Exascale Climate Modeling |
| 12 Social Events | 75 High Performance at Massive Scale |
| 12 Registration Pass Access | 76 Scalable Algorithms and Applications |
| 13 Daily Schedules/Maps | 77 Panels |
| 27 Invited Speakers | 78 Building 3D Internet Experiences |
| 28 Opening Address | 78 3D Internet Panel |
| 28 Plenary and Kennedy Award Speakers | 78 Cyberinfrastructure in Healthcare Management |
| 29 Cray/Fernbach Award Speakers | 79 Energy Efficient Datacenters for HPC: How Lean and Green do we need to be? |
| 29 Keynote Address | 79 Applications Architecture Power Puzzle |
| 31 Papers | 79 Flash Technology in HPC: Let the Revolution Begin |
| 32 GPU/SIMD Processing | 80 Preparing the World for Ubiquitous Parallelism |
| 32 Large-Scale Applications | 80 The Road to Exascale: Hardware and Software Challenges |
| 33 Autotuning and Compilers | 81 Awards |
| 34 Cache Techniques | 82 ACM Gordon Bell Finalists I |
| 35 Sparse Matrix Computation | 83 ACM Gordon Bell Finalists II |
| 36 Particle Methods | 83 ACM Gordon Bell Finalists III |
| 37 Performance Tools | 84 Doctoral Research Showcase I |
| 37 Virtual Wide-Area Networking | 86 ACM Student Research Competition |
| 38 Acceleration | 87 Doctoral Research Showcase II |
| 38 Grid Scheduling | 89 Challenges |
| 39 High Performance Filesystems and I/O | 90 Storage Challenge Finalists |
| 40 GPU Applications | 93 Exhibitor Forum |
| 41 Networking | 107 Birds-of-a-Feather |
| 42 Performance Analysis and Optimization | 123 Tutorials |
| 43 Sustainability and Reliability | 131 Workshops |
| 43 Metadata Management and Storage Cache Allocation | 137 SCinet |
| 44 Multicore Task Scheduling | 141 Conference Services |
| 45 System Performance Evaluation | 147 Acknowledgements |
| 46 Dynamic Task Scheduling | |
| 47 Future HPC Architectures | |

STATE OF OREGON
PROCLAMATION
OFFICE OF THE GOVERNOR

WHEREAS: We live in a dramatically changing world and the application of advanced computing technologies is key to the improvement, sustainability, and vitality of all Oregonians and society as a whole; and

WHEREAS: The attendees of SC09 at the Oregon Convention Center are the heartbeat of a technological ecosystem that leads the way in creating and adapting 'Computing for a Changing World' – the theme of SC09; and

WHEREAS: The city of Portland has been selected to host the 22nd annual gathering of the world's leading computational scientists, researchers and computing experts, for the third time.

NOW,
THEREFORE: I, Theodore R. Kulongoski, Governor of the State of Oregon, hereby proclaim **November 14 – 20, 2009** to be

SUPERCOMPUTING WEEK

in Oregon and encourage all Oregonians to join in this observance.

IN WITNESS WHEREOF, I hereunto set my hand and cause the Great Seal of the State of Oregon to be affixed. Done at the Capitol in the City of Salem in the State of Oregon on this day, September 22, 2009.



Theodore R. Kulongoski
Theodore R. Kulongoski, Governor

Kate Brown
Kate Brown, Secretary of State

CITY OF
PORTLAND

Whereas

Whereas, Supercomputing returns to the City of Portland for its SC09 conference; and

Whereas, SC09 is recognized globally as the premier international conference on High Performance Computing (HPC), networking, storage and analysis; and

Whereas, SC09 will feature the most interesting and innovative HPC scientific and technical applications from around the world; and

Whereas, the SC09 conference is also respected as the premier global networking event for stakeholders throughout the HPC ecosystem; and

Whereas, no other conference brings together this many scientists, engineers, researchers, educators, programmers, system administrators, application developers, and program managers; and

Whereas, SC09 will bring to the City of Portland, 11, 000 attendees from around the world, representing an estimated economic impact of more than \$10 million;

Now, therefore, I, Sam Adams, Mayor of the City of Portland, Oregon, the "City of Roses," do hereby proclaim November 14-20, 2009 to be

Supercomputing Week

in Portland, and encourage all residents to recognize not only the accomplishments and contributions of those attending SC09, but also the economic importance of having this distinguished conference return to our city.



THEODORE R. KULONGOSKI
GOVERNOR



November 14 - 20, 2009

Greetings:

As Governor of Oregon, I am pleased to welcome you to the SC09, International Conference for High Performance Computing (HPC), Networking, Storage and Analysis. I also invite you to enjoy our beautiful scenery, warm hospitality and the special attractions Oregon and the exciting city of Portland have to offer.

With professionals from around the globe, this conference is an opportunity to meet new people and experience the progress and innovations of HPC scientific and technical applications. This event gives participants an opportunity to explore the ever-changing frontier of computer technology and to further delve into bio-computing, sustainability and 3-D Internet. The conference joins colleagues, peers and industry experts together to focus today's efforts on tomorrow's advancements as experts address "Computing for a Changing World."

On behalf of the citizens of Oregon, best wishes for a successful event, and may your travel home be safe.

Sincerely



THEODORE R. KULONGOSKI
Governor

STATE CAPITOL, SALEM 97301-4047 (503) 378-3111 FAX (503) 378-4863 TTY (503) 378-4895
WWW.GOVERNOR.STATE.OR.US



OFFICE OF MAYOR SAM ADAMS
CITY OF PORTLAND

Greetings:

It is my sincere pleasure to welcome you to the City of Portland for SC 09, the 22nd annual International Conference for High Performance Computing, Networking, Storage and Analysis.

As I have learned more about this vitally important conference, I have been truly amazed at the dedicated following that has continued to grow over the past 21 years. Portland is honored to host this conference composed of such a diverse, global community of participants including researchers, scientists, computing center staff members, IT and data center management, application developers, computer manufacturing personnel, program managers, journalists and congressional staffers.

Truly, diversity is one of this conference's main strengths, making it a yearly "must attend" forum for stakeholders throughout the scientific and technical computing community.

SC09 is a gathering like no other – where technology and brain trust come together to further the advancement of the sciences and enable the technology that will make significant future contributions to society.

We applaud your theme of, "Computing for a Changing World" and your special focus areas this year which include Bio-Computing, 3D Internet, and a topic that Portland is very familiar with – sustainability.

Our city is proud to have been chosen as the host location for SC09.

For the thousands of you visiting from out of town, we hope you enjoy the Rose City's numerous attractions. Whether you visit our Waterfront Park or enjoy shopping downtown, dine at one of hundreds of restaurants in our "Pearl District," or just take in the views of Mt. Hood from Portland's Rose Garden, we are sure you will find your visit an unforgettable and enjoyable experience.

Again, on the behalf of the beautiful City of Portland, Oregon – welcome, thank you for being here, and best wishes for a successful conference. We hope you will return again soon.

Sincerely,

Mayor Sam Adams

1221 SW FOURTH AVENUE, SUITE 340 ♦ PORTLAND, OREGON 97204
(503) 823-4120 ♦ mayorsamadams.com

Computing for a Changing World



Portland,
OR | 2009

Welcome



Wilfred Pinfold
General Chair

Welcome to SC09, the 22nd Annual conference on high performance computing, networking, storage and analysis. Sponsored by Association of Computer Machinery (ACM) and the IEEE Computer Society, the conference provides the premier forum for the exchange of information in the global HPC community through a highly respected technical program, a vibrant exhibit floor and a set of strong community programs.

This year's technical program offers a broad spectrum of presentations and discussions. Rigorously reviewed tutorials, papers, panels, workshops and posters provide access to new research results from laboratories and institutions around the world. As an example this year's committee selected 59 technical papers from 261 submissions in six areas: applications, architecture, networks, grids/clouds, performance, storage and systems software. The quality of this year's technical program is exceptionally high.

For 2009, we have added three Technology Thrust areas: 3D Internet, Bio-Computing, and Sustainability. These highlight ways in which HPC is changing our world—and we recognize this in the conference theme, “Computing for a Changing World.” Opening speakers on Tuesday, Wednesday and Thursday will introduce each of the thrust areas, and related content will be highlighted in the technical program and on the exhibit floor.

This year's exhibits make full use of the 255,000 square feet of trade show floor in the recently remodeled and expanded Oregon Convention Center. As of late summer we have 275 organizations participating, with 40 first-time exhibitors.

SC09 welcomes a tremendous diversity of attendees from all professions related to high performance computing, including corporate, research and educational institutions. Four special programs have been developed to improve the SC conference experience. This year we inaugurated the SC Ambassadors program to focus on the needs of our growing community from outside the USA. The SC Broader Engagement Program facilitates participation and engagement of individuals who have been traditionally underrepresented in the field of HPC. The SC Education Program supports undergraduate faculty, administrators and college students, as well as collaborating high school teachers who are interested in bringing HPC technologies to the next generation of scientists, technologists, engineers, mathematicians and teachers. The SC Student Volunteer Program supports undergraduate and graduate students who work as volunteers at the conference, where they have the opportunity to discuss the latest technologies and meet leading researchers from around the world.

Portland is an exceptionally welcoming city. As you will see as the local people and businesses extend their warmth and hospitality to you during the week. Please enjoy the many parks, gardens and public walkways. The city's “Pearl District” is noteworthy as home to many exceptional restaurants, galleries and retail stores. While in Portland, you should take some time to enjoy the local microbrews, Oregon wines, Northwest coffee shops and fantastic ethnic restaurants. And finally, if possible take the time to travel into the extraordinary beautiful surrounding areas, both west toward the coast and east toward the Cascade Mountains. Oregon's scenery is truly something to experience.

The conference week is the culmination of three years of planning by an outstanding committee of your peers. In addition to this committee the conference is supported by a small army of student volunteers. Please join me in thanking them all for making this event possible.

We hope you enjoy every aspect of SC09 and the great city of Portland, Oregon.

Wilfred Pinfold
General Chair, SC09

Thrust Areas

Bio-Computing

Often referred to as the “Decade of Biology,” high performance computing is leveraging the onslaught of biological information to meet societal needs. Advanced modeling and simulation, data management and analysis, and intelligent connectivity are enabling advances in such fields as medicine, environment, defense, energy and the food industry.

Biological modeling is progressing rapidly due not only to data availability, but also to the increases in computer power and availability. This is leading to the growth of systems biology as a quantitative field, in which computer simulation and analysis addresses the interoperation of biological processes and functions at multiple spatial, temporal, and structural scales. Detailed, reductionist models of single processes and cells are being integrated to develop an understanding of whole systems or organisms. SC09 will highlight the state of the art in biological simulation, introducing the unique complexities of living organisms.

The past decade has seen a rapid change in how we practice biology. The science is evolving from single-lab data generation and discovery to global data access and consortium-driven discovery. High-throughput instruments and assays deliver increasingly large volumes of data at throughputs that make biology a data intensive science, fundamentally challenging our traditional approaches to storing, managing, sharing and analyzing data while maintaining a meaningful biological context. We will highlight experts in the field of genomics that will address the core needs and challenges of genome research and discuss how we can leverage new paradigms and trends in distributed computing.

The medical industry is changing rapidly by integrating the latest in computing technologies to guide personnel in making timely and informed decisions on medical diagnoses and procedures. By providing a systems-level perspective, integration, interoperability, and secured access to biomedical data on a national scale we are positioned to transform the nation's healthcare. Doctors can utilize simulation-based methods, initialized with patient-specific data, to design improved treatments for individuals based on optimizing predicted outcomes. We will present issues and projects across clinical care, public health, education and research with particular focus on transformative efforts enabled by high performance computing.

HPC hardware must be available and appropriate in architecture to support heterogeneously-structured modeling. HPC systems

must support interdisciplinary collaborations with software that facilitates construction of simulations that are efficient, flexible, and robustly address inquiry in experimental science. SC09 sessions will discuss how architecture is impacting the broad field of biology.



Watch for this icon identifying activities related to bio-computing.

Sustainability

HPC is playing a lead role in modeling, simulations and visualization to support primary research, engineering programs and business tools that are helping to create a more “sustainable” future for humankind. Simultaneously, those involved in High Performance Computing are also scrutinizing how to minimize their immediate impact on energy consumption, water consumption, carbon-emissions and other non-sustainable activities.

The “sustainability” thrust area at SC09 illuminates how HPC is impacting societies’ drive toward understanding and minimizing the impacts of our activities and in the search for sustainable solutions that address societal needs and goals. As our computer centers grow to tens of megawatts in power, we are, of course, concerned with the impacts of our own activities. Even with respect to this conference, you will see how the convention center and the conference committee have teamed to put “reduce, reuse, recycle” into play.

We have assembled numerous technical events on the topic of sustainability. Those of you interested in the broad topic of future technologies for peta- and exascale systems will find information on component technologies of the future designed with energy-sipping technologies to enable broad deployment in large systems. Data center enthusiasts will find numerous sessions that deal with data center design, retrofit, and operational efficiency. For those interested in topics on the use of energy and impact on the environment, we have assembled a number of world experts in climate modeling, green energy and the future of fossil fuels for our generation.

Many of these topics are being strongly debated in the press these days so we anticipate good interaction in the sessions. We are pleased to have been able to bring you such high-quality information and hope you are able to learn and make good connections to satisfy your interests.



Watch for this icon identifying activities related to sustainability.

3D Internet

3D Internet, enabled by supercomputing technology, will change the way we collaborate, visualize, learn and play. This thrust area explores the use of HPC and its impact and application in delivering high quality, 3D visualization and analysis solutions. The technologists of both HPC and 3D have a great deal to learn from each other. Through submitted papers, in-depth workshops, demonstrations and panel discussions, we hope to stimulate a wide range of discussion on this topic.

Two panels, “3D across HPC,” will explore the business and technology implications in emerging 3D internet applications that are utilizing HPC hardware and software solutions. The technology discussion will focus on the technical impact of HPC in improving scalability, parallelism and real time performance of emerging gaming, 3D content, education and scientific visualization experiences delivered via the internet. The business discussion will focus on emerging business models to bring mass adoption of 3D internet technology.

A half day workshop titled “Building 3D Internet Experiences” will have technology practitioners walk through the development process for compelling scientific simulations and visualizations and collaborative educational environments.

SC09 is leading the development of Sciencesim, an open source platform for the development of 3D Simulation and collaboration experiences and welcomes submissions of applications of this technology for community exploration and review.



Watch for this icon identifying activities related to 3D Internet

Disruptive Technologies

For last half century, we’ve witnessed exponential growth in the scale and capability of semiconductor electronics and related technologies, a phenomenon commonly referred to as Moore’s Law. This institutionalized revolution in technology has enabled extraordinary changes in almost all facets of modern society. Steady and predictable improvements in Silicon VLSI have largely squeezed out other, potentially disruptive competing technologies. The ITRS roadmap suggests that this will continue to be true for another decade.

This is not to suggest that another decade of Moore’s Law will not bring disruptive change. Power density has constrained processor clock rates, and increased performance now comes from an exponential growth in the number of cores. The density

of logic is increasing faster than that of DRAM, and the familiar ratio of one byte per flop/s is increasingly unaffordable. Technologies such as NAND-Flash are appearing in systems, and we may soon enjoy persistent main memory.

Therefore, the first objective of this year’s talks will be present a vision of how another decade of Moore’s Law will change the technology underlying today’s systems. Wilfried Haensch from IBM Research will present how VLSI will likely evolve out to the end of the next decade, and the impact that will have on the logic that underpins computing and networking systems. Dean Klien from Micron will describe how DRAM will evolve, and what new technologies might soon compliment it. Finally, Keren Bergman from Columbia University will discuss how similar rapid changes in optics will impact the networks which tie our systems together.

The second objective of Disruptive Technologies is to understand how another decade of Moore’s Law will impact system software and applications. Thomas Sterling from Louisiana State University will describe how the exponential growth in the number of cores could lead to a new model of execution. Vivek Sarkar from Rice will discuss how systems software, such as the operating system, will likely evolve. Finally, Tim Mattson from Intel will discuss changes in how the developers of applications, ranging games to large-scale scientific codes, are already adapting to the challenge of programming extreme parallel systems.

Don’t miss the Disruptive Technology exhibit this year (lobby area, Exhibit Hall D-E), which will feature innovative new technology, often a preview of what will be presented in larger SC09 exhibits. The technology presented will address a number of concerns ranging from maximizing the PUE of servers to accelerating MATLAB with GPUs.

Datacenter of the Future

This booth showcases design elements of energy efficient HPC datacenters from diverse locations around the globe. These are innovative and current design features, many of which are from datacenters that are not yet fully built and on-line. The line up includes datacenters from government labs, as well as universities from the United States, Asia and Europe. Also included in the booth are the submission results from the Datacenter of the Future Challenge. Representatives from the showcase datacenters, the Green500, the Energy Efficient HPC Working Group and authors Lauri Minas and Brad Ellison (Energy Efficient Datacenter) will be available in the booth to answer questions and engage in collegial discussions.

The booth is located in the lobby area, Exhibit Hall D-E.

SC09 Communities

SC-Your Way

SC can be an overwhelming experience for attendees, particularly first-timers. With each passing year, the conference gets bigger - more exhibits, more technical program sessions and more attendees—making it harder than ever to keep track of things. It's too easy to miss out on events important to you!

Add to that a new city and finding your way around the area, like Portland this year. To help SC09 attendees make sense of it all, we're repeating the web-based service called SC-Your Way that was launched last year.

This new system integrates two portals:

- SC-Your Way helps you navigate the conference itself by identifying every session related to a particular topic or assisting you to plan a personalized route through the exhibits
- Portland-Your Way helps you make the most of your visit to Portland, assisting you to choose the conference hotel that's right for you and informing you about the local music scene, restaurants, and recreational activities

Information is available on the web and at a booth located near the registration area.

Education Program

The Education Program is a year-round effort designed to introduce HPC and computational tools, resources and methods to undergraduate faculty and pre-college educators. The program also assists the educators in integrating HPC and computational techniques into their classrooms and research programs.

The Education Program offers ten week-long summer workshops on campuses across the country covering a range of topics: parallel/distributed/grid computing, computational chemistry, computational biology, computational thinking and curriculum development. Educators are encouraged to attend these in-depth hands-on workshops as teams to foster sustained incorporation of computation into their curriculum and institution.

During the annual SC conference in November, the Education Program hosts a four-day intensive program that further immerses participants in high-performance computing, networking, storage and analysis. The program offers guided tours of the exhibit hall, focused hands-on tutorials and birds-of-a-feather gatherings, as well as formal and informal opportunities to interact with other SC communities and exhibitors.

Our goal is to encourage SC Education Program participants to take the knowledge they have garnered from these workshops and the conference back to their campuses and energize their students and colleagues to advance the preparation of today's students to be tomorrow's leaders in the HPC community.

Broader Engagement

The aim of the SC09 Broader Engagement program is to increase the involvement of individuals who have been traditionally underrepresented in the HPC field. The program offers a number of special activities at SC09 to stimulate interest within the SC conferences and the HPC community in general.

A number of specialized sessions have been designed specifically for students and early-career faculty and professionals. These sessions will include discussions on topics such as mentoring, cybersecurity and computational biology.

Mentor/Protégé Program

Most people participating in the Broader Engagement, Student Volunteers and Student Cluster Challenge programs are experiencing SC for the first time. It can be an exciting but sometimes overwhelming experience. To help them make the most of their experience, we have developed a Mentor/Protégé Program associated with Broader Engagement. This program matches each participant (protégé) with a mentor who has attended SC before and who is willing to share their experiences. Matches are made on the basis of similar technical backgrounds and interests.

SC Ambassadors

Special Resources for International Attendees

SC09 has launched a new program intended specifically to assist attendees from outside the USA. The SC conference series has always had a large number of international attendees, both in the technical program and the exhibit hall. This year, a special effort is being made to provide special assistance and facilities to enhance the SC09 experience of international participants. A new group, known as SC Ambassadors, has been set up to implement this new initiative.

At the conference itself, an International Attendees Center will provide a place for international attendees to meet, learn more about the conference and interact with international colleagues, as well as organizers of the conference.

We extend a special welcome to international attendees and hope your conference experience is academically and socially rewarding.

Registration

Registration/Conference Store

The registration area and conference store are located in the convention center lobby.

Registration/Conference Store Hours

| | |
|--------------------------|-------------------|
| Saturday, Nov. 14 | 1pm-6pm |
| Sunday, Nov. 15 | 7am- 6pm |
| Monday, Nov 16 | 7am-9pm |
| Tuesday, Nov 17 | 7:30am-6pm |
| Wednesday, Nov 18 | 7:30am-6pm |
| Thursday, Nov 19 | 7:30am-5pm |
| Friday, Nov 20 | 8am-11am |

Registration Levels

Technical Program

Technical Program registration provides access to plenary talks, posters, panels, BOFs, papers, exhibits, challenges, awards, Masterworks, the Doctoral Showcase, and workshops.

Exhibitor

Exhibitor registration provides access to the exhibit floor and to limited Technical Program events for attendees affiliated with organizations with exhibits on the show floor.

Exhibits Only

Exhibits Only registration provides access to the exhibit floor for all three days of the exhibition during regular exhibit hours. It does not provide access to the Monday Night Gala Opening.

Member, Retired Member and Student Registration Discounts

To qualify for discounted registration rates, present your current IEEE, IEEE Computer Society, ACM, or ACM SIGARCH membership number or a copy of a valid full-time student identification card when registering. You may complete the IEEE Computer Society and/or ACM membership application provided in your conference bag and return it to the Special Assistance desk in the registration area to receive the member discounted registration rate.

Tutorials

Full-day and half-day tutorials are offered on Sunday and Monday, November 16 and 17. Tutorials are not included in the Technical Program registration fee and require separate registration and fees. Attendees may choose a one-day or two-day passport, allowing them to move freely between tutorials on the selected day(s). Tutorial notes and luncheons are provided for each registered tutorial attendee.

Registered tutorial attendees will receive a copy of all tutorial notes on a computer-readable medium; no hardcopy notes will be distributed or available. Some of the tutorials will have hands-on components. For these, attendees must bring their own laptops with SSH software installed. Rooms used for hands-on tutorials will be equipped with wired network drops, Ethernet cables, SCinet wireless networking, and power drops, but there will be no computer support available. Please arrive early, as there may be tutorial-specific software to install on your laptop.

Children

Children under age 12 are not permitted on the floor except during Family Hours (4-6pm, Wednesday, November 18), and must be accompanied by a family member who is a registered conference attendee.

Proceedings

Attendees registered for the Technical Program will receive one copy of the SC08 proceedings on a USB flash drive.

Lost Badge

There is a \$40 processing fee to replace lost badges.

Exhibit Floor Hours

| | |
|---------------------------|--------------------|
| Tuesday, Nov. 17 | 10am - 6pm |
| Wednesday, Nov. 18 | 10 am - 6pm |
| Thursday, Nov. 19 | 10am - 4pm |

Media Registration

Location: C125

Media representatives and industry analysts should visit the Media Room for onsite registration. A valid media identification card, business card, or letter from a media outlet verifying a freelance assignment writing, researching and filing their stories, or interviewing conference participants and exhibitors. The Media Room is also available to exhibitors who wish to provide materials to, or arrange interviews with, media representatives

and industry analysts covering the conference. A separate room will be available for conducting interviews and may be reserved through the media room coordinator during the conference. Credentialed media representatives are allowed to photograph or video-record the SC exhibits and technical program sessions as long as they are not disruptive. Whenever possible, the communications committee will provide an escort to assist with finding appropriate subjects.

Media Room Hours:

| | |
|---------------------------|----------------|
| Sunday, Nov. 15 | 1pm-4pm |
| Monday, Nov. 16 | 9am-6pm |
| Tuesday, Nov. 17 | 8am-5pm |
| Wednesday, Nov. 18 | 9am-5pm |
| Thursday, Nov. 19 | 8am-4pm |

Social Events

- **Exhibitor Party:** From 7:30pm-10pm Sunday November 15, SC09 will host an exhibitor party for registered exhibitors. Get ready for a night of food and fun at the Portland Union Station. Featured entertainment will be the Midnight Serenaders (www.midnightserenaders.com). There will be other surprises such as Paparazzi Tonight! Your exhibitor badge is all that is required to participate.

- **Gala Opening Reception:** On Monday, November 16, SC09 will host its annual Grand Opening Gala in the Exhibit Hall, 7-9pm. A new feature at this year's Opening Gala is an area where you can taste regional wines. Look in the Exhibit Hall for signs indicating the location. Bring your drink tickets and join us in tasting some wonderful Oregon wines. (Requires Technical Program registration; guest tickets may be purchased in the SC09 Store.)
- **Poster Reception:** From 5:15-7pm Tuesday, take the opportunity to visit the area outside the Oregon Ballroom and discuss late-breaking research results with research poster presenters. The event is open to technical program attendees and registered exhibitors.
- **SC09 Conference Reception:** The Thursday night conference party is a traditional highlight of SC09. This year's party will be held at the Portland Center for the Performing Arts (PCPA), a short light-rail trip from the convention center and within easy walking distance from many of the downtown hotels. PCPA is the premier arts and entertainment venue in the Pacific Northwest and is nationally recognized as one of the top 10 performing arts centers in the nation. From 6-9pm, you will be able to sample the best of Oregon cuisine, wine and beer. There will be local music and entertainment and plenty of comfortable, quiet space to meet with your colleagues. (Requires Technical Program registration; guest tickets may be purchased in the SC09 Store.)

Registration Pass Access

| Type of Event | Tutorials | Technical Program | Exhibits Only | Exhibitor |
|------------------------------|-----------|-------------------|---------------|-----------|
| All Tutorial Sessions | • | | | |
| Tutorial Lunch | • | | | |
| Exhibitor Party | | | | • |
| Monday Exhibits/Gala Opening | | • | | • |
| Tuesday Opening Address | | • | | • |
| Tuesday Poster Reception | | • | | |
| Wednesday Plenary | | • | | |
| Cray/Fernbach/Kennedy Awards | | • | | |
| Thursday Keynote Address | | • | | • |
| Invited Speakers | | | | |
| Thursday Night Reception | | • | | • |
| Birds-of-a-Feather | | • | | • |
| Challenge Presentations | | • | | • |
| Exhibitor Forum | | • | • | • |
| Exhibit Floor | | • | • | • |
| Masterworks | | • | | |
| Panels (Friday Only) | | • | • | • |
| Panels (Except Friday) | | • | | |
| Paper Sessions | | • | | |
| Poster Reception | | • | | |
| Poster Presentations | | • | • | • |
| SCinet Access | • | • | • | • |
| Workshops | | • | | |



Computing for a Changing World

Daily Schedules and Maps

Daily Schedules and Maps

Sunday, November 15

| Time | Event | Title | Location |
|-------------|--------------|---|---------------|
| 8:30am-5pm | Tutorial | S01: Application Supercomputing and the Many-Core Paradigm Shift | * |
| 8:30am-5pm | Tutorial | S02: Parallel Computing 101 | * |
| 8:30am-5pm | Tutorial | S03: A Hands-on Introduction to OpenMP | * |
| 8:30am-5pm | Tutorial | S04: High Performance Computing with CUDA | * |
| 8:30am-5pm | Tutorial | S05: Parallel I/O in Practice | * |
| 8:30am-5pm | Tutorial | S06: Open-Source Stack for Cloud Computing | * |
| 8:30am-Noon | Tutorial | S07: InfiniBand and 10-Gigabit Ethernet for Dummies | * |
| 8:30am-Noon | Tutorial | S08: Principles and Practice of Application Performance Measurement and Analysis on Parallel Systems | * |
| 8:30am-Noon | Tutorial | S09: VisIt - Visualization and Analysis for Very Large Data Sets | * |
| 8:30am-Noon | Tutorial | S10: Power and Thermal Management in Data Centers | * |
| 9am-5:30pm | Workshop | 4th Petascale Data Storage Workshop | A106 |
| 9am-5:30pm | Workshop | 5th International Workshop on High Performance Computing for Nano-science and Technology (HPCNano09) | A107 |
| 9am-5:30pm | Workshop | Workshop on High Performance Computational Finance | A105 |
| 9am-5:30pm | Workshop | Component-Based High Performance Computing (Day 1) | A103 |
| 9am-5:30pm | Workshop | Third International Workshop on High-Performance Reconfigurable Computing Technology and Applications (HPRCTA'09) | A108 |
| 1:30pm-5pm | Tutorial | S11: Emerging Technologies and Their Impact on System Design | * |
| 1:30pm-5pm | Tutorial | S12: Large Scale Visualization with ParaView | * |
| 1:30pm-5pm | Tutorial | S13: Large Scale Communication Analysis: An Essential Step in Understanding Highly Scalable Codes | * |
| 1:30pm-5pm | Tutorial | S14: Designing High-End Computing Systems with InfiniBand and 10-Gigabit Ethernet | * |
| 7:30pm-10pm | Social Event | Exhibitor Party (Exhibitors Only) | Union Station |

* Tutorial locations were not available at this printing. Please go to <http://scyourway.supercomputing.org> for room assignments.

Monday, November 16

| Time | Event | Title | Location |
|-------------|--------------|--|--------------|
| 8:30am-5pm | Tutorial | M01: A Practical Approach to Performance Analysis and Modeling of Large-Scale Systems | * |
| 8:30am-5pm | Tutorial | M02: Advanced MPI | * |
| 8:30am-5pm | Tutorial | M03: Developing Scientific Applications using Eclipse and the Parallel Tools Platform | * |
| 8:30am-5pm | Tutorial | M04: Programming using the Partitioned Global Address Space (PGAS) Model | * |
| 8:30am-5pm | Tutorial | M05: Productive Performance Engineering of Petascale Applications with POINT and VI-HPS | * |
| 8:30am-5pm | Tutorial | M06: Linux Cluster Construction | * |
| 8:30am-Noon | Tutorial | M07: Cloud Computing Architecture and Application Programming | * |
| 8:30am-Noon | Tutorial | M08: Modeling the Memory Hierarchy Performance of Current and Future Multicore Systems | * |
| 8:30am-Noon | Tutorial | M09: Hybrid MPI and OpenMP Parallel Programming | * |
| 8:30am-Noon | Tutorial | M10: Expanding Your Debugging Options | * |
| 9am-5:30pm | Workshop | 2009 Workshop on Ultra-Scale Visualization | A107 |
| 9am-5:30pm | Workshop | 2nd Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS) | A105 |
| 9am-5:30pm | Workshop | 4th Workshop on Workflows in Support of Large-Scale Science (WORKS09) | A108 |
| 9am-5:30pm | Workshop | Component-Based High Performance Computing 2009 (Day 2) | A103 |
| 9am-5:30pm | Workshop | User Experience and Advances in Bridging Multicore's Programmability Gap | A106 |
| 9am-5:30pm | Workshop | Using Clouds for Parallel Computations in Systems Biology | A104 |
| 1:30pm-5pm | Tutorial | M11: Configuring and Deploying GridFTP for Managing Data Movement in Grid/HPC Environments | * |
| 1:30pm-5pm | Tutorial | M12: Python for High Performance and Scientific Computing | * |
| 1:30pm-5pm | Tutorial | M13: OpenCL: A Standard Platform for Programming Heterogeneous Parallel Computers | * |
| 1:30pm-5pm | Tutorial | M14: Memory Debugging Parallel Applications | * |
| 7pm-9pm | Social Event | SC09 Opening Gala | Exhibit Hall |

* Tutorial locations were not available at this printing. Please go to <http://scyourway.supercomputing.org> for room assignments.

Tuesday, November 17

| Time | Event | Title | Location |
|-----------------------------|--------------------|--|---------------------------------|
| 8:30am-10am | Opening Address | The Rise of the 3D Internet: Advancements in Collaborative and Immersive Sciences by Justin Rattner (Intel Corporation) | PB253-254 |
| 10am-6pm | Exhibit | Industry and Research Exhibits | Exhibit Hall |
| 10am-6pm | Exhibit | Disruptive Technologies | Lobby area, Exhibit Halls D & E |
| 10am-6pm | Exhibit | DataCenter of the Future | Lobby area, Exhibit Halls D & E |
| 10:30am-Noon | Technical Papers | GPU/SIMD Processing <ul style="list-style-type: none"> Increasing Memory Latency Tolerance for SIMD Cores Triangular Matrix Inversion on Graphics Processing Units Auto-Tuning 3-D FFT Library for CUDA GPUs | PB252 |
| 10:30am-Noon | Technical Papers | Large-Scale Applications <ul style="list-style-type: none"> Terascale Data Organization for Discovering Multivariate Climatic Trends A Configurable Algorithm for Parallel Image-Compositing Applications Scalable Computation of Streamlines on Very Large Datasets | PB255 |
| 10:30am-Noon | Storage Challenge | Low Power Amdahl-Balanced Blades for Data Intensive Computing <ul style="list-style-type: none"> Accelerating Supercomputer Storage I/O Performance Data Intensive Science: Solving Scientific Unknowns by Solving Storage Problems An Efficient and Flexible Parallel I/O Implementation for the CFD General Notation System | PB251 |
| 10:30am-Noon | Masterworks | Future Energy Enabled by HPC <ul style="list-style-type: none"> HPC and the Challenge of Achieving a Twenty-fold Increase in Wind Energy The Outlook for Energy: Enabled with Supercomputing | PB253-254 |
| 10:30am-Noon | Exhibitor Forum | Software Tools and Libraries for C, C++ and C# <ul style="list-style-type: none"> Vector C++: C++ and Vector Code Generation by Transformation A Methodology to Parallelize Code without Parallelization Obstacles Parallelizing C+ Numerical Algorithms for Improved Performance | E147-148 |
| 10:30pm-Noon | Exhibitor Forum | Storage Solutions I <ul style="list-style-type: none"> Panasas: pNFS, Solid State Disks and RoadRunner Solving the HPC I/O Bottleneck: Sun Lustre Storage System Benefits of an Appliance Approach to Parallel File Systems | E143-144 |
| 10:30am-Noon; 1:30pm-3pm | Special Event | Building 3D Internet Experiences | D135-136 |
| 12:15pm-1:15pm | Birds-of-a-Feather | 2009 HPC Challenge Awards | E145-146 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Blue Gene/P User Forum | A103-104 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Breaking the Barriers to Parallelization at Mach Speed BoF | D139-140 |
| 12:15pm-1:15pm | Birds-of-a-Feather | CIFTS: A Coordinated Infrastructure for Fault Tolerant Systems | D137-138 |

Tuesday, November 17

| Time | Event | Title | Location |
|----------------|--------------------|--|-----------|
| 12:15pm-1:15pm | Birds-of-a-Feather | Developing and Teaching Courses in Computational Science | D133-134 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Next Generation Scalable Adaptive Graphics Environment (SAGE) for Global Collaboration | A107-108 |
| 12:15pm-1:15pm | Birds-of-a-Feather | NSF Strategic Plan for a Comprehensive National CyberInfrastructure | E141-142 |
| 1:30pm-3pm | Technical Papers | Autotuning and Compilers <ul style="list-style-type: none"> • Autotuning Multigrid with PetaBricks • Compact Multi-Dimensional Kernel Extraction for Register Tiling • Automating the Generation of Composed Linear Algebra Kernels | PB256 |
| 1:30pm-3pm | Technical Papers | Cache Techniques <ul style="list-style-type: none"> • Flexible Cache Error Protection using an ECC FIFO • A Case for Integrated Processor-Cache Partitioning in Chip Multiprocessors • Enabling Software Management for Multicore Caches with a Lightweight Hardware Support | PB255 |
| 1:30pm-3pm | Technical Papers | Sparse Matrix Computation <ul style="list-style-type: none"> • Minimizing Communication in Sparse Matrix Solvers • Implementing Sparse Matrix-Vector Multiplication on Throughput-Oriented Processors • Sparse Matrix Factorization on Massively Parallel Computers | PB252 |
| 1:30pm-3pm | Masterworks | Data Challenges in Genome Analysis <ul style="list-style-type: none"> • Big Data and Biology: The Implications of Petascale Science • The Supercomputing Challenge to Decode the Evolution and Diversity of Our Genomes | PB253-254 |
| 1:30pm-3pm | Birds-of-a-Feather | Scalable Fault-Tolerant HPC Supercomputers | D135-136 |
| 1:30pm-3pm | Exhibitor Forum | HPC Architectures: Toward Exascale Computing <ul style="list-style-type: none"> • Cray: Impelling Exascale Computing • Scalable Architecture for the Many-Core and Exascale Era | E143-144 |
| 1:30pm-3pm | Exhibitor Forum | Software Tools for Multi-core, GPUs and FPGAs <ul style="list-style-type: none"> • Acumem: Getting Multicore Efficiency • A Programming Language for a Heterogeneous Many-Core World • PGI Compilers for Heterogeneous Many-Core HPC Systems | E147-148 |
| 3:30pm-5pm | Exhibitor Forum | Networking I <ul style="list-style-type: none"> • Managing the Data Stampede: Securing High Speed, High Volume Research Networks • Juniper Networks Showcases Breakthrough 100 Gigabit Ethernet Interface for T Series Routers • Update on the Delivery of 100G Wavelength Connectivity | E147-148 |
| 3:30pm-5pm | Exhibitor Forum | Storage Solutions II <ul style="list-style-type: none"> • Storage and Cloud Challenges • Tape: Looking Ahead • Dynamic Storage Tiering: Increase Performance without Penalty | E143-144 |
| 3:30pm-5pm | Awards | ACM Gordon Bell Finalist I <ul style="list-style-type: none"> • Beyond Homogeneous Decomposition: Scaling Long-Range Forces on Massively Parallel Architectures | E145-146 |

Tuesday, November 17

Schedules

| Time | Event | Title | Location |
|---------------|--------------------|---|-----------------------|
| | | <ul style="list-style-type: none"> • A Scalable Method for Ab Initio Computation of Free Energies in Nanoscale Systems • Liquid Water: Obtaining the Right Answer for the Right Reasons | |
| 3:30pm-5pm | Technical Papers | Particle Methods <ul style="list-style-type: none"> • A Massively Parallel Adaptive Fast-Multipole Method on Heterogeneous Architectures • Efficient Band Approximation of Gram Matrices for Large Scale Kernel Methods on GPUs • Memory-Efficient Optimization of Gyrokinetic Particle-to-Grid Interpolation for Multicore Processors | PB255 |
| 3:30pm-5pm | Technical Papers | Performance Tools <ul style="list-style-type: none"> • FACT: Fast Communication Trace Collection for Parallel Applications through Program Slicing • Evaluating Similarity-Based Trace Reduction Techniques for Scalable Performance Analysis • Space-Efficient Time-Series Call-Path Profiling of Parallel Applications | PB252 |
| 3:30pm-5pm | Technical Papers | Virtual Wide-Area Networking <ul style="list-style-type: none"> • Improving GridFTP Performance Using the Phoebus Session Layer • On the Design of Scalable, Self-Configuring Virtual Networks | PB256 |
| 3:30pm-5pm | Masterworks | Finite Elements and Your Body <ul style="list-style-type: none"> • μFinite Element Analysis of Human Bone Structures • Virtual Humans: Computer Models for Vehicle Crash Safety | PB253-254 |
| 3:30pm-5pm | Special Event | 3D Cross HPC: Technology and Business Implications | PB251 |
| 5:15pm-7pm | Posters | Poster Reception and ACM Student Research Competition Posters | Oregon Ballroom Lobby |
| 5:30-7pm | Exhibitor Forum | Top 500 Supercomputers | PB253-254 |
| 5:30pm-7pm | Birds-of-a-Feather | Accelerating Discovery in Science and Engineering through Petascale Simulations and Analysis: The NSF PetaApps Program | D133-134 |
| 5:30pm-7pm | Birds-of-a-Feather | Art of Performance Tuning for CUDA and Manycore Architectures | E141-142 |
| 5:30pm-7pm | Birds-of-a-Feather | Data Curation | D137-138 |
| 5:30pm-7pm | Birds-of-a-Feather | European HPC and Grid Infrastructures | D135-136 |
| 5:30pm-7pm | Birds-of-a-Feather | Low Latency Ethernet through New Concept of RDMA over Ethernet | PB252 |
| 5:30pm-7pm | Birds-of-a-Feather | Lustre, ZFS, and End-to-End Data Integrity | E145-146 |
| 5:30pm-7pm | Birds-of-a-Feather | Micro-Threads and Exascale Systems | PB251 |
| 5:30pm-7pm | Birds-of-a-Feather | MPI Acceleration in Hardware | D139-140 |
| 5:30pm-7pm | Birds-of-a-Feather | NSF High End Computing University Research Activity (HECURA) | PB255 |
| 5:30pm-7pm | Birds-of-a-Feather | PGAS: The Partitioned Global Address Space Programming Model | PB256 |
| 5:30pm-7pm | Birds-of-a-Feather | pNFS: Parallel Storage Client and Server Development Panel Update | E143-144 |
| 5:30pm-7pm | Birds-of-a-Feather | SLURM Community Meeting | A103-104 |
| 5:30pm-7pm | Birds-of-a-Feather | Users of EnSight Visualization Software | E147-148 |
| 5:30pm-7:30pm | Birds-of-a-Feather | Productivity Tools for Multicore and Heterogeneous Systems | A107-108 |

Wednesday, November 18

| Time | Event | Title | Location |
|----------------|----------------------------------|--|------------------------------|
| 8:30am-10am | Invited Speaker/Awards | Plenary/Kennedy Awards Speakers <ul style="list-style-type: none"> • Systems Medicine, Transformational Technologies and the Emergence of Predictive, Personalized, Preventive and Participatory (P4) Medicine, Leroy Hood (Institute for Systems Biology) • Kennedy Award Presentation | PB253-254 |
| 10am-6pm | Exhibits | Industry and Research Exhibits | Exhibit Hall |
| 10am-6pm | Exhibit | Disruptive Technologies | Lobby area, Exhibit Hall D-E |
| 10am-6pm | Exhibit | Datacenter of the Future | Lobby area, Exhibit Hall D-E |
| 10:30am-Noon | Awards | Seymour Cray and Sidney Fernbach Award Presentations | PB253-254 |
| 10am-Noon | Exhibitor Forum | Software Tools: Scalable 4GL Environments <ul style="list-style-type: none"> • MATLAB: The Parallel Technical Computing Environment • Supercomputing Engine for Mathematica • Solving Large Graph-Analytic Problems from Productivity Languages with Many Hardware Accelerators | E143-144 |
| 10:30pm-Noon | Exhibitor Forum | Storage Systems, Networking and Supercomputing Applications <ul style="list-style-type: none"> • InfiniStor: Most Feature-Rich Cluster Storage System • Ethernet Data Center: Evolving to a Flat Network and a Single Switch • Smith Waterman Implementation for the SX2000 Reconfigurable Compute Platform | E147-148 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Benchmark Suite Construction for Multicore and Accelerator Architectures | B119 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Best Practices for Deploying Parallel File Systems | D137-138 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Building Parallel Applications using Microsoft's Parallel Computing Models, Tools, and Platforms | A107-108 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Deploying HPC and Cloud Computing Services for Interactive Simulation | D133-134 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Developing Bioinformatics Applications with BioHDF | D139-140 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Early Access to the Blue Waters Sustained Petascale System | A103-104 |
| 12:15pm-1:15pm | Birds-of-a-Feather | HPC Centers | B118 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Network Measurement | B117 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Open MPI Community Meeting | E145-146 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Practical HPC Considerations for Advanced CFD | E141-142 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Trends and Directions in Workload and Resource Management using PBS | D135-136 |
| 1:30pm-3pm | Award | ACM Gordon Bell Finalist II <ul style="list-style-type: none"> • Enabling High-Fidelity Neutron Transport Simulations on Petascale Architectures • Scalable Implicit Finite Element Solver for Massively Parallel Processing with Demonstration to 160K cores • 42 TFlops Hierarchical N-body Simulations on GPUs with Applications in both Astrophysics and Turbulence | D135-136 |
| 1:30pm-3pm | ACM Student Research Competition | Award Finalists <ul style="list-style-type: none"> • On the Efficacy of Haskell for High-Performance Computational Biology • A Policy Based Data Placement Service • Hiding Communication and Tuning Scientific Applications using Graph-Based Execution • An Automated Air Temperature Analysis and Prediction System for the Blue Gene/P | PB251 |

Wednesday, November 18

Schedules

| Time | Event | Title | Location |
|------------|------------------|---|-----------|
| | | <ul style="list-style-type: none"> • CUSUMMA: Scalable Matrix-Matrix Multiplication on GPUs with CUDA • BGPSStatusView Overview • Communication Optimizations of SCEC CME AWP-Olsen Application for Petascale Computing • IO Optimizations of SCEC AWP-Olsen Application for Petascale Earthquake Computing • A Hierarchical Approach for Scalability Enhancement in Distributed Network Simulations • Large-Scale Wavefront Parallelization on Multiple Cores for Sequence Alignment • A Feature Reduction Scheme for Obtaining Cost-Effective High-Accuracy Classifiers for Linear Solver Selection • Parallelization of Tau-Leaping Coarse-Grained Monte Carlo Method for Efficient and Accurate Simulations on GPUs | |
| 1:30pm-3pm | Exhibitor Forum | HPC Architectures: Microprocessor and Cluster Technology <ul style="list-style-type: none"> • Scaling Performance Forward with Intel Architecture Platforms in HPC • AMD: Enabling the Path to Cost Effective Petaflop Systems • Aurora Highlights: Green Petascale Performance | E143-144 |
| 1:30pm-3pm | Exhibitor Forum | Networking II <ul style="list-style-type: none"> • Network Automation: Advances in ROADM and GMPLS Control Plane Technology • Ethernet: <i>The</i> Converged Network • More Performance with Less Hardware through Fabric Optimization | E147-148 |
| 3:30-5pm | Exhibitor Forum | Infiniband, Memory and Cluster Technology <ul style="list-style-type: none"> • Driving InfiniBand Technology to Petascale Computing and Beyond • Meeting the Growing Demands for Memory Capacity and Available Bandwidth in Server and HPC Applications • Open High Performance and High Availability Supercomputer | E143-144 |
| 3:30pm-5pm | Exhibitor Forum | Parallel Programming and Visualization <ul style="list-style-type: none"> • HPC and Parallel Computing at Microsoft • VizSchema: A Unified Interface for Visualization of Scientific Data | E147-148 |
| 1:30pm-3pm | Panel | Cyberinfrastructure in Healthcare Management | PB252 |
| 1:30pm-3pm | Technical Papers | Acceleration <ul style="list-style-type: none"> • A 32x32x32, Spatially Distributed 3D FFT in Four Microseconds on Anton • SCAMPI: A Scalable Cam-based Algorithm for Multiple Pattern Inspection | PB255 |
| 1:30pm-3pm | Technical Papers | Grid Scheduling <ul style="list-style-type: none"> • Evaluating the Impact of Inaccurate Information in Utility-Based Scheduling • Predicting the Execution Time of Grid Workflow Applications through Local Learning • Supporting Fault-Tolerance for Time-Critical Events in Distributed Environments | PB256 |
| 1:30pm-3pm | Technical Papers | High Performance Filesystems and I/O <ul style="list-style-type: none"> • I/O Performance Challenges at Leadership Scale • Scalable Massively Parallel I/O to Task-Local Files • PLFS: A Checkpoint Filesystem for Parallel Applications | E145-146 |
| 1:30pm-3pm | Masterworks | HPC in Modern Medicine <ul style="list-style-type: none"> • Grid Technology Transforming Healthcare • Patient-specific Finite Element Modeling of Blood Flow and Vessel Wall Dynamics | PB253-254 |
| 3:30pm-5pm | Awards | ACM Gordon Bell Finalist III <ul style="list-style-type: none"> • Indexing Genomic Sequences on the IBM Blue Gene • The Cat is Out of the Bag: Cortical Simulations with 10^9 Neurons, 10^{13} Synapses • Millisecond-Scale Molecular Dynamics Simulations on Anton | D135-136 |

Wednesday, November 18

| Time | Event | Title | Location |
|------------|----------------------------|---|-----------|
| 3:30pm-5pm | Technical Papers | GPU Applications <ul style="list-style-type: none"> • Multi-core Acceleration of Chemical Kinetics for Simulation and Prediction • Towards a Framework for Abstracting Accelerators in Parallel Applications: Experience with Cell • A Microdriver Architecture for Error Correcting Codes inside the Linux Kernel | PB256 |
| 3:30pm-5pm | Technical Papers | Networking <ul style="list-style-type: none"> • HyperX: Topology, Routing, and Packaging of Efficient Large-Scale Networks • Router Designs for Elastic Buffer On-Chip Networks • Allocator Implementations for Network-on-Chip Routers | E145-146 |
| 3:30pm-5pm | Masterworks | Multi-Scale Simulations in Bioscience <ul style="list-style-type: none"> • Big Science and Computing Opportunities: Molecular Theory, Models and Simulation • Fighting Swine Flu through Computational Medicine | PB253-254 |
| 3:30pm-5pm | Doctoral Research Showcase | Doctoral Research Showcase I <ul style="list-style-type: none"> • Scalable Automatic Topology Aware Mapping for Large Supercomputers • Performance Analysis of Parallel Programs: From Multicore to Petascale • Energy Efficiency Optimizations using Helper Threads in Chip Multiprocessors • Consistency Aware, Collaborative Workflow Developer Environments • Architecture and Performance of Runtime Environments for Data Intensive Scalable Computing • Providing Access to Large Scientific Datasets on Clustered Databases | PB252 |
| 3:30pm-5pm | Panel | Disruptive Technologies: Hardware | PB251 |
| 5:30pm-7pm | Birds-of-a-Feather | Campus Champions: Your Road to Free HPC Resources | E141-142 |
| 5:30pm-7pm | Birds-of-a-Feather | Can OpenCL Save HPC? | E145-146 |
| 5:30pm-7pm | Birds-of-a-Feather | Communicating Virtual Science | E147-148 |
| 5:30pm-7pm | Birds-of-a-Feather | Eclipse Parallel Tools Platform | D137-138 |
| 5:30pm-7pm | Birds-of-a-Feather | FAST-OS | B118 |
| 5:30pm-7pm | Birds-of-a-Feather | HPC Advisory Council Initiative | PB252 |
| 5:30pm-7pm | Birds-of-a-Feather | International Exascale Software Program | PB256 |
| 5:30pm-7pm | Birds-of-a-Feather | iPlant Collaborative: Computational Scaling Challenges in Plant Biology | B117 |
| 5:30pm-7pm | Birds-of-a-Feather | MPI Forum: Preview of the MPI 3 Standard (Comment Session) | D135-136 |
| 5:30pm-7pm | Birds-of-a-Feather | Network for Earthquake Engineering Simulation (NEES): Open Invitation to Build Bridges with Related Virtual Organizations | E143-144 |
| 5:30pm-7pm | Birds-of-a-Feather | OpenMP: Evolving in an Age of Extreme Parallelism | PB255 |
| 5:30pm-7pm | Birds-of-a-Feather | OSCAR Community Meeting | B119 |
| 5:30pm-7pm | Birds-of-a-Feather | Python for High Performance and Scientific Computing | A103-104 |
| 5:30pm-7pm | Birds-of-a-Feather | Simplify Your Data Parallelization Woes with Ct: C++ for Throughput Computing | D133-134 |
| 5:30pm-7pm | Birds-of-a-Feather | Solving Interconnect Bottlenecks with Low Cost Optical Technologies | D139-140 |
| 5:30pm-7pm | Birds-of-a-Feather | Update on OpenFabrics Software (OFED) for Linux and Windows Latest Releases | PB251 |
| 5:30pm-7pm | Birds-of-a-Feather | What Programs <i>Really</i> Work for Students Interested in Research and Computing? | A107-108 |

Thursday, November 19

| Time | Event | Title | Location |
|----------------|--------------------|--|-------------------|
| 8:30pm-10am | Keynote Address | Building Solutions: Energy, Climate and Computing for a Changing World by Former U.S. Vice President Al Gore | Portland Ballroom |
| 10am-4pm | Exhibits | Industry and Research Exhibits | Exhibit Hall |
| 10:30am-Noon | Masterworks | Toward Exascale Climate Modeling <ul style="list-style-type: none"> • Toward Climate Modeling in the ExaFlop Era • Green Flash: Exascale Computing for Ultra-High Resolution Climate Modeling | PB252 |
| 10:30am-Noon | Technical Papers | Performance Analysis and Optimization <ul style="list-style-type: none"> • Performance Evaluation of NEC SX-9 using Real Science and Engineering Applications • Early Performance Evaluation of “Nehalem” Cluster using Scientific and Engineering Applications • Machine Learning-Based Prefetch Optimization for Data Center Applications | D135-136 |
| 10:30am-Noon | Technical Papers | Sustainability and Reliability <ul style="list-style-type: none"> • FALCON: A System for Reliable Checkpoint Recovery in Shared Grid Environments • Scalable Temporal Order Analysis for Large Scale Debugging • Optimal Real Number Codes for Fault Tolerant Matrix Operations | PB251 |
| 10:30am-Noon | Panel | Energy Efficient Data Centers for HPC: How Lean and Green do we need to be? | PB256 |
| 10:30am-Noon | Exhibitor Forum | Grid Computing, Cyber Infrastructures and Benchmarking <ul style="list-style-type: none"> • Building a Real Business Model around the Distributed Grid • Contribution of Cyberinfrastructure to Economic Development in South Africa: Update on Developments • Common Application Benchmarks on Current Hardware Platforms | E147-148 |
| 10:30am-Noon | Exhibitor Forum | Virtualization and Cloud Computing <ul style="list-style-type: none"> • High-End Virtualization as a Key Enabler for the HPC Cloud or HPC as a Service • Managing HPC Clouds • Maximizing the Potential of Virtualization and the Cloud: How to Unlock the Traditional Storage Bottleneck | E143-144 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Energy Efficient High Performance Computing Working Group | D139-140 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Extending Global Arrays to Future Architectures | B118 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Getting Started with Institution-Wide Support for Supercomputing | D133-134 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Green500 List | A107-108 |
| 12:15pm-1:15pm | Birds-of-a-Feather | HDF5: State of the Union | D137-138 |
| 12:15pm-1:15pm | Birds-of-a-Feather | HPC Saving the Planet, One Ton of CO ₂ at a Time | D137-138 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Jülich Research on Petaflops Architectures Project | D135-136 |
| 12:15pm-1:15pm | Birds-of-a-Feather | MPICH: A High-Performance Open-Source MPI Implementation | E145-146 |
| 12:15pm-1:15pm | Birds-of-a-Feather | Securing High Performance Government Networks with Open Source Deep Packet Inspection Applications | B119 |

Thursday, November 19

| Time | Event | Title | Location |
|----------------|--------------------|---|----------|
| 12:15pm-1:15pm | Birds-of-a-Feather | What's New about INCITE in 2010? | B117 |
| 1:30pm-3pm | Technical Papers | SmartStore: A New Metadata Organization Paradigm with Semantic-Awareness for Next-Generation File Systems <ul style="list-style-type: none"> • Adaptive and Scalable Metadata Management to Support A Trillion Files • Dynamic Storage Cache Allocation in Multi-Server Architectures | PB251 |
| 1:30pm-3pm | Technical Papers | Multicore Task Scheduling <ul style="list-style-type: none"> • Dynamic Task Scheduling for Linear Algebra Algorithms on Distributed-Memory Multicore • PFunc: Modern Task Parallelism for Modern High Performance Computing • Age-Based Scheduling for Asymmetric Multiprocessors | PB255 |
| 1:30pm-3pm | Technical Papers | System Performance Evaluation <ul style="list-style-type: none"> • Instruction-Level Simulation of a Cluster at Scale • Diagnosing Performance Bottlenecks in Emerging Petascale Applications • Comparative Study of One-Sided Factorizations with Multiple Software Packages on Multi-Core Hardware | PB256 |
| 1:30pm-3pm | Technical Papers | Dynamic Task Scheduling <ul style="list-style-type: none"> • VGrADS: Enabling e-Science Workflows on Grids and Clouds with Fault Tolerance • GridBot: Execution of Bags of Tasks in Multiple Grids • Scalable Work Stealing | E145-146 |
| 1:30am-3pm | Exhibitor Forum | GPUs and Software Tools for Heterogeneous Architectures <ul style="list-style-type: none"> • Tesla: Fastest Processor Adoption in HPC History • Debugging the Future: GPUs and Petascale • Developing Software for Heterogeneous and Accelerated Systems | E143-144 |
| 1:30am-3pm | Exhibitor Forum | HPC Architectures: Future Technologies and Systems <ul style="list-style-type: none"> • Convey's Hybrid-Core Computing: Breaking Through the Power/Performance Wall • Fujitsu's Technologies for Sustained Petascale Computing • Next Generation High Performance Computer | E143-144 |
| 1:30pm-3pm | Masterworks | High Performance at Massive Scale <ul style="list-style-type: none"> • Warehouse-Scale Computers • High Performance at Massive Scale: Lessons Learned at Facebook | PB252 |
| 3:30pm-5pm | Masterworks | Scalable Algorithms and Applications <ul style="list-style-type: none"> • Scalable Parallel Solvers in Computational Electrophysiology • Simulation and Animation of Complex Flows on 10,000 Processor Cores | PB252 |
| 3:30pm-5pm | Technical Papers | Future Scaling of Processor-Memory Interfaces <ul style="list-style-type: none"> • A Design Methodology for Domain-Optimized Power-Efficient Supercomputing • Leveraging 3D PCRAM Technologies to Reduce Checkpoint Overhead for Future Exascale Systems | PB256 |
| 3:30pm-5pm | Technical Papers | Future HPC Architectures <ul style="list-style-type: none"> • Future Scaling of Processor-Memory Interfaces • A Design Methodology for Domain-Optimized Power-Efficient Supercomputing • Leveraging 3D PCRAM Technologies to Reduce Checkpoint Overhead for Future Exascale Systems | PB256 |

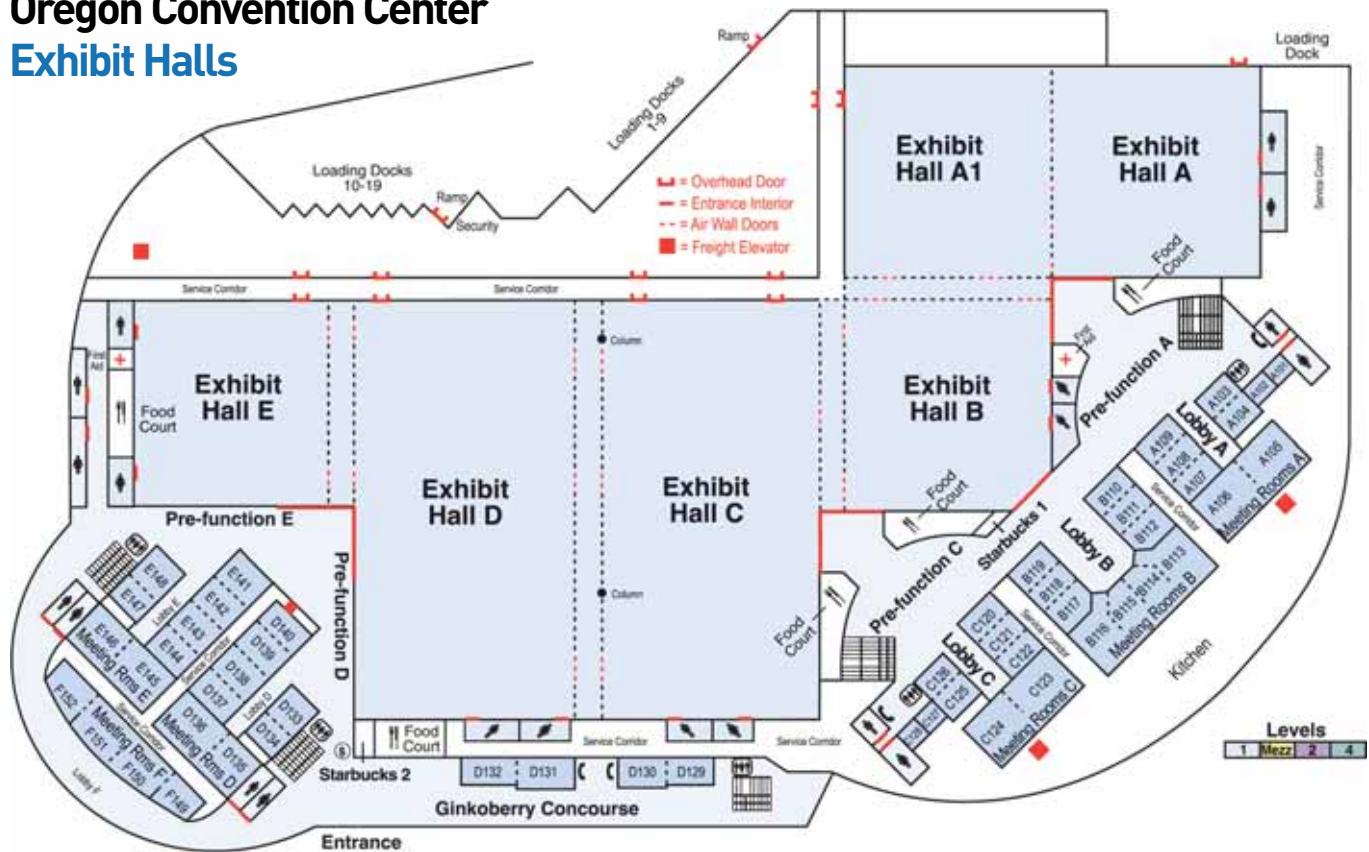
Thursday, November 19

| Time | Event | Title | Location |
|------------|-------------------|---|---|
| 3:30pm-5pm | Exhibitor Forum | Technologies for Data and Computer Centers <ul style="list-style-type: none"> • Effective Data Center Physical Infrastructure Management • Using Air and Water Cooled Miniature Loop Heat Pipes to Save Up to 50% in Cluster and Data Center Cooling Costs • 48V VR12 Solution for High Efficiency Data Centers | E147-148 |
| 3:30pm-5pm | Doctoral Research | Doctoral Research Showcase II <ul style="list-style-type: none"> • Processing Data Intensive Queries in Global-Scale Scientific Database Federations • GPGPU and Cloud Computing for DNA Sequence Analysis • Providing QoS for Heterogeneous Workloads in Large, Volatile, and Non-Dedicated Distributed Systems • Computer Generation of FFT Libraries for Distributed Memory Computing Platforms • Adaptive Runtime Optimization of MPI Binaries • An Integrated Framework for Parameter-Based Optimization of Scientific Workflows | PB251 |
| 3:30pm-5pm | Panel | Disruptive Technologies: Software | PB255 |
| 6pm-9pm | Social Event | SC09 Conference Reception | Portland Center for the Performing Arts |

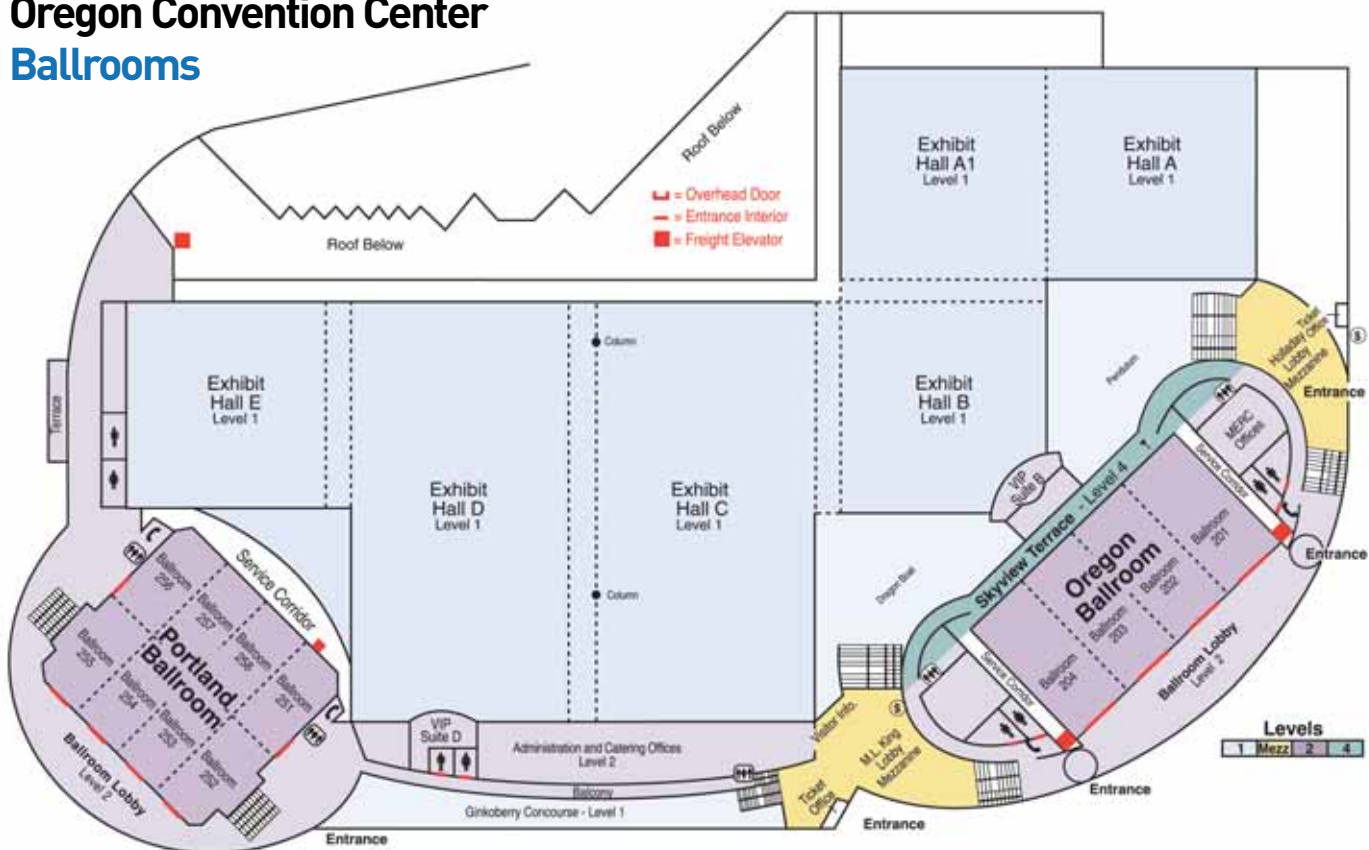
Friday, November 20

| Time | Event | Title | Location |
|----------------|----------|---|----------|
| 8:30am-5pm | Workshop | ATIP 1st Workshop on HPC in India: Research Challenges on Computing in India | E141-142 |
| 8:30am-5pm | Workshop | Grid Computing Environments | D139-140 |
| 8:30am-10pm | Panel | Applications Architecture Power Puzzle | PB252 |
| 8:30am-10pm | Panel | Flash Technology in HPC: Let the Revolution Begin | PB251 |
| 10:30am-Noon | Panel | Preparing the World for Ubiquitous Parallelism | PB252 |
| 10:30am-Noon | Panel | The Road to Exascale: Hardware and Software Challenges | PB251 |
| 10:30am-1:30pm | Workshop | Early Adopters PhD Workshop: Building the Next Generation of Application Scientists | D137-138 |

Oregon Convention Center Exhibit Halls



Oregon Convention Center Ballrooms



Maps





Computing for a Changing World

Invited Speakers

This year SC09 features three invited presentations by internationally recognized leaders in their fields that discuss challenges and opportunities in high-performance computing, featuring innovative uses of high performance computing and networking.

Invited Speakers

Invited Speakers

Tuesday, November 17

Opening Address

8:30am-10am

Room: PB253-254



The Rise of the 3D Internet: Advancements in Collaborative and Immersive Sciences

Speaker: Justin Rattner (Intel Corporation)



Forty Exabytes of unique data will be generated worldwide in 2009. This data can help us understand scientific and engineering phenomenon as well as operational trends in business and finance. The best way to understand, navigate and communicate these phenomena is through visualization. In his opening address, Intel CTO Justin Rattner will talk about today's data deluge and how high performance computing is being used to deliver cutting edge, 3D collaborative visualizations. He will also discuss how the 2D Internet started and draw parallels to the rise of the 3D Internet today. With the help of demonstrations, he will show how rich visualization of scientific data is being used for discovery, collaboration and education.

Wednesday, November 18

Plenary and Kennedy Award Speakers

8:30am-10am

Room: PB253-254



Plenary: Systems Medicine, Transformational Technologies and the Emergence of Predictive, Personalized, Preventive and Participatory (P4) Medicine

Speaker: Lee Hood (Institute for Systems Biology)



The challenge for biology in the 21st century is the need to deal with its incredible complexity. One powerful way to think of biology is to view it as an informational science. This view leads to the conclusion that biological information is captured, mined, integrated by biological networks and finally passed off to molecular machines for execution. Hence the challenge in understanding biological complexity is that of deciphering the operation of dynamic biological networks across the three time scales of life's evolution, development and physiological

responses. Systems approaches to biology are focused on delineating and deciphering dynamic biological networks and their interactions with simple and complex molecular machines. I will focus on our efforts at a systems approach to disease—looking at prion disease in mice. We have just published a study that has taken more than 5 years that lays out the principles of a systems approach to disease including dealing with the striking signal to noise problems of high throughput biological measurements and biology itself (e.g. polymorphisms). I will also discuss the emerging technologies (measurement and visualization) that will transform medicine over the next 10 years, including next generation DNA sequencing, microfluidic protein chips and single-cell analyses. I will also discuss some of the computational and mathematical challenges that are fundamental to the revolution in medicine—those that deal with medical sciences and those that deal in a general way with healthcare. It appears that systems medicine, together with pioneering changes such as next-generation DNA sequencing and blood protein measurements (nanotechnology), as well as the development of powerful new computational and mathematical tools, will transform medicine over the next 5-20 years from its currently reactive state to a mode that is predictive, personalized, preventive and participatory (P4). This will in turn lead to the digitalization of medicine, with ultimately a profound decrease in the cost of healthcare. It will also transform the business strategies for virtually every sector of the health care industry. These considerations have led ISB to begin formulating a series of national and international strategic partnerships that are focused on accelerating the realization of P4 medicine. I will discuss some of these strategic partnerships and discuss the implications for healthcare arising from P4 medicine.

Kennedy Award Presentation

The new Ken Kennedy Award recognizes substantial contributions to programmability and productivity in computing and substantial community service or mentoring contributions. The award honors the remarkable research, service, and mentoring contributions of the late Ken Kennedy. This is the first presentation of this award, which includes a \$5,000 honorarium.

This award is co-sponsored by ACM and IEEE Computer Society.

Cray/Fernbach Award Speakers

10:30-Noon

Room: PB253-254

Cray Award Presentation

The Seymour Cray Computer Science and Engineering Award recognizes innovative contributions to HPC systems that best exemplify the creative spirit of Seymour Cray. Sponsored by IEEE Computer Society, this prestigious honor is presented annually during a special session held at the SC conference.

Fernbach Award Presentation

The Sidney Fernbach Memorial Award honors innovative uses of HPC in problem solving. Sponsored by IEEE Computer Society, this prestigious honor is presented annually during a special session held at the SC conference.

Thursday, November 19

Keynote Address

8:30am-10am

Room: Portland Ballroom



Building Solutions, Energy, Climate and Computing for a Changing World

Speaker: The Honorable Al Gore, 45th Vice President of the United States, Nobel Laureate, Author (An Inconvenient Truth), Chairman, Generation Investment Management, and Chairman, Current TV



Biography: The world's most influential voice on climate change, an advisor to the President of the United States, leaders in Congress, and heads of state throughout the world, Former Vice President Al Gore offers a unique perspective on national and international affairs.

Vice President Gore is co-founder and Chairman of Generation Investment Management, a firm that is focused on a new approach to Sustainable Investing.

He is also co-founder and Chairman of Current TV, an independently owned cable and satellite television network for young people based on viewer-created content and citizen journalism. A member of the Board of Directors of Apple Computer, Inc. and a Senior Advisor to Google, Inc. Gore is also Visiting Professor at Middle Tennessee State University in Murfreesboro, Tennessee. Vice President Gore is the author of *An Inconvenient Truth*, a

best-selling book on the threat of and solutions to global warming, and the subject of the movie of the same title, which has already become one of the top documentary films in history. In 2007, *An Inconvenient Truth* was awarded two Academy Awards for Best Documentary Feature and Best Original Song.

Since his earliest days in the U. S. Congress 30 years ago, Al Gore has been the leading advocate for confronting the threat of global warming. His pioneering efforts were outlined in his best-selling book *Earth in the Balance: Ecology and the Human Spirit* (1992). He led the Clinton-Gore Administration's efforts to protect the environment in a way that also strengthens the economy.

Al Gore was born on March 31, 1948, the son of former U.S. Senator Albert Gore, Sr. and Pauline Gore. Raised in Carthage, Tennessee, and Washington, D.C., he received a degree in government with honors from Harvard University in 1969. After graduation, he volunteered for enlistment in the U.S. Army and served in the Vietnam War. Upon returning from Vietnam, Al Gore became an investigative reporter with the Tennessean in Nashville, where he also attended Vanderbilt University's Divinity School and then Law School.

Computing for a Changing World



Portland,
OR | 2009





Computing for a Changing World

Papers

The SC09 Technical Papers program received 261 high quality submissions this year covering a variety of advanced research topics in HPC, spanning six areas: Applications, Architecture/Networks, Grids/Clouds, Performance, Storage and Systems Software. After an extremely rigorous peer review process, in which all the papers were subject to at least four careful reviews, a two-day face-to-face committee meeting was held on June 1-2 in Portland, physically attended by over 100 technical paper committee members to discuss every paper and finalize the selections. The quality of the paper being exceptionally high this year sometimes caused extensive but civil debates as SC strives to further improve its reputation as the top HPC conference. At the conclusion of the meeting, 59 papers were

accepted for presentation, covering hot topics of today, such as how multicore/GPUs/storage/sustainability/clouds are leading to the exascale era of the future, resulting in one of the most exciting papers program in the history of SC.

Amongst the excellent sets of work, three outstanding papers were selected as the "Best Paper Finalists," and four papers as the "Best Student Paper Finalists."

The Best Paper and the Best Student Paper awards, along with other technical program awards, will be announced at the conference awards ceremony (invitation only) at noon on Thursday.

Tuesday, November 17

GPU/SIMD Processing

Chair: John L. Gustafson (Intel Corporation)

10:30am-Noon

Room: PB252

Increasing Memory Latency Tolerance for SIMD Cores

Authors: David Tarjan (University of Virginia), Jiayuan Meng (University of Virginia), Kevin Skadron (University of Virginia)

Manycore processors with wide SIMD cores are becoming a popular choice for the next generation of throughput oriented architectures. We introduce a hardware technique called "diverge on miss" that allows SIMD cores to better tolerate memory latency for workloads with non-contiguous memory access patterns. Individual threads within a SIMD "warp" are allowed to slip behind other threads in the same warp, letting the warp continue execution even if a subset of threads are waiting on memory. Diverge on miss can either increase the performance of a given design by up to a factor of 3.14 for a single warp per core, or reduce the number of warps per core needed to sustain a given level of performance from 16 to 2 warps, reducing the area per core by 35%.

Awards: Best Student Paper Finalist

Triangular Matrix Inversion on Graphics Processing Units

Authors: Florian Ries (University of Bologna), Tommaso De Marco (University of Bologna), Matteo Zivieri (University of Bologna), Roberto Guerrieri (University of Bologna)

Dense matrix inversion is a basic procedure in many linear algebra algorithms. A computationally arduous step in most dense matrix inversion methods is the inversion of triangular matrices as produced by factorization methods such as LU decomposition. In this paper, we demonstrate how triangular matrix inversion (TMI) can be accelerated considerably by using commercial Graphics Processing Units (GPU) in a standard PC. Our implementation is based on a divide and conquer type recursive TMI algorithm, efficiently adapted to the GPU architecture. Our implementation obtains a speedup of 34x versus a CPU-based LAPACK reference routine, and runs at up to 54 gigaflops/s on a GTX 280 in double precision. Limitations of the algorithm are discussed, and strategies to cope with them are introduced. In addition, we show how inversion of an L- and U-matrix can be performed concurrently on a GTX 295 based dual-GPU system at up to 90 gigaflops/s.

Auto-Tuning 3-D FFT Library for CUDA GPUs

Authors: Akira Nukada (Tokyo Institute of Technology), Satoshi Matsuoka (Tokyo Institute of Technology)

Existing implementations of FFTs on GPUs are optimized for specific transform sizes like powers of two, and exhibit unstable and peaky performance i.e., do not perform as well in other sizes that appear in practice. Our new auto-tuning 3-D FFT on CUDA generates high performance CUDA kernels for FFTs of varying transform sizes, alleviating this problem. Although auto-tuning has been implemented on GPUs for dense kernels such as DGEMM and stencils, this is the first instance that has been applied comprehensively to bandwidth intensive and complex kernels such as 3-D FFTs. Bandwidth intensive optimizations such as selecting the number of threads and inserting padding to avoid bank conflicts on shared memory are systematically applied. Our resulting auto-tuner is fast and results in performance that essentially beats all 3-D FFT implementations on a single processor to date, and moreover exhibits stable performance irrespective of problem sizes or the underlying GPU hardware.

Large-Scale Applications

Chair: Michael A Heroux (Sandia National Laboratories)

10:30am-Noon

Room: PB255

Terascale Data Organization for Discovering Multivariate Climatic Trends

Authors: Wesley Kendall (University of Tennessee, Knoxville), Markus Glatter (University of Tennessee, Knoxville), Jian Huang (University of Tennessee, Knoxville), Tom Peterka (Argonne National Laboratory), Robert Latham (Argonne National Laboratory), Robert Ross (Argonne National Laboratory)

Current visualization tools lack the ability to perform full-range spatial and temporal analysis on terascale scientific datasets. Two key reasons exist for this shortcoming: I/O and postprocessing on these datasets are being performed in suboptimal manners, and the subsequent data extraction and analysis routines have not been studied in depth at large scales. We resolved these issues through advanced I/O techniques and improvements to current query-driven visualization methods. We show the efficiency of our approach by analyzing over a terabyte of multivariate satellite data and addressing two key issues in climate science: time-lag analysis and drought assessment. Our methods allowed us to reduce the end-to-end execution times on these problems to one minute on a Cray XT4 machine.



A Configurable Algorithm for Parallel Image-Compositing Applications

Authors: Tom Peterka (Argonne National Laboratory), David Goodell (Argonne National Laboratory), Robert Ross (Argonne National Laboratory), Han-Wei Shen (Ohio State University), Rajeev Thakur (Argonne National Laboratory)

Collective communication operations can dominate the cost of large scale parallel algorithms. Image compositing in parallel scientific visualization is one such reduction operation where this is the case. We present a new algorithm that in many cases performs better than existing compositing algorithms. It can do this via a set of configurable parameters, the radices, that determine the number of communication partners in each message round. The algorithm embodies and unifies binary swap and direct-send, two of the best-known compositing methods, and enables numerous other configurations via appropriate choices of radices. While general-purpose and not tied to a particular computing architecture or network topology, the selection of radix values allows the algorithm to take advantage of new supercomputer interconnect features such as multi-porting. We show scalability across image size and system size, including both powers of two and non-powers of two process counts.

Scalable Computation of Streamlines on Very Large Datasets

Authors: David Pugmire (Oak Ridge National Laboratory), Hank Childs (Lawrence Berkeley National Laboratory), Christoph Garth (University of California, Davis), Sean Ahern (Oak Ridge National Laboratory), Gunther Weber (Lawrence Berkeley National Laboratory)

Understanding vector fields resulting from large scientific simulations is an important and often difficult task. Streamlines, curves that are tangential to a vector field at each point, are a powerful visualization method in this context. Application of streamline-based visualization to very large vector field data represents a significant challenge due to the non-local and data-dependent nature of streamline computation, and requires careful balancing of computational demands placed on I/O, memory, communication, and processors. In this paper we review two parallelization approaches based on established parallelization paradigms (static decomposition and on-demand loading) and present a novel hybrid algorithm for computing streamlines. Our algorithm is aimed at good scalability and performance across the widely varying computational characteristics of streamline-based problems. We perform performance and scalability studies of all three algorithms on a number of prototypical application problems and demonstrate that our hybrid scheme is able to perform well in different settings.

Autotuning and Compilers

Chair: Barbara Chapman (University of Houston)

Time: 1:30pm-3pm

Room: PB256

Autotuning Multigrid with PetaBricks

Authors: Cy P Chan (Massachusetts Institute of Technology), Jason Ansel (Massachusetts Institute of Technology), Yee Lok Wong (Massachusetts Institute of Technology), Saman Amarasinghe (Massachusetts Institute of Technology), Alan Edelman (Massachusetts Institute of Technology)

Algorithmic choice is essential in any problem domain to realizing optimal computational performance. We present a programming language and autotuning system that address issues of algorithmic choice, accuracy requirements, and portability for multigrid applications in a near-optimal and efficient manner. We search the space of algorithmic choices and cycle shapes efficiently by utilizing a novel dynamic programming method to build tuned algorithms from the bottom up. The results are optimized multigrid algorithms that invest targeted computational power to yield the accuracy required by the user. Our implementation uses PetaBricks, an implicitly parallel programming language where algorithmic choices are exposed in the language. The PetaBricks compiler uses these choices to analyze, autotune, and verify the PetaBricks program. These language features, most notably the autotuner, were key in enabling our implementation to be clear, correct, and fast.

Compact Multi-Dimensional Kernel Extraction for Register Tiling

Authors: Lakshminarayanan Renganarayanan (IBM T.J. Watson Research Center), Uday Bondhugula (IBM T.J. Watson Research Center), Salem Derisavi (IBM Toronto Research Laboratory), Alexandre E. Eichenberger (IBM T.J. Watson Research Center), Kevin O'Brien (IBM T.J. Watson Research Center)

To achieve high performance on multi-cores, modern loop optimizers apply long sequences of transformations that produce complex loop structures. Downstream optimizations such as register tiling (unroll-and-jam plus scalar promotion) typically provide a significant performance improvement. Typical register tilers provide this performance improvement only when applied on simple loop structures. They often fail to operate on complex loop structures leaving a significant amount of performance on the table. We present a technique called compact multi-dimensional kernel extraction (COMDEX) which can make register tilers operate on arbitrarily complex loop structures and enable them to provide the performance benefits. COMDEX extracts

compact unrollable kernels from complex loops. We show that by using COMDEX as a pre-processing to register tiling we can (i) enable register tiling on complex loop structures and (ii) realize a significant performance improvement on a variety of codes.

Automating the Generation of Composed Linear Algebra Kernels

Authors: Geoffrey Belter (University of Colorado), E. R. Jessup (University of Colorado), Ian Karlin (University of Colorado), Jeremy G. Siek (University of Colorado)

Memory bandwidth limits the performance of important kernels in many scientific applications. Such applications often use sequences of Basic Linear Algebra Subprograms (BLAS), and highly efficient implementations of those routines enable scientists to achieve high performance at little cost. However, tuning the BLAS in isolation misses opportunities for memory optimization that result from composing multiple subprograms. Because it is not practical to create a library of all BLAS combinations, we have developed a domain-specific compiler that generates them on demand. In this paper, we describe a novel algorithm for compiling linear algebra kernels and searching for the best combination of optimization choices. We also present a new hybrid analytic/empirical method for quickly evaluating the profitability of each optimization. We report experimental results showing speedups of up to 130% relative to the GotoBLAS on an AMD Opteron and up to 137% relative to MKL on an Intel Core 2.

Cache Techniques

Chair: William Harrod (DARPA)

Time: 1:30pm-3pm

Room: PB255

Flexible Cache Error Protection using an ECC FIFO

Authors: Doe Hyun Yoon (University of Texas at Austin), Mattan Erez (University of Texas at Austin)

We present ECC FIFO, a mechanism enabling two-tiered last-level cache error protection using an arbitrarily strong tier-2 code without increasing on-chip storage. Instead of adding redundant ECC information to each cache line, our ECC FIFO mechanism off-loads the extra information to off-chip DRAM. We augment each cache line with a tier-1 code, which provides error detection consuming limited resources. The redundancy required for strong protection is provided by a tier-2 code placed in off-chip memory. Because errors that require tier-2 correction are rare, the overhead of accessing DRAM is unimportant. We show

how this method can save 15 _ 25% and 10 _ 17% of on-chip cache area and power respectively while minimally impacting performance, which decreases by 1% on average across a range of scientific and consumer benchmarks.

A Case for Integrated Processor-Cache Partitioning in Chip Multiprocessors

Authors: Shekhar Srikantaiah (Pennsylvania State University), Reetuparna Das (Pennsylvania State University), Asit K. Mishra (Pennsylvania State University), Chita R. Das (Pennsylvania State University), Mahmut Kandemir (Pennsylvania State University)

This paper examines an operating system directed integrated processor-cache partitioning scheme that partitions both the available processors and the shared last level cache in a chip multiprocessor among different multi-threaded applications. Extensive simulations using a full system simulator and a set of multiprogrammed workloads show that our integrated processor-cache partitioning scheme facilitates achieving better performance isolation as compared to state of the art hardware/software based solutions. Specifically, our integrated processor-cache partitioning approach performs, on an average, 20.83% and 14.14% better than equal partitioning and the implicit partitioning enforced by the underlying operating system, respectively, on the fair speedup metric on an 8 core system. We also compare our approach to processor partitioning alone and a state-of-the-art cache partitioning scheme and our scheme fares 8.21% and 9.19% better than these schemes on a 16 core system.

Enabling Software Management for Multicore Caches with a Lightweight Hardware Support

Authors: Jiang Lin (Iowa State University), Qingda Lu (Ohio State University), Xiaoning Ding (Ohio State University), Zhao Zhang (Iowa State University), Xiaodong Zhang (Ohio State University), Ponnuswamy Sadayappan (Ohio State University)

The management of shared caches in multicore processors is a critical and challenging task. Many hardware and OS-based methods have been proposed. However, they may be hardly adopted in practice due to their non-trivial overheads, high complexities, and/or limited abilities to handle increasingly complicated scenarios of cache contention caused by many-cores. In order to turn cache partitioning methods into reality in the management of multicore processors, we propose to provide an affordable and lightweight hardware support to coordinate with OS-based cache management policies. The proposed methods are scalable to many-cores, and perform comparably with other proposed hardware solutions, but have much lower overheads, therefore can be easily adopted in commodity processors. Having

conducted extensive experiments with 37 multi-programming workloads, we show the effectiveness and scalability of the proposed methods. For example on 8-core systems, one of our proposed policies improves performance over LRU-based hardware cache management by 14.5% on average.

Sparse Matrix Computation

Chair: George Biros (Georgia Institute of Technology)

Time: 1:30pm-3pm

Room: PB252

Minimizing Communication in Sparse Matrix Solvers

Authors: Marghoob Mohiyuddin (University of California, Berkeley), Mark Hoemmen (University of California, Berkeley), James Demmel (University of California, Berkeley), Katherine Yelick (University of California, Berkeley)

Data communication within the memory system of a single processor node and between multiple nodes in a system is the bottleneck in many iterative sparse matrix solvers like CG and GMRES: here k iterations of a conventional implementation perform k sparse-matrix-vector-multiplications and $\Omega(k)$ vector operations like dot products, resulting in communication that grows by a factor of $\Omega(k)$ in both the memory and network. By reorganizing the sparse-matrix kernel to compute a set of matrix-vector products at once and reorganizing the rest of the algorithm accordingly, we can perform k iterations by sending $O(\log P)$ messages instead of $O(k \log P)$ messages on a parallel machine, and reading matrix A from DRAM to cache just once instead of k times on a sequential machine. This reduces communication to the minimum possible. We combine these techniques to implement GMRES on an 8-core Intel Clovertown, and get speedups of up to 4.3x over standard GMRES, without sacrificing convergence rate or numerical stability.

Implementing Sparse Matrix-Vector Multiplication on Throughput-Oriented Processors

Authors: Nathan Bell (NVIDIA Research), Michael Garland (NVIDIA Research)

Sparse matrix-vector multiplication (SpMV) is of singular importance in sparse linear algebra. In contrast to the uniform regularity of dense linear algebra, sparse operations encounter a broad spectrum of matrices ranging from the regular to the highly irregular. Harnessing the tremendous potential of throughput-oriented processors for sparse operations requires that we expose substantial fine-grained parallelism and impose sufficient regularity on execution paths and memory access patterns. We explore SpMV methods that are well-suited to throughput-oriented

architectures like the GPU and which exploit several common sparsity classes. The techniques we propose are efficient, successfully utilizing large percentages of peak bandwidth. Furthermore, they deliver excellent total throughput, averaging 16 GFLOP/s and 10 GFLOP/s in double precision for structured grid and unstructured mesh matrices, respectively, on a GeForce GTX 285. This is roughly 2.8 times the throughput previously achieved on Cell BE and more than 10 times that of a quad-core Intel Clovertown system.

Sparse Matrix Factorization on Massively Parallel Computers

Authors: Anshul Gupta (IBM T.J. Watson Research Center), Seid Koric (National Center for Supercomputing Applications), Thomas George (Texas A&M University)

Direct methods for solving sparse systems of linear equations have a high asymptotic computational and memory requirements relative to iterative methods. However, systems arising in some applications, such as structural analysis, can often be too ill-conditioned for iterative solvers to be effective. We cite real applications where this is indeed the case, and using matrices extracted from these applications to conduct experiments on three different massively parallel architectures, show that a well designed sparse factorization algorithm can attain very high levels of performance and scalability. We present strong scalability results for test data from real applications on up to 8,192 cores, along with both analytical and experimental weak scalability results for a model problem on up to 16,384 core—an unprecedented number for sparse factorization. For the model problem, we also compare experimental results with multiple analytical scaling metrics and distinguish between some commonly used weak scaling methods.

Particle Methods

Chair: Subhash Saini (NASA Ames Research Center)

Time: 3:30pm-5pm

Room: PB255

A Massively Parallel Adaptive Fast-Multipole Method on Heterogeneous Architectures

Authors: Ilya Lashuk (Georgia Institute of Technology), Aparna Chandramowlishwaran (Georgia Institute of Technology), Harper Langston (Georgia Institute of Technology), Tuan-Anh Nguyen (Georgia Institute of Technology), Rahul Sampath (Georgia Institute of Technology), Aashay Shringarpure (Georgia Institute of Technology), Rich Vuduc (Georgia Institute of Technology), Lexing Ying (University of Texas at Austin), Denis Zorin (New York University), George Biros (Georgia Institute of Technology)

We present new scalable algorithms and a new implementation of our kernel-independent fast multipole method (Ying et al. ACM/IEEE SC '03), in which we employ both distributed memory parallelism (via MPI) and shared memory/streaming parallelism (via GPU acceleration) to rapidly evaluate two-body non-oscillatory potentials. On traditional CPU-only systems, our implementation scales well up to 30 billion unknowns on 65K cores (AMD/CRAY-based Kraken system at NSF/NICS) for highly non-uniform point distributions. On GPU-enabled systems, we achieve 30X speedup for problems of up to 256 million points on 256 GPUs (Lincoln at NSF/NCSA) over a comparable CPU-only based implementations. We use a new MPI-based tree construction and partitioning, and a new reduction algorithm for the evaluation phase. For the sub-components of the evaluation phase, we use NVIDIA's CUDA framework to achieve excellent performance. Taken together, these components show promise for ultrascale FMM in the petascale era and beyond.

Award: Best Paper Finalist

Efficient Band Approximation of Gram Matrices for Large Scale Kernel Methods on GPUs

Authors: Mohamed Hussein (University of Maryland), Wael Abd-Elmageed (University of Maryland)

Kernel-based methods require $O(N^2)$ time and space complexities to compute and store non-sparse Gram matrices, which is prohibitively expensive for large scale problems. We introduce a novel method to approximate a Gram matrix with a band matrix. Our method relies on the locality preserving properties of space filling curves, and the special structure of Gram matrices. Our approach has several important merits. First, it computes only those elements of the Gram matrix that lie within the

projected band. Second, it is simple to parallelize. Third, using the special band matrix structure makes it space efficient and GPU-friendly. We developed GPU implementations for the Affinity Propagation (AP) clustering algorithm using both our method and the COO sparse representation. Our band approximation is about 5 times more space efficient and faster to construct than COO. AP gains up to 6x speedup using our method without any degradation in its clustering performance.

Memory-Efficient Optimization of Gyrokinetic Particle-to-Grid Interpolation for Multicore Processors

Authors: Kamesh Madduri (Lawrence Berkeley National Laboratory), Samuel Williams (Lawrence Berkeley National Laboratory), Stephane Ethier (Princeton Plasma Physics Laboratory), Leonid Oliker (Lawrence Berkeley National Laboratory), John Shalf (Lawrence Berkeley National Laboratory), Erich Strohmaier (Lawrence Berkeley National Laboratory), Katherine Yelick (Lawrence Berkeley National Laboratory)

We present multicore parallelization strategies for the particle-to-grid interpolation step in the Gyrokinetic Toroidal Code (GTC), a 3D particle-in-cell (PIC) application to study turbulent transport in magnetic-confinement fusion devices. Particle-grid interpolation is a known performance bottleneck in several PIC applications. In GTC, this step involves particles depositing charges to a 3D toroidal mesh, and multiple particles may contribute to the charge at a grid point. We design new parallel algorithms for the GTC charge deposition kernel, and analyze their performance on three leading multicore platforms. We implement thirteen different variants for this kernel and identify the best-performing ones given typical PIC parameters such as the grid size, number of particles per cell, and the GTC-specific particle Larmor radius variation. We find that our best strategies can be 2X faster than the reference optimized MPI implementation, and our analysis provides insight into desirable architectural features for high-performance PIC simulation codes.

Performance Tools

Chair: David Lowenthal (University of Arizona)

Time: 3:30pm-5pm

Room: PB252

FACT: Fast Communication Trace Collection for Parallel Applications through Program Slicing

Authors: Jidong Zhai (Tsinghua University), Tianwei Sheng (Tsinghua University), Jiangzhou He (Tsinghua University), Wenguang Chen (Tsinghua University), Weimin Zheng (Tsinghua University)

Communication pattern of parallel applications is important to optimize application performance and design better communication subsystems. Communication patterns can be obtained by analyzing communication traces. However, existing approaches to generate communication traces need to execute the entire parallel applications on full-scale systems which are time-consuming and expensive. We propose a novel technique, called FACT, which can perform FAsT Communication Trace collection for large-scale parallel applications on small-scale systems. Our idea is to reduce the original program to obtain a program slice through static analysis, and to execute the program slice to acquire communication traces. We have implemented FACT and evaluated it with NPB programs and Sweep3D. The results show that FACT can reduce resource consumptions by two orders of magnitude in most cases. For example, FACT collects communication traces of 512-process Sweep3D in just 6.79 seconds, consuming 1.25GB memory, while the original program takes 256.63 seconds and consumes 213.83GB memory.

Evaluating Similarity-Based Trace Reduction Techniques for Scalable Performance Analysis

Authors: Kathryn Mohror (Portland State University), Karen L. Karavanic (Portland State University)

Event traces are required to correctly diagnose a number of performance problems that arise on today's highly parallel systems. Unfortunately, the collection of event traces can produce a large volume of data that is difficult, or even impossible, to store and analyze. One approach for compressing a trace is to identify repeating trace patterns and retain only one representative of each pattern. However, determining the similarity of sections of traces, i.e., identifying patterns, is not straightforward. In this paper, we investigate pattern-based methods for reducing traces that will be used for performance analysis. We evaluate the different methods against several criteria, including size reduction, introduced error, and retention of performance trends, using both benchmarks with carefully chosen performance behaviors, and a real application.

Space-Efficient Time-Series Call-Path Profiling of Parallel Applications

Authors: Zoltán Szebenyi (Juelich Supercomputing Centre), Felix Wolf (Juelich Supercomputing Centre), Brian J. N. Wylie (Juelich Supercomputing Centre)

The performance behavior of parallel simulations often changes considerably as the simulation progresses with potentially process-dependent variations of temporal patterns. While call-path profiling is an established method of linking a performance problem to the context in which it occurs, call paths reveal only little information about the temporal evolution of performance phenomena. However, generating call-path profiles separately for thousands of iterations may exceed available buffer space --- especially when the call tree is large and more than one metric is collected. In this paper, we present a runtime approach for the semantic compression of call-path profiles based on incremental clustering of a series of single-iteration profiles that scales in terms of the number of iterations without sacrificing important performance details. Our approach offers low runtime overhead by using only a condensed version of the profile data when calculating distances and accounts for process-dependent variations by making all clustering decisions locally.

Virtual Wide-Area Networking

Chair: Kate Keahey (Argonne National Laboratory)

Time: 3:30pm-5pm

Room: PB256

Improving GridFTP Performance Using The Phoebus Session Layer

Authors: Ezra Kissel (University of Delaware), Aaron Brown (University of Delaware), Martin Swany (University of Delaware)

Phoebus is an infrastructure for improving end-to-end throughput in high-bandwidth, long-distance networks by using a "session layer" protocol and "Gateways" in the network. Phoebus has the ability to dynamically allocate resources and to use segment specific transport protocols between Gateways, as well as applying other performance-improving techniques on behalf of the user. One of the key data movement applications in high-performance and Grid computing is GridFTP from the Globus project. GridFTP features a modular library interface called XIO that allows it to use alternative transport mechanisms. To facilitate use of the Phoebus system, we have implemented a Globus XIO driver for Phoebus. This paper presents tests of the Phoebus-enabled GridFTP over a network testbed that allows us to modify latency and loss rates. We discuss use of various transport connections, both end-to-end and hop-by-hop, and evaluate the performance of a variety of cases.

On the Design of Scalable, Self-Configuring Virtual Networks

Authors: David Isaac Wolinsky (University of Florida), Yonggang Liu (University of Florida), Pierre St. Juste (University of Florida), Girish Venkatasubramanian (University of Florida), Renato Figueiredo (University of Florida)

Virtual networks (VNs) enable global addressability of VN-connected machines through either a common Ethernet (LAN) or a NAT-free layer 3 IP network (subnet), providing methods that simplify resource management, deal with connectivity constraints, and support legacy applications in distributed systems. This paper presents a novel VN design that supports dynamic, seamless addition of new resources with emphasis on scalability in a unified private IP address space. Key features of this system are: (1) Scalable connectivity via a P2P overlay with the ability to bypass overlay routing in LAN communications, (2) support for static and dynamic address allocation in conjunction with virtual nameservers through a distributed data store, and (3) support for transparent migration of IP endpoints across wide-area networks. The approach is validated by a prototype implementation which has been deployed in grid and cloud environments. We present both a quantitative and qualitative discussion of our findings.

Wednesday, November 18

Acceleration

Chair: Scott Pakin (Los Alamos National Laboratory)

Time: 1:30pm-3pm

Room: PB255

A 32x32x32, Spatially Distributed 3D FFT in Four Microseconds on Anton

Authors: Cliff Young (D.E. Shaw Research), Joseph A. Bank (D.E. Shaw Research), Ron O. Dror (D.E. Shaw Research), J.P. Grossman (D.E. Shaw Research), John K. Salmon (D.E. Shaw Research), David E. Shaw (D.E. Shaw Research)

Anton, a massively parallel special-purpose machine for molecular dynamics simulations, performs a 32x32x32 FFT in 3.7 microseconds and a 64x64x64 FFT in 13.3 microseconds on a configuration with 512 nodes—an order of magnitude faster than all other FFT implementations of which we are aware. Achieving this FFT performance requires a coordinated combination of computation and communication techniques that leverage Anton's underlying hardware mechanisms. Most significantly, Anton's communication subsystem provides over 300 gigabits per second of bandwidth per node, message latency in

the hundreds of nanoseconds, and support for word-level writes and single-ended communication. In addition, Anton's general-purpose computation system incorporates primitives that support the efficient parallelization of small 1D FFTs. Although Anton was designed specifically for molecular dynamics simulations, a number of the hardware primitives and software implementation techniques described in this paper may also be applicable to the acceleration of FFTs on general-purpose high-performance machines.

SCAMPI: A Scalable Cam-based Algorithm for Multiple Pattern Inspection

Authors: Fabrizio Petrini (IBM T.J. Watson Research Center), Virat Agarwal (IBM T.J. Watson Research Center), Davide Pasetto (IBM Corporation)

In this paper we present SCAMPI, a ground-breaking string searching algorithm that is fast, space-efficient, scalable and resilient to attacks. SCAMPI is designed with a memory-centric model of complexity in mind, to minimize memory traffic and enhance data reuse with a careful compile-time data layout. The experimental evaluation executed on two families of multicore processors, Cell B.E and Intel Xeon E5472, shows that it is possible to obtain a processing rate of more than 2 Gbits/sec per core with very large dictionaries and heavy hitting rates. In the largest tested configuration, SCAMPI reaches 16 Gbits/sec on 8 Xeon cores, reaching, and in some cases exceeding, the performance of special-purpose processors and FPGA. Using SCAMPI we have been able to scan an input stream using a dictionary of 3.5 millions keywords at a rate of more than 1.2 Gbits/sec per processing core.

Grid Scheduling

Chair: Weicheng Huang (National Center for High-Performance Computing Taiwan)

Time: 1:30pm-3pm

Room: PB256

Evaluating the Impact of Inaccurate Information in Utility-Based Scheduling

Authors: Alvin AuYoung (University of California, San Diego), Amin Vahdat (University of California, San Diego), Alex Snoeren (University of California, San Diego)

Proponents of utility-based scheduling policies have shown the potential for a 100-1400% increase in value-delivered to users when used in lieu of traditional approaches such as FCFS, back-fill or priority queues. However, perhaps due to concerns about

their potential fragility, these policies are rarely used in practice. We present an evaluation of a utility-based scheduling policy based upon real workload data from both an auction-based resource infrastructure, and a supercomputing cluster. We model two potential sources of imperfect operating conditions for a utility-based policy: user uncertainty and wealth inequity. Through simulation, we find that in the worst-case, the value delivered by a utility-based policy can degrade to half that of traditional approaches, but that such a policy can provide 20-100% improvement in expected operating conditions. We conclude that future efforts in designing utility-based allocation mechanisms and policies must explicitly consider the fidelity of elicited job value information from users.

Predicting the Execution Time of Grid Workflow Applications through Local Learning

Authors: Farrukh Nadeem (University of Innsbruck), Thomas Fahringer (University of Innsbruck)

Workflow execution time prediction is widely seen as a key service to understand the performance behavior and support the optimization of Grid workflow applications. In this paper, we present a novel approach for estimating the execution time of workflows based on Local Learning. The workflows are characterized in terms of different attributes describing structural and runtime information about workflow activities, control and data flow dependencies, number of Grid sites, problem size, etc. Our local learning framework is complemented by a dynamic weighing scheme that assigns weights to workflow attributes reflecting their impact on the workflow execution time. Predictions are given through intervals bounded by the minimum and maximum predicted values, which are associated with a confidence value indicating the degree of confidence about the prediction accuracy. Evaluation results for three real world workflows on a real Grid are presented to demonstrate the prediction accuracy and overheads of the proposed method.

Supporting Fault-Tolerance for Time-Critical Events in Distributed Environments

Authors: Qian Zhu (Ohio State University), Gagan Agrawal (Ohio State University)

In this paper, we consider the problem of supporting fault tolerance for adaptive and time-critical applications in heterogeneous and unreliable grid computing environments. Our goal for this class of applications is to optimize a user-specified benefit function while meeting the time deadline. Our first contribution in this paper is a multi-objective optimization algorithm for scheduling the application onto the most efficient and reliable

resources. In this way, the processing can achieve the maximum benefit while also maximizing the success-rate, which is the probability of finishing execution without failures. However, when failures do occur, we have developed a hybrid failure-recovery scheme to ensure the application can complete within the time interval. Our experimental results show that our scheduling algorithm can achieve better benefit when compared to several heuristics-based greedy scheduling algorithms with a negligible overhead. Benefit is further improved by the hybrid failure recovery scheme, and the success-rate becomes 100%.

Award: Best Student Paper Finalist

High Performance Filesystems and I/O

Chair: Scott Klasky (Oak Ridge National Laboratory)

Time: 1:30pm-3pm

Room: E145-146

I/O Performance Challenges at Leadership Scale

Authors: Samuel Lang (Argonne National Laboratory), Philip Carns (Argonne National Laboratory), Kevin Harms (Argonne National Laboratory), Robert Latham (Argonne National Laboratory), Robert Ross (Argonne National Laboratory), William Allcock (Argonne National Laboratory)

Today's top high performance computing systems run applications with hundreds of thousands of processes, contain hundreds of storage nodes, and must meet massive I/O requirements for capacity and performance. These leadership-class systems face daunting challenges to deploying scalable I/O systems. In this paper we present a case study of the I/O challenges to performance and scalability on Intrepid, the IBM Blue Gene/P system at the Argonne Leadership Computing Facility. Listed in the top 5 fastest supercomputers of 2008, Intrepid runs computational science applications with intensive demands on the I/O system. We show that Intrepid's file and storage system sustain high performance under varying workloads as the applications scale with the number of processes.

Scalable Massively Parallel I/O to Task-Local Files

Authors: Wolfgang Frings (Juelich Supercomputing Centre), Felix Wolf (Juelich Supercomputing Centre), Ventsislav Petkov (Technical University Munich)

Parallel applications often store data in multiple task-local files, for example, to remember checkpoints, to circumvent memory limitations, or to record performance data. When operating at very large processor configurations, such applications often experience scalability limitations when the simultaneous creation of

thousands of files causes metadata-server contention or simply when large file counts complicate file management or operations on those files even destabilize the file system. SIONlib is a parallel I/O library that addresses this problem by transparently mapping a large number of task-local files onto a small number of physical files via internal metadata handling and block alignment to ensure high performance. While requiring only minimal source code changes, SIONlib significantly reduces file creation overhead and simplifies file handling without penalizing read and write performance. We evaluate SIONlib's efficiency with up to 288 K tasks and report significant performance improvements in two application scenarios.

PLFS: A Checkpoint Filesystem for Parallel Applications

Authors: John Bent (Los Alamos National Laboratory), Garth Gibson (Carnegie Mellon University), Gary Grider (Los Alamos National Laboratory), Ben McClelland (Los Alamos National Laboratory), Paul Nowoczynski (Pittsburgh Supercomputing Center), James Nunez (Los Alamos National Laboratory), Milo Polte (Carnegie Mellon University), Meghan Wingate (Los Alamos National Laboratory)

Parallel applications running across thousands of processors must protect themselves from inevitable system failures. Many applications insulate themselves from failures by checkpointing. For many applications, checkpointing into a shared single file is most convenient. With such an approach, the size of writes are often small and not aligned with file system boundaries. Unfortunately for these applications, this preferred data layout results in pathologically poor performance from the underlying file system which is optimized for large, aligned writes to non-shared files. To address this fundamental mismatch, we have developed a virtual parallel log structured file system, PLFS. PLFS remaps an application's preferred data layout into one which is optimized for the underlying file system. Through testing on PanFS, Lustre, and GPFS, we have seen that this layer of indirection and reorganization can reduce checkpoint time by an order of magnitude for several important benchmarks and real applications.

Award: Best Paper Finalist

GPU Applications

Chair: Esmond G Ng (Lawrence Berkeley National Laboratory)

Time: 3:30pm-5pm

Room: PB256

Multi-core Acceleration of Chemical Kinetics for Simulation and Prediction

Authors: John C. Linford (Virginia Tech), John Michalakes (National Center for Atmospheric Research), Manish Vachharajani (University of Colorado at Boulder), Adrian Sandu (Virginia Tech)

This work implements a computationally expensive chemical kinetics kernel from a large-scale community atmospheric model on three multi-core platforms: NVIDIA GPUs using CUDA, the Cell Broadband Engine, and Intel Quad-Core Xeon CPUs. A comparative performance analysis for each platform in double and single precision on coarse and fine grids is presented. Platform-specific design and optimization is discussed in a mechanism-agnostic way, permitting the optimization of many chemical mechanisms. The implementation of a three-stage Rosenbrock solver for SIMD architectures is discussed. When used as a template mechanism in the Kinetic PreProcessor, the multi-core implementation enables the automatic optimization and porting of many chemical mechanisms on a variety of multi-core platforms. Speedups of 5.5x in single precision and 2.7x in double precision are observed when compared to eight Xeon cores. Compared to the serial implementation, the maximum observed speedup is 41.1x in single precision.

Towards a Framework for Abstracting Accelerators in Parallel Applications: Experience with Cell

Authors: David M. Kunzman (University of Illinois at Urbana-Champaign), Laxmikant V. Kale (University of Illinois at Urbana-Champaign)

While accelerators have become more prevalent in recent years, they are still considered hard to program. In this work, we extend a framework for parallel programming so that programmers can easily take advantage of the Cell processor's Synergistic Processing Elements (SPEs) as seamlessly as possible. Using this framework, the same application code can be compiled and executed on multiple platforms, including x86-based and Cell-based clusters. Furthermore, our model allows independently developed libraries to efficiently time-share one or more SPEs by interleaving work from multiple libraries. To demonstrate the framework, we present performance data for an example molecular dynamics (MD) application. When compared to a single Xeon core utilizing streaming SIMD extensions (SSE), the MD program achieves a

speedup of 5.74 on a single Cell chip (with 8 SPEs). In comparison, a similar speedup of 5.89 is achieved using six Xeon (x86) cores.

Award: Best Student Paper Finalist

A Microdriver Architecture for Error Correcting Codes inside the Linux Kernel

Authors: Dominic Eschweiler (Forschungszentrum Juelich), Andre Brinkmann (University of Paderborn)

Linux is often used in conjunction with parallel file systems in high performance cluster environments and the tremendous storage growth in these environments leads to the requirement of multi-error correcting codes. This work investigated the potential of graphic cards for such coding applications in the Linux kernel. For this purpose, a special micro driver concept (Barracuda) has been designed that can be integrated into Linux without changing kernel APIs. For the investigation of the performance of this concept, the Linux RAID 6-system and the applied Reed-Solomon code have been exemplary extended and studied. The resulting measurements outline opportunities and limitations of our microdriver concept. On the one hand, the concept achieves a speed-up of 72 for complex, 8-failure correcting codes, while no additional speed-up can be generated for simpler, 2-error correcting codes. An example application for Barracuda could therefore be replacement of expensive RAID systems in cluster storage environments.

Networking

Chair: Dennis Abts (Google)

Time: 3:30pm-5pm

Room: E145-146

HyperX: Topology, Routing, and Packaging of Efficient Large-Scale Networks

Authors: Jung Ho Ahn (Hewlett-Packard), Nathan Binkert (Hewlett-Packard), Al Davis (Hewlett-Packard), Moray McLaren (Hewlett-Packard), Robert S. Schreiber (Hewlett-Packard)

In the push to achieve exascale performance, systems will grow to over 100,000 sockets, as growing cores-per-socket and improved single-core performance provide only part of the speedup needed. These systems will need affordable interconnect structures that scale to this level. To meet the need, we consider an extension of the hypercube and flattened butterfly topologies, the HyperX, and give an adaptive routing algorithm, DAL. HyperX takes advantage of high-radix switch components that integrated photonics will make available. Our main contributions include a formal descriptive framework, enabling a search method that finds optimal HyperX configurations; DAL; and a low cost pack-

aging strategy for an exascale HyperX. Simulations show that HyperX can provide performance as good as a folded Clos, with fewer switches. We also describe a HyperX packaging scheme that reduces system cost. Our analysis of efficiency, performance, and packaging demonstrates that the HyperX is a strong competitor for exascale networks.

Router Designs for Elastic Buffer On-Chip Networks

Authors: George Micheliogiannakis (Stanford University), William J. Dally (Stanford University)

This paper explores the design space of elastic buffer (EB) routers by evaluating three representative designs. We propose an enhanced two-stage EB router which maximizes throughput by achieving a 42% reduction in cycle time and 20% reduction in occupied area by using look-ahead routing and replacing the three-slot output EBs in the baseline router of [17] with two-slot EBs. We also propose a single-stage router which merges the two pipeline stages to avoid pipelining overhead. This design reduces zero-load latency by 24% compared to the enhanced two-stage router if both are operated at the same clock frequency; moreover, the single-stage router reduces the required energy per transferred bit and occupied area by 29% and 30% respectively, compared to the enhanced two-stage router. However, the cycle time of the enhanced two-stage router is 26% smaller than that of the single-stage router.

Allocator Implementations for Network-on-Chip Routers

Authors: Daniel U. Becker (Stanford University), William J. Dally (Stanford University)

The present contribution explores the design space for virtual channel (VC) and switch allocators in network-on-chip (NoC) routers. Based on detailed RTL-level implementations, we evaluate representative allocator architectures in terms of matching quality, delay, area and power and investigate the sensitivity of these properties to key network parameters. We introduce a scheme for sparse VC allocation that limits transitions between groups of VCs based on the function they perform, and reduces the VC allocator's delay, area and power by up to 41%, 90% and 83%, respectively. Furthermore, we propose a pessimistic mechanism for speculative switch allocation that reduces switch allocator delay by up to 23% compared to a conventional implementation without increasing the router's zero-load latency. Finally, we quantify the effects of the various design choices discussed in the paper on overall network performance by presenting simulation results for two exemplary 64-node NoC topologies.

Thursday, November 19

Performance Analysis and Optimization

Chair: Kirk W. Cameron (Virginia Tech)

Time: 10:30am-Noon

Room: D135-136

Performance Evaluation of NEC SX-9 using Real Science and Engineering Applications

Authors: Takashi Soga (Tohoku University), Akihiro Musa (NEC Corporation), Youichi Shimomura (NEC Corporation), Koki Okabe (Tohoku University), Ryusuke Egawa (Tohoku University), Hiroyuki Takizawa (Tohoku University), Hiroaki Kobayashi (Tohoku University), Ken'ichi Itakura (Japan Agency for Marine-Earth Science & Technology)

This paper describes a new-generation vector parallel supercomputer, NEC SX-9 system. The SX-9 processor has an outstanding core to achieve over 100Gflop/s, and a software-controllable on-chip cache to keep the high ratio of the memory bandwidth to the floating-point operation rate. Moreover, its large SMP nodes of 16 vector processors with 1.6Tflop/s performance and 1TB memory are connected with dedicated network switches, which can achieve inter-node communication at 128GB/s per direction. The sustained performance of the SX-9 processor is evaluated using six practical applications in comparison with conventional vector processors and the latest scalar processor such

as Nehalem-EP. Based on the results, this paper discusses the performance tuning strategies for new-generation vector systems. An SX-9 system of 16 nodes is also evaluated by using the HPC challenge benchmark suite and a CFD code. Those evaluation results clarify the highest sustained performance and scalability of the SX-9 system.

Early Performance Evaluation of “Nehalem” Cluster using Scientific and Engineering Applications

Authors: Subhash Saini (NASA Ames Research Center), Andrey Naraikin (Intel Corporation), Rupak Biswas (NASA Ames Research Center), David Barkai (Intel Corporation), Timothy Sandstrom (NASA Ames Research Center)

In this paper, we present an early performance evaluation of a 512-core cluster based on Intel third-generation quad-core processor—the Intel® Xeon® 5500 Series (code named Nehalem-EP). This is Intel's first processor with non-uniform memory access (NUMA) architecture managed by on-chip integrated memory controllers. It employs a point-to-point intercon-

nect called Quick Path Interconnect (QPI) between processors and to the input/output (I/O) hub. Two other features of the processor are the introduction of simultaneous multi threading (SMT) to Intel quad-core and “Turbo Mode”—the ability to dynamically increase the clock frequency as long as the power consumed remains within the designed thermal envelope. We critically evaluate all these Nehalem features using the High Performance Computing Challenge (HPCC) benchmarks, NAS Parallel Benchmarks (NPB), and four full-scale scientific applications. We compare and contrast Nehalem results against an SGI Altix ICE 8200EX® platform and an Intel® cluster of previous generation processors.

Machine Learning-Based Prefetch Optimization for Data Center Applications

Authors: Shih-wei Liao (Google), Tzu-Han Hung (Princeton University), Donald Nguyen (University of Texas at Austin), Chinyen Chou (National Taiwan University), Chiaheng Tu (National Taiwan University), Hucheng Zhou (National Taiwan University)

Performance tuning for data centers is essential and complicated. It is important since a data center comprises thousands of machines and thus a single-digit performance improvement can significantly reduce cost and power consumption. Unfortunately, it is extremely difficult as data centers are dynamic environments where applications are frequently released and servers are continually upgraded. In this paper, we study the effectiveness of different processor prefetch configurations, which can greatly influence the performance of memory system and the overall data center.

We observe a wide performance gap when comparing the worst and best configurations, from 1.4% to 75.1%, for 11 important data center applications. We then develop a tuning framework which attempts to predict the optimal configuration based on hardware performance counters. The framework achieves performance within 1% of the best performance of any single configuration for the same set of applications.

Sustainability and Reliability

Chair: Franck Cappello (INRIA)

Time: 10:30am-Noon

Room: PB251

FALCON: A System for Reliable Checkpoint Recovery in Shared Grid Environments

Authors: Tanzima Z Islam (Purdue University), Saurabh Bagchi (Purdue University), Rudolf Eigenmann (Purdue University)

In Fine-Grained Cycle Sharing (FGCS) systems, machine owners voluntarily share their unused CPU cycles with guest jobs, as

long as their performance degradation is tolerable. However, unpredictable evictions of guest jobs lead to fluctuating completion times. Checkpoint-recovery is an attractive mechanism for recovering from such "failures". Today's FGCS systems often use expensive, high-performance dedicated checkpoint servers. However, in geographically distributed clusters, this may incur high checkpoint transfer latencies. In this paper we present a system called FALCON that uses available disk resources of the FGCS machines as shared checkpoint repositories. However, an unavailable storage host may lead to loss of checkpoint data. Therefore, we model failures of storage hosts and develop a prediction algorithm for choosing reliable checkpoint repositories. We experiment with FALCON in the university-wide Condor testbed at Purdue and show improved and consistent performance for guest jobs in the presence of irregular resource availability.

Award: Best Student Paper Finalist

Scalable Temporal Order Analysis for Large Scale Debugging

Authors: Dong H. Ahn (Lawrence Livermore National Laboratory), Bronis R. de Supinski (Lawrence Livermore National Laboratory), Ignacio Laguna (Purdue University), Gregory L. Lee (Lawrence Livermore National Laboratory), Ben Liblit (University of Wisconsin-Madison), Barton P. Miller (University of Wisconsin-Madison), Martin Schulz (Lawrence Livermore National Laboratory)

We present a scalable temporal order analysis technique that supports debugging of large scale applications by classifying MPI tasks based on their logical program execution order. Our approach combines static analysis techniques with dynamic analysis to determine this temporal order scalably. It uses scalable stack trace analysis techniques to guide selection of critical program execution points in anomalous application runs. Our novel temporal ordering engine then leverages this information along with the application's static control structure to apply data flow analysis techniques to determine key application data such as

loop control variables. We then use lightweight techniques to gather the dynamic data that determines the temporal order of the MPI tasks. Our evaluation, which extends the Stack Trace Analysis Tool (STAT), demonstrates that this temporal order analysis technique can isolate bugs in benchmark codes with injected faults as well as a real world hang case with AMG2006.

Optimal Real Number Codes for Fault Tolerant Matrix Operations

Authors: Zizhong Chen (Colorado School of Mines)

It has been demonstrated recently that single fail-stop process failure in ScaLAPACK matrix multiplication can be tolerated without checkpointing. Multiple simultaneous processor failures can be tolerated without checkpointing by encoding matrices using a real-number erasure correcting code. However, the floating-point representation of a real number in today's high performance computer architecture introduces round off errors which can be enlarged and cause the loss of precision of possibly all effective digits during recovery when the number of processors in the system is large. In this paper, we present a class of Reed-Solomon style real-number erasure correcting codes which have optimal numerical stability during recovery. We analytically construct the numerically best erasure correcting codes for 2 erasures and develop an approximation method to computationally construct numerically good codes for 3 or more erasures. Experimental results demonstrate that the proposed codes are numerically much more stable than existing codes.

Metadata Management and Storage Cache Allocation

Chair: Karsten Schwan (Georgia Institute of Technology)

Time: 1:30pm-3pm

Room: PB251

SmartStore: A New Metadata Organization Paradigm with Semantic-Awareness for Next-Generation File Systems

Authors: Yu Hua (Huazhong University of Science and Technology), Hong Jiang (University of Nebraska-Lincoln), Yifeng Zhu (University of Maine), Dan Feng (Huazhong University of Science and Technology), Lei Tian (Huazhong University of Science and Technology / University of Nebraska-Lincoln)

Existing storage systems using hierarchical directory tree do not meet scalability and functionality requirements for exponentially growing datasets and increasingly complex queries in Exabyte-level systems with billions of files. This paper proposes semantic-aware organization, called SmartStore, which exploits metadata

semantics of files to judiciously aggregate correlated files into semantic-aware groups by using information retrieval tools. Decentralized design improves system scalability and reduces query latency for complex queries (range and top-k queries), which is conducive to constructing semantic-aware caching, and conventional filename-based query. SmartStore limits search scope of complex query to a single or a minimal number of semantically related groups and avoids or alleviates brute-force search in entire system. Extensive experiments using real-world traces show that SmartStore improves system scalability and reduces query latency over basic database approaches by one thousand times. To the best of our knowledge, this is the first study implementing complex queries in large-scale file systems.

Adaptive and Scalable Metadata Management to Support A Trillion Files

Authors: Jing Xing (Chinese Academy of Sciences), Jin Xiong (Chinese Academy of Sciences), Ninghui Sun (Chinese Academy of Sciences), Jie Ma (Chinese Academy of Sciences)

How to provide high access performance to a single file system or directory with billions or more files is big challenge for cluster file systems. However, limited by a single directory index organization, exist file systems will be prohibitory slow for billions of files. In this paper, we present a scalable and adaptive metadata management system that aims to maintain trillions of files efficiently by an adaptive two-level directory partitioning based on extendible hashing. Moreover, our system utilizes fine-grained parallel processing within a directory to improve performance of concurrent updates, a multi-level metadata cache management to improve memory utilization, and a dynamic load-balance mechanism based on consistent hashing to improve scalability. Our performance tests on 32 metadata servers show that our system can create more than 74,000 files per second and can fstat more than 270,000 files per second in a single directory with 100 million files.

Dynamic Storage Cache Allocation in Multi-Server Architectures

Authors: Ramya Prabhakar (Pennsylvania State University), Shekhar Srikantaiah (Pennsylvania State University), Christina Patrick (Pennsylvania State University), Mahmut Kandemir (Pennsylvania State University)

We introduce a novel storage cache allocation algorithm, called maxperf, that manages the aggregate cache space in multi-server storage architectures such that the service level objectives (SLOs) of concurrently executing applications are satisfied and any spare cache capacity is proportionately allocated according to the mar-

ginal gains of the applications to maximize performance. We use a combination of Neville's algorithm and a linear programming model in our scheme to discover the required storage cache partition size, on each server, for every application accessing that server. Experimental results show that our algorithm enforces partitions to provide stronger isolation to applications, meets application level SLOs even in the presence of dynamically changing storage cache requirements, and improves I/O latency of individual applications as well as the overall I/O latency significantly compared to two alternate storage cache management schemes, and a state-of-the-art single server storage cache management scheme extended to multi-server architecture.

Multicore Task Scheduling

Chair: Mitsuhiro Sato (University of Tsukuba)

Time: 1:30pm-3pm

Room: PB255

Dynamic Task Scheduling for Linear Algebra Algorithms on Distributed-Memory Multicore Systems

Authors: Fengguang Song (University of Tennessee, Knoxville), Asim YarKhan (University of Tennessee, Knoxville), Jack Dongarra (University of Tennessee, Knoxville)

This paper presents a dynamic task scheduling approach to executing dense linear algebra algorithms on multicore systems (either shared-memory or distributed-memory). We use a task-based library to replace the existing linear algebra subroutines such as PBLAS to transparently provide the same interface and computational function as the ScaLAPACK library. Linear algebra programs are written with the task-based library and executed by a dynamic runtime system. We mainly focus our runtime system design on the metric of performance scalability. We propose a distributed algorithm to solve data dependences without process cooperation. We have implemented the runtime system and applied it to three linear algebra algorithms: Cholesky, LU, and QR factorizations. Our experiments on both shared-memory machines (16, 32 cores) and distributed-memory machines (1024 cores) demonstrate that our runtime system is able to achieve good scalability. Furthermore, we provide analytical analysis to show why the tiled algorithms are scalable and the expected execution time.

PFunc: Modern Task Parallelism for Modern High Performance Computing

Authors: Prabhanjan Kambadur (Indiana University), Anshul Gupta (IBM T.J. Watson Research Center), Amol Ghoting (IBM T.J. Watson Research Center), Haim Avron (Tel-Aviv University), Andrew Lumsdaine (Indiana University)

HPC faces new challenges from paradigm shifts in both hardware and software. The ubiquity of multi-cores, many-cores, and GPGPUs is forcing traditional applications to be parallelized for these architectures. Emerging applications (e.g. in informatics) are placing unique requirements on parallel programming tools that remain unaddressed. Although, of all the available parallel programming models, task parallelism appears to be the most promising in meeting these new challenges, current solutions are inadequate. In this paper, we introduce PFunc, a new library for task parallelism that extends the feature set of current solutions for task parallelism with custom task scheduling, task priorities, task affinities, multiple completion notifications and task groups. These features enable one to naturally parallelize a wide variety of modern HPC applications and to support the SPMD model. We present three case studies: demand-driven DAG execution, frequent pattern mining and sparse iterative solvers to demonstrate the utility of PFunc's new features.

Age Based Scheduling for Asymmetric Multiprocessors

Authors: Nagesh B. Lakshminarayana (Georgia Institute of Technology), Jaekyu Lee (Georgia Institute of Technology), Hyesoon Kim (Georgia Institute of Technology)

Asymmetric Multiprocessors (AMPs) are becoming popular in the current era of multicores due to their power efficiency and potential performance and energy efficiency. However, scheduling of multithreaded applications in AMPs is still a challenge. Scheduling algorithms for AMPs must not only be aware of asymmetry in processor performance, but should also consider characteristics of application threads. In this paper, we propose a new scheduling policy, Age based scheduling, that assigns a thread with a larger remaining execution time to a fast core. Age based scheduling predicts the remaining execution time of threads based on their age, i.e., when the threads were created. These predictions are based on the insight that most threads that are created together tend to have similar execution durations. Using Age-based scheduling, we improve the performance of several multithreaded applications by 13% on average and up to 37% compared to the best of previously proposed mechanisms.

System Performance Evaluation

Chair: Xian-He Sun (Illinois Institute of Technology)

1:30pm-3pm

Room: PB256

Instruction-Level Simulation of a Cluster at Scale

Authors: Edgar Leon (University of New Mexico), Rolf Riesen (Sandia National Laboratories), Arthur B. Maccabe (Oak Ridge National Laboratory), Patrick G. Bridges (University of New Mexico)

Instruction-level simulation is necessary to evaluate new architectures. However, single-node simulation cannot predict the behavior of a parallel application on a supercomputer. We present a scalable simulator that couples a cycle-accurate node simulator with a supercomputer network model. Our simulator executes individual instances of IBM's Mambo PowerPC simulator on hundreds of cores. We integrated a NIC emulator into Mambo and model the network instead of fully simulating it. This decouples the individual node simulators and makes our design scalable. Our simulator runs unmodified parallel message-passing applications on hundreds of nodes. We can change network and detailed node parameters, inject network traffic directly into caches, and use different policies to decide when that is an advantage. This paper describes our simulator in detail, evaluates it, and demonstrates its scalability. We show its suitability for architecture research by evaluating the impact of cache injection on parallel application performance.

Diagnosing Performance Bottlenecks in Emerging Petascale Applications

Authors: Nathan Tallent (Rice University), John Mellor-Crummey (Rice University), Laksono Adhianto (Rice University), Michael Fagan (Rice University), Mark Krentel (Rice University)

Cutting-edge science and engineering applications require petascale computing. It is, however, a significant challenge to use petascale computing platforms effectively. Consequently, there is a critical need for performance tools that enable scientists to understand impediments to performance on emerging petascale systems. In this paper, we describe HPCToolkit—a suite of multi-platform tools that supports sampling-based analysis of application performance on emerging petascale platforms. HPCToolkit uses sampling to pinpoint and quantify both scaling and node performance bottlenecks. We study several emerging petascale applications on the Cray XT and IBM BlueGene/P platforms and use HPCToolkit to identify specific source lines—in their full calling context—associated with performance bottlenecks in these codes. Such information is exactly what application developers need to know to improve their applications to take full advantage of the power of petascale systems.

Comparative Study of One-Sided Factorizations with Multiple Software Packages on Multi-Core Hardware

Authors: Emmanuel Agullo (University of Tennessee, Knoxville), Bilel Hadri (University of Tennessee, Knoxville), Hatem Ltaief (University of Tennessee, Knoxville), Jack Dongarra (University of Tennessee, Knoxville)

The emergence and continuing use of multi-core architectures require changes in the existing software and sometimes even a redesign of the established algorithms in order to take advantage of now prevailing parallelism. The Parallel Linear Algebra for Scalable Multi-core Architectures (PLASMA) is a project that aims to achieve both high performance and portability across a wide range of multi-core architectures. We present in this paper a comparative study of PLASMA's performance against established linear algebra packages (LAPACK and ScaLAPACK), against new approaches at parallel execution (Task Based Linear Algebra Subroutines-TBLAS), and against equivalent commercial software offerings (MKL, ESSL and PESSL). Our experiments were conducted on one-sided linear algebra factorizations (LU, QR and Cholesky) and used multi-core architectures (based on Intel Xeon EMT64 and IBM Power6). The performance results show improvements brought by new algorithms on up to 32 cores-the largest multi-core system we could access.

Dynamic Task Scheduling

Chair: David Abramson (Monash University)

Time: 3:30pm-5pm

Room: E145-146

VGrADS: Enabling e-Science Workflows on Grids and Clouds with Fault Tolerance

Authors: Lavanya Ramakrishnan (Indiana University), Daniel Nurmi (University of California, Santa Barbara), Anirban Mandal (Renaissance Computing Institute), Charles Koelbel (Rice University), Dennis Gannon (Microsoft Research), T. Mark Huang (University of Houston), Yang-Suk Kee (Oracle), Graziano Obertelli (University of California, Santa Barbara), Kiran Thyagaraja (Rice University), Rich Wolski (University of California, Santa Barbara), Asim Yarkhan (University of Tennessee, Knoxville), Dmitrii Zagorodnov (University of California, Santa Barbara)

Today's scientific workflows use distributed heterogeneous resources through diverse grid and cloud interfaces that are often hard to program. In addition, especially for time-sensitive critical applications, predictable quality of service is necessary across these distributed resources. VGrADS' virtual grid execution system (vgES) provides an uniform qualitative resource abstraction over grid and cloud systems. We apply vgES for scheduling a set

of deadline sensitive weather forecasting workflows. Specifically, this paper reports on our experiences with (1) virtualized reservations for batch-queue systems, (2) coordinated usage of TeraGrid (batch queue), Amazon EC2 (cloud), our own clusters (batch queue) and Eucalyptus (cloud) resources, and (3) fault tolerance through automated task replication. The combined effect of these techniques was to enable a new workflow planning method to balance performance, reliability and cost considerations. The results point toward improved resource selection and execution management support for a variety of e-Science applications over grids and cloud systems.

GridBot: Execution of Bags of Tasks in Multiple Grids

Authors: Mark Silberstein (Technion), Artyom Sharov (Technion), Dan Geiger (Technion), Assaf Schuster (Technion)

We present a holistic approach for efficient execution of bags-of-tasks (BOTs) on multiple grids, clusters, and volunteer computing grids virtualized as a single computing platform. The challenge is twofold: to assemble this compound environment and to employ it for execution of a mixture of throughput- and performance-oriented BOTs, with a dozen to millions of tasks each. Our mechanism allows per BOT specification of dynamic arbitrary scheduling and replication policies as a function of the system state, BOT execution state or BOT priority. We implement our mechanism in the GridBot system and demonstrate its capabilities in a production setup. GridBot has executed hundreds of BOTs with total of over 9 million jobs during three months alone; these have been invoked on 25,000 hosts, 15,000 from the Superlink@Technion community grid and the rest from the Technion campus grid, local clusters, the Open Science Grid, EGEE, and the UW Madison pool.

Scalable Work Stealing

Authors: James Dinan (Ohio State University), Sriram Krishnamoorthy (Pacific Northwest National Laboratory), D. Brian Larkins (Ohio State University), Jarek Nieplocha (Pacific Northwest National Laboratory), P. Sadayappan (Ohio State University)

Irregular and dynamic parallel applications pose significant challenges to achieving scalable performance on large-scale multicore clusters. These applications often require ongoing, dynamic load balancing in order to maintain efficiency. Scalable dynamic load balancing on large clusters is a challenging problem which can be addressed with distributed dynamic load balancing systems. Work stealing is a popular approach to distributed dynamic load balancing; however its performance on large-scale clusters is not well understood. Prior work on work stealing has largely focused on shared memory machines. In this work we investigate the

design and scalability of work stealing on modern distributed memory systems. We demonstrate high efficiency and low overhead when scaling to 8,192 processors for three benchmark codes: a producer-consumer benchmark, the unbalanced tree search benchmark, and a multiresolution analysis kernel.

Future HPC Architectures

Chair: Ron Brightwell (Sandia National Laboratories)

Time: 3:30pm-5pm

Room: PB256



Future Scaling of Processor-Memory Interfaces

Authors: Jung Ho Ahn (Hewlett-Packard), Norm P. Jouppi (Hewlett-Packard), Christos Kozyrakis (Stanford University), Jacob Leverich (Stanford University), Robert S. Schreiber (Hewlett-Packard)

Continuous evolution in process technology brings energy-efficiency and reliability challenges, which are harder for memory system designs since chip multiprocessors demand high bandwidth and capacity, global wires improve slowly, and more cells are susceptible to hard and soft errors. Recently, there are proposals aiming at better main-memory energy efficiency by dividing a memory rank into subsets. We holistically assess the effectiveness of rank subsetting in the context of system-wide performance, energy-efficiency, and reliability perspectives. We identify the impact of rank subsetting on memory power and processor performance analytically, then verify the analyses by simulating a chip-multiprocessor system using multithreaded and consolidated workloads. We extend the design of Multicore DIMM, one proposal embodying rank subsetting, for high-reliability systems and show that compared with conventional chipkill approaches, it can lead to much higher system-level energy efficiency and performance at the cost of additional DRAM devices.

A Design Methodology for Domain-Optimized Power-Efficient Supercomputing

Authors: Marghoob Mohiyuddin (University of California, Berkeley / Lawrence Berkeley National Laboratory), Mark Murphy (University of California, Berkeley), Leonid Oliker (Lawrence Berkeley National Laboratory), John Shalf (Lawrence Berkeley National Laboratory), John Wawrzyniec (University of California, Berkeley), Samuel Williams (Lawrence Berkeley National Laboratory)

As power has become the pre-eminent design constraint for future HPC systems, computational efficiency is being emphasized over simply peak performance. Recently, static benchmark codes have been used to find a power efficient architecture.

Unfortunately, because compilers generate sub-optimal code, benchmark performance can be a poor indicator of the performance potential of architecture design points. Therefore, we present hardware/software co-tuning as a novel approach for system design, in which traditional architecture space exploration is tightly coupled with software auto-tuning for delivering substantial improvements in area and power efficiency. We demonstrate the proposed methodology by exploring the parameter space of a Tensilica-based multi-processor running three of the most heavily used kernels in scientific computing, each with widely varying micro-architectural requirements: sparse matrix vector multiplication, stencil-based computations, and general matrix-matrix multiplication. Results demonstrate that co-tuning significantly improves hardware area and energy efficiency -- a key driver for next generation of HPC system design.

Leveraging 3D PCRAM Technologies to Reduce Checkpoint Overhead for Future Exascale Systems

Authors: Xiangyu Dong (Pennsylvania State University), Naveen Muralimanohar (HP Labs), Norm Jouppi (HP Labs), Richard Kaufmann (HP Labs), Yuan Xie (Pennsylvania State University)

The scalability of future massively parallel processing (MPP) systems is challenged by high failure rates. Current hard disk drive (HDD) checkpointing results in overhead of 25% or more at the petascale. With a direct correlation between checkpoint frequencies and node counts, novel techniques that can take more frequent checkpoints with minimum overhead are critical to implement a reliable exascale system. In this work, we leverage the upcoming Phase-Change Random Access Memory (PCRAM) technology and propose a hybrid local/global checkpointing mechanism after a thorough analysis of MPP systems failure rates and failure sources. We propose three variants of PCRAM-based hybrid checkpointing schemes: DIMM+HDD, DIMM+DIMM, and 3D+3D, to reduce the checkpoint overhead and offer a smooth transition from the conventional pure HDD checkpoint to the ideal 3D PCRAM mechanism. The proposed pure 3D PCRAM-based mechanism can ultimately take checkpoints with overhead less than 4% on a projected exascale system.

Computing for a Changing World



Portland,
OR | 2009



A Scalable Domain Decomposition Method for Ultra-Parallel Arterial Flow Simulations

Leopold Grinberg (Brown University), John Cazes (Texas Advanced Computing Center), Greg Foss (Pittsburgh Supercomputing Center), George Karniadakis (Brown University)

High resolution 3D simulations of blood flow in the human arterial tree, which consists of hundreds of vessels, require the use of thousands of processors. Ultra-parallel flow simulations on hundreds of thousands of processors require new multi-level domain decomposition methods. We present a new two-level method that couples discontinuous and continuous Galerkin formulations. At the coarse level the domain is subdivided into several big overlapping patches and within each patch a spectral/hp element discretization is employed. New interface conditions for the Navier-Stokes equations are developed to connect the patches, relaxing the C^0 continuity and minimizing data transfer at the patch interface. A Multilevel Communicating Interface (MCI) has been developed to enhance the parallel efficiency. Excellent weak scaling in solving billions degrees of freedom problems on up to 32,768 cores was observed on BG/P and XT5. Results and details of the simulation will be presented in the poster.

Quantum Monte Carlo Simulation of Materials using GPUs

Kenneth Esler (University of Illinois at Urbana-Champaign), Jeongnim Kim (University of Illinois at Urbana-Champaign), David Ceperley (University of Illinois at Urbana-Champaign)

Continuum quantum Monte Carlo (QMC) has proved to be an invaluable tool for predicting the properties of matter from fundamental principles. By solving the many-body Schroedinger equation through a stochastic projection, it achieves greater accuracy than DFT methods and much better scalability than quantum chemical methods, enabling scientific discovery across a broad spectrum of disciplines. The multiple forms of parallelism afforded by QMC algorithms make it an ideal candidate for acceleration in the many-core paradigm. We present the results of our effort to port the QMCPACK simulation code to the NVIDIA CUDA platform. We restructure the CPU algorithms to express additional parallelism, minimize GPU-CPU communication, and efficiently utilize the GPU memory hierarchy. Using mixed precision on G200 GPUs and MPI for intercommunication, we observe typical full-application speedups of approximately 10x to 15x relative to quad-core Xeon CPUs alone, while reproducing the double-precision CPU results within statistical error.

HPC Application Fault-Tolerance using Transparent Redundant Computation

Kurt B Ferreira (Sandia National Laboratories), Rolf Riesen (Sandia National Laboratories), Ron A. Oldfield (Sandia National Laboratories), Ron Brightwell (Sandia National Laboratories), James H. Laros (Sandia National Laboratories), Kevin T. Pedretti (Sandia National Laboratories)

As the core count of HPC machines continue to grow in size, issues such as fault tolerance and reliability are becoming limiting factors for application scalability. Current techniques to ensure progress across faults, for example coordinated checkpoint-restart, are unsuitable for machines of this scale due to their predicted high overheads. In this study, we present the design and implementation of a novel system for ensuring reliability which uses transparent, rank-level, redundant computation. Using this system, we show the overheads involved in redundant computation for a number of real-world HPC applications. Additionally, we relate the communication characteristics of an application to the overheads observed.



Visualization System for the Evaluation of Numerical Simulation Results

Hiroko Nakamura Miyamura (Japan Atomic Energy Agency), Sachiko Hayashi (Japan Atomic Energy Agency), Yoshio Suzuki (Japan Atomic Energy Agency), Hiroshi Takemiya (Japan Atomic Energy Agency)

With the improvement in the performance of supercomputers, numerical simulations have become larger and more complex, which has made the interpretation of simulation results more difficult. Moreover, occasionally, users cannot evaluate numerical simulation results, even though they have spent a great deal of time, because interactive visualization is impossible for such large-scale data. Therefore, we herein propose a visualization system with which users evaluate large-scale time-series data that has been obtained in a parallel and distributed environment. The proposed system allows users to simultaneously visualize and analyze data in both the global and local scales.



ClassCloud: Building Energy-Efficient Experimental Cloud Infrastructure using DRBL and Xen

Yao-Tsung Wang (National Center for High-Performance Computing Taiwan), Che-Yuan Tu (National Center for High-Performance Computing Taiwan), Wen-Chieh Kuo (National Center for High-Performance Computing Taiwan), Wei-Yu Chen (National Center for High-Performance Computing Taiwan), Steven Shiau (National Center for High-Performance Computing Taiwan)

Green Computing is a growing research topic in recent years. Its goal is to increase energy efficiency and reduce resource consumption. Building infrastructure for Cloud Computing is one of the methods to achieve these goals. The key concept of Cloud Computing is to provide a resource-sharing model based on virtualization, distributed filesystem and web services. In this poster, we will propose an energy efficient architecture for Cloud Computing using Diskless Remote Boot in Linux (DRBL) and Xen. We call this architecture ClassCloud because it is suitable to build an experimental cloud infrastructure in a PC classroom. According to our power consumption experiments, the diskless design of DRBL provides a way to implement a power economization computing platform. Computers booting from network without hard disk installed will reduce power consumption from 8% to 26% compared to computers with hard disk installed.



Eigenvalue Solution on the Heterogeneous Multicore Cluster for Nuclear Fusion Reactor Monitoring

Noriyuki Kushida (Japan Atomic Energy Agency), Hiroshi Takemiya (Japan Atomic Energy Agency), Shinji Tokuda (Japan Atomic Energy Agency)

In this study, we developed a high speed eigenvalue solver that is the necessity of plasma stability analysis system for International Thermo-nuclear Experimental Reactor (ITER) on Cell cluster system. Our stability analysis system is developed in order to prevent the damage of the ITER from the plasma disruption. Our stability analysis system requires solving the eigenvalue of matrices whose dimension is hundred thousand in a few seconds. However, current massively parallel processor (MPP) type supercomputer is not applicable for such short-term calculation, because of the network overhead become dominant. Therefore, we employ Cell cluster system, because we can obtain sufficient processing power with small number of processors. Furthermore, we developed novel eigenvalue solver with the consideration of hierarchical parallelism of Cell cluster. Finally, we succeeded to solve the block tri-diagonal Hermitian matrix, which had 1024 diagonal blocks and each block size was 128x128, within one second.



Massively Parallel Simulations of Large DNA Molecules Using the IBM Blue Gene Supercomputer

Amanda Peters (Harvard University), Greg Lakatos (Harvard University), Maria Fyta (Technical University Munich), Efthimios Kaxiras (Harvard University)

The conformational variability of DNA remains a problem of critical importance, especially given recent experimental studies of DNA translocation through nanopores and DNA interaction with nanostructures such as carbon nanotubes. Naturally occurring DNA molecules are large, containing potentially billions of individual chemical units. We have developed a coarse-grained molecular level description of DNA that captures both the structural and dynamical properties. Using a domain-decomposition approach, we implemented an efficient, parallel molecular dynamics code enabling the simulation of DNA molecules ~10K base pairs long. We evaluated the performance using a simulated dataset on a BlueGene/L. Our results show a 97% overall time reduction and 98% efficiency at 512 processors. This work represents a significant advance in the development multi-scale simulation tools for large DNA molecules. These tools will provide insights into the behavior of DNA, both in living cells, and in a new generation of novel DNA sequencing and sorting devices.

Automatic Live Job Migration for HPC Fault Tolerance with Virtualization

Yi Lun Pan (National Center for High-Performance Computing Taiwan)

The reliability of large-scale parallel jobs within a cluster or even across multi-clusters under the Grid environment is a long term issue due to its difficulties involving the monitoring and managing of a large number of compute nodes. To contribute to the issue, an Automatic Live Job Migration (ALJM) middleware with fault tolerance feature has been developed by the Distributed Computing Team in the National Center for High-performance Computing (NCHC). The proposed approach relies on the virtualization techniques exemplified by the OpenVZ [1], which is an open source implementation of virtualization. The approach provides automatically and transparently the fault tolerance capability to the parallel HPC applications. An extremely lightweight approach helps the acquisition of local scheduler services to provide checkpoint/restart mechanism. The approach leverages virtualization techniques combined with cluster queuing system and load balance migration mechanism.

Real-Time Tsunami Simulation on a Multi-Node GPU Cluster

Marlon Rodolfo Arce Acuna (Tokyo Institute of Technology), Takayuki Aoki (Tokyo Institute of Technology)

Tsunamis are destructive forces of nature and thus their accurate forecast and early warning is extremely important. In order to predict a Tsunami, the Shallow Water Equations must be solved in real-time. With the introduction of GPGPU, a new revolution has been opened in high performance computing. We used CUDA to compute the simulation on the GPU which drastically accelerated the computation. A single-GPU calculation had been found to be 62-times faster than using a single CPU core. Moreover the domain was decomposed and solved on a multi-node GPU cluster. Overlapping transfers and computation further accelerated the process by hiding communication. For the GPU transfers an asynchronous copy model was used. The MPI library was used to transfer the data between nodes. A domain representing real bathymetry, with grid size 4096x8192 was used as our dataset. Our tests on TSUBAME showed excellent scalability.

Simulation Caching for Interactive Remote Steering Framework

Shunsaku Tezuka (University of Fukui), Kensuke Hashimoto (University of Fukui), Shinji Fukuma (University of Fukui), Shin-ichiro Mori (University of Fukui), Shinji Tomita (Kyoto University)

We have been investigating the human-in-the-loop scientific computing environment where the scientific simulation running on a high performance computing server (a remote server) located far away from the operator is interactively visualized and steered beyond the internet. In order to realize such an environment, not only the remote server should have a sufficient computing throughput, but also there should have some mechanism to hide the communication latency between the remote server and the operation terminal. For this purpose, we propose a simulation model which we call "Simulation Caching." The Simulation Caching 1) utilizes a moderate scale computing server (a local server), associate with local operation terminal, to perform a sort of simulation, low resolution simulation for example, to make an immediate and reasonable response to the operator's intervention, and 2) keeps the accuracy of the cached simulation by weakly cooperating with the original simulation running on the remote server.



Multicore Computers Can Protect Your Bones: Rendering the Computerized Diagnosis of Osteoporosis a Routine Clinical Practice

Costas Bekas (IBM Research), Alessandro Curioni (IBM Research), Ralph Mueller (ETH Zürich)

Coupling recent imaging capabilities with microstructural finite element analysis offers a powerful tool to determine bone stiffness and strength. It shows high potential to improve individual fracture risk prediction, a tool much needed in the diagnosis and treatment of Osteoporosis that is, according to the World Health Organization, second only to cardiovascular disease as a leading health care problem. We present a high performance computational framework that aims to render the computerized diagnosis of Osteoporosis an everyday, routine clinical practice. In our setting the CAT scanner is fitted with a high performance multicore computing system. Images from the scanner are fed into the computational model and several load conditions are simulated in a matter of minutes. Low time to solution is achieved by means of a novel mixed precision iterative refinement linear solver together with low memory consumption preconditioners. The result is an accurate strength profile of the bone.



GPU Accelerated Electrophysiology Simulations

Fred V. Lionetti (University of California, San Diego), Andrew D. McCulloch (University of California, San Diego), Scott B. Baden (University of California, San Diego)

Numerical simulations of cellular membranes are useful for basic science and increasingly for clinical diagnostic and therapeutic applications. A common bottleneck arises from solving large stiff systems of ODEs at thousands of integration points in a three-dimensional whole organ model. When performing a modern electrophysiology simulation on a single conventional core, the ODE bottleneck consumed 99.4% of the running time. Using an nVidia GTX 295 GPU, we reduced the time to 2%, eliminating the bottleneck. This speedup comes at a small loss in accuracy, due to the use of single precision. By comparison, a multithreaded implementation running on the CPU yielded a speedup of 3.3 only. We are currently investigating the new bottleneck—a PDE solver. We plan to present results at the conference demonstrating the benefits of hybrid execution, whereby the PDE solver runs as multiple threads on the front end CPU. This choice balances development costs against performance.

GPUs: TeraFLOPS or TeraFLAWED?

Imran S. Haque (Stanford University), Vijay S. Pande (Stanford University)

A lack of error checking and correcting (ECC) capability in the memory subsystems of graphics cards has been cited as hindering acceptance of GPUs as high-performance coprocessors, but no quantification has been done to assess the impact of this design. In this poster we present MemtestG80, our software for assessing memory error rates on NVIDIA G80-architecture-based GPUs. Furthermore, we present the results of a large-scale assessment of GPU error rate, conducted by running MemtestG80 on over 20,000 hosts on the Folding@home distributed computing network. Our control experiments on consumer-grade and dedicated-GPGPU hardware in a controlled environment found no errors. However, our survey over consumer-grade cards on Folding@home finds that, in their installed environments, a majority of tested GPUs exhibit a non-negligible, pattern-sensitive rate of memory soft errors. We demonstrate that these errors persist even after controlling for overclocking or environmental proxies for temperature, but depend strongly on board architecture.



Whole Genome Resequencing Analysis in the Clouds

Michael C. Schatz (University of Maryland), Ben Langmead (Johns Hopkins University), Jimmy Lin (University of Maryland), Mihai Pop (University of Maryland), Steven L. Salzberg (University of Maryland)

Biological researchers have a critical need for highly efficient methods for analyzing vast quantities of DNA resequencing data. For example, the 1000 Genomes Project aims to characterize the variations within 1000 human genomes by aligning and analyzing billions of short DNA sequences from each individual. Each genome will utilize ~100GB of compressed sequence data and require ~400 hours computation. Crossbow is our new high-throughput pipeline for whole genome resequencing analysis. It combines Bowtie, an ultrafast and memory efficient short read aligner, and SoapSNP, a highly accurate Bayesian genotyping algorithm, within the distributed processing framework Hadoop to accelerate the computation using many compute nodes. Our results show the pipeline is extremely efficient, and can accurately analyze an entire genome in one day on a small 10-node local cluster, or in one afternoon and for less than \$250 in the Amazon EC2 cloud. Crossbow is available open-source at <http://bowtie-bio.sourceforge.net/crossbow>.

Performance Analysis and Optimization of Parallel I/O in a Large Scale Groundwater Application on the Cray XT5

Vamsi Sripathi (North Carolina State University), Glenn E. Hammond (Pacific Northwest National Laboratory), G. (Kumar) Mahinthakumar (North Carolina State University), Richard T. Mills (Oak Ridge National Laboratory), Patrick H. Worley (Oak Ridge National Laboratory), Peter C. Lichtner (Los Alamos National Laboratory)

We describe in this poster the performance analysis and optimization of I/O within a massively parallel groundwater application, PFLOTRAN, on the Cray XT5 at ORNL. A strong scaling study with a 270 million cell test problem from 2,048 to 65,536 cores indicated that a high volume of independent I/O disk access requests and file access operations would severely limit the I/O performance scalability. To avoid the performance penalty at higher processor counts, we implemented a two-phase I/O approach at the application level by splitting the MPI global communicator into multiple sub-communicators. The root process in each sub-communicator is responsible for performing the I/O operations for the entire group and then distributing the data to rest of the group. With this approach we were able to achieve 25X improvement in read I/O and 3X improvement in write I/O resulting in an overall application performance improvement of over 5X at 65,536 cores.



Simulation of Biochemical Pathways using High Performance Heterogeneous Computing System

Michael Maher (Brunel University), Simon Kent (Brunel University), Xuan Liu (Brunel University), David Gilbert (Brunel University), Dairis Latimer (Petapath), Neil Hickey (Petapath)

Biochemical systems are responsible for processing signals, inducing the appropriate cellular responses and sequence of internal events. However these systems are not fully or even poorly understood. At the core of biochemical systems research, is computational modelling using Ordinary Differential Equations, to create abstract models of biochemical systems for subsequent analysis. Due to the complexity of these models it is often necessary to use high-performance computing to keep the experimental time within tractable bounds. We plan to use the Petapath prototype e740 accelerator card, to parallelize various aspects of the computational analysis. This approach can greatly reduce the running time of an ODEs-driven biochemical simulator, e.g. BioNessie. Using a highly effective SIMD parallel processing architecture, this approach has the additional benefit of being incredibly power efficient, consuming only one watt per 3.84 GFLOPS of double precision computation. This reduces not only running costs but also has a positive environmental impact.

Teramem: A Remote Swapping System for High-Performance Computing

Kazunori Yamamoto (University of Tokyo), Yutaka Ishikawa (University of Tokyo)

Some application programs, e.g. ones in bioinformatics and natural language processing, need huge memory spaces. However, most supercomputers only support several tens of gigabytes of physical memory per compute node. It is almost impossible to use more memory than installed due to the lack of disk swap space or the severe performance degradation caused by disk swapping. We introduce Teramem, a remote swapping system for providing large virtual memory spaces on 64-bit commodity architectures. Teramem has mainly the following two advantages: (1) Its kernel-level implementation enables Teramem to use efficient page replacement algorithms by utilizing memory management information, and (2) Its design is independent of Linux swap mechanism and optimized for remote memory transfer. The evaluation shows that it achieves 603 MB/s sequential read bandwidth, and the GNU sort benchmark runs more than 40 times as fast as using disk swapping.

A Resilient Runtime Environment for HPC and Internet Core Router Systems

Ralph Castain (Cisco Systems), Joshua Hursey (Indiana University), Timothy I. Mattox (Cisco Systems), Chase Cotton (University of Delaware), Robert M. Broberg (Cisco Systems), Jonathan M. Smith (University of Pennsylvania)

Core routers, with aggregate I/O capabilities now approaching 100 terabits/second, are closely analogous to modern HPC systems (i.e., highly parallel with various types of processor interconnects). Maintaining or improving availability while continuing to scale demands integration of resiliency techniques into the supporting runtime environments (RTEs). Open MPI's Runtime Environment (ORTE) is a modular open source RTE implementation which we have enhanced to provide resilience to both HPC and core router applications. These enhancements include proactive process migration and automatic process recovery services for applications, including unmodified MPI applications. We describe the distributed failure detection, prediction, notification, and recovery components required for resilient operations. During recovery, the fault topology aware remapping of processes on the machine (based on the Fault Group model) reduces the impact of cascading failures on applications. We present preliminary results and plans for future extensions.

Scalability of Quantum Chemistry Codes on BlueGene/P and Challenges for Sustained Petascale Performance

Jeff R. Hammond (Argonne National Laboratory)

The path to sustained petascale performance of quantum chemistry codes is frequently discussed but rarely analyzed in detail. We present performance data of four popular quantum chemistry codes on BlueGene/P - NWChem, GAMESS, MPQC and Dalton - which cover a diverse collection of scientific capabilities. Unlike evaluations which rely only on performance within a single code, our analysis considers both the underlying algorithms as well as the quality of their implementation, both of which vary significantly between codes. The analysis will be broken down into four components: (1) single-node performance of atomic integral codes, (2) use of memory hierarchies in tensor contractions, (3) communication patterns and scope of synchronization and (4) the trade-offs between IO and on-the-fly computation. The results reveal that benign neglect has introduced significant hurdles towards achieving sustained petascale performance but suggest ways to overcome them using techniques such as auto-tuning and hierarchical parallelism.



Highly Scalable Bioinformatics Tools on Supercomputers

Bhanu Rekapalli (University of Tennessee, Knoxville), Christian Halloy (University of Tennessee, Knoxville), Igor B. Jouline (University of Tennessee, Knoxville)

Only a few thousand genomes have been sequenced since 1998, even though there are millions of known organisms on the planet. It is expected that new sequencing technologies will be able to sequence nearly 1000 microbial genomes per hour. This mind-boggling growth of genomic data will require vastly improved computational methods for the pre-processing, analysis, storage and retrieval of sequence data. We have made significant headway in this direction by porting some of the most popular bioinformatics tools such as BLAST and HMMER on supercomputers. Our Highly Scalable Parallel versions of these tools scale very well up to thousands of cores on a CrayXT5 supercomputers (UTK-ORNL). We discuss how we resolved performance issues by cleverly restructuring the input data sets of protein sequences, randomizing and reducing the I/O bottlenecks. For example, HSP-HMMER code can identify the functional domains of millions of proteins 100 times faster than the publicly available MPI-HMMER.

The Texas Test Problem Server

Victor Eijkhout (University of Texas at Austin), James Kneeland (University of Texas at Austin)

The Texas Test Problem Server is a resource for algorithm researchers in computational science, in particular in numerical linear algebra. In contrast to several static collections of test matrices, the TxTPS hosts a number of applications that generate computationally interesting problems. A user can run these applications through a web interface, and receive a download link to the problem generated. This approach has the advantage that the user can obtain test problems that come from related test problems, through varying PDE parameters, or from the same test problem with multiple discretizations. Additional features of the TxTPS are that test problems can be generated as assembled or unassembled matrices, and as mass matrix / stiffness matrix pairs. For each test problem, extensive numerical properties are computed, and the database of stored problem can be searched by these. We present the interface to the server, and an overview of currently available applications.

Corral: A Resource Provisioning System for Scientific Workflows on the Grid

Gideon Juve (Information Sciences Institute), Ewa Deelman (Information Sciences Institute), Karan Vahi (Information Sciences Institute), Gaurang Mehta (Information Sciences Institute)

The development of grid and workflow technologies has enabled complex, loosely coupled scientific applications to be executed on distributed resources. Many of these applications consist of large numbers of short-duration tasks whose execution times are heavily influenced by scheduling policies and overheads. Such applications often perform poorly on the grid because grid sites implement scheduling policies that favor tightly-coupled applications and grid middleware imposes large scheduling overheads and delays. We will present a provisioning system that enables workflow applications to allocate resources from grid sites using conventional tools, and control the scheduling of tasks on those resources. We will demonstrate how the system has been used to significantly improve the runtime of several scientific workflow applications.

Exploring Options in Mobile Supercomputing for Image Processing Applications

Fernando Ortiz (EM Photonics), Carmen Carrano (Lawrence Livermore National Laboratory)

Digital video enhancement is a necessity within many applications in law enforcement, the military and space exploration. As sensor technology advances increasing amounts of digital data is being collected. In addition, processing algorithms increase in complexity to counter a growing list of effects such as platform instabilities, image noise, atmospheric blurring and other perturbations. Current desktop computers, however, are incapable of applying these algorithms to high-definition video at real-time speeds. In this poster, we compare traditional supercomputing approaches with digital image processing in novel portable platforms like FPGAs and GPUs. The supercomputer runs were performed at Lawrence Livermore National Laboratory, and the GPU/FPGA runs at EM Photonics, Inc. We demonstrate that the same programming paradigms can be shared among supercomputers and commodity platforms, with the latter having the added benefit of lower power consumption and limited portability.

Knowledge Discovery from Sensor Data for Environmental Information

Frank Kendall (Kean University), Brian Sinnicke (Kean University), Ryan Suleski (Kean University), Patricia Morreale (Kean University)

Street Corners is a wireless sensor network application which supports the contextual presentation of data gathered from an urban setting. The Street Corners application offers real-time data display and provides support for predictive algorithms suitable for anticipating, detecting and defending urban communities, among others, from environmental threats such as declining air quality and urban flash floods. The network design and deployment of Street Corners, using Crossbow and Sun SPOT sensors, is presented. Data patterns are extracted from the very large datasets developed from wireless sensor data and identified using data mining on wireless sensor streams and filtering algorithms. The goal of this research is to visually present locations identified as being at risk for environmental harm in near real-time, so that appropriate preventive measures can be taken.

Automating the Knowledge Discovery Process for Computational Quality of Service Research with Scientific Workflow Systems

Meng-Shiou Wu (Ames Laboratory), Colin Campbell (University of Tennessee, Knoxville)

Computational Quality of Service (CQoS) research requires knowledge of performance behaviors acquired from running scientific applications on a variety of platforms in order to develop tuning mechanisms. However, the complexity of current high performance architectures and large-scale scientific packages, the granularity of the 'right' performance data required combined with the numerous options of compilers and external libraries all hamper the discovery of performance knowledge. Without automation of the knowledge-discovering process, CQoS researchers will spend most of their time in the laborious task of data management and analysis. In this research we present our experiences and efforts in automating the process of knowledge discovery for CQoS research. We will discuss the fundamental modules in this process, present our prototype infrastructure, and show how to integrate scientific workflow systems into our infrastructure to automate the complex process of knowledge discovery.

Seamlessly Scaling a Desktop MATLAB Application to an Experimental TeraGrid Resource

Zoya Dimitrova (Centers for Disease Control and Prevention), David Campo (Centers for Disease Control and Prevention), Elizabeth Neuhaus (Centers for Disease Control and Prevention), Yuri Khudyakov (Centers for Disease Control and Prevention), David Lifka (Cornell Center for Advanced Computing (CAC), Nate Woody (Cornell Center for Advanced Computing (CAC))

MATLAB(R) was used to study networks of coordinated amino acid variation in Hepatitis C virus (HCV), a major cause of liver disease worldwide. Mapping of coordinated variations in the viral polyprotein has revealed a small collection of amino acid sites that significantly impacts Hepatitis viral evolution. Knowledge of these sites and their interactions may help devise novel molecular strategies for disrupting Hepatitis viral functions and may also be used to find new therapeutic targets for HCV. Statistical verification of HCV coordinated mutation networks requires generation of thousands of random amino acid alignments, a computationally intensive process that greatly benefits from parallelization. Cornell recently deployed an experimental TeraGrid computing resource with 512 dedicated cores running MATLAB Distributed Computing Server that may be accessed by any TeraGrid user running MATLAB and the Parallel Computing Toolbox on their desktop. This resource enabled a great reduction in HCV simulation times.

Multicore-Optimized Temporal Blocking for Stencil Computations

Markus Wittmann (Erlangen Regional Computing Center), Georg Hager (Erlangen Regional Computing Center)

The "DRAM Gap," i.e. the discrepancy between theoretical peak performance and memory bandwidth of a processor chip, keeps worsening with the advent of multi- and many-core processors. New algorithms and optimization techniques must be developed to diminish the applications' hunger for memory bandwidth. Stencil codes, which are frequently used at the core of fluid flow simulations (like Lattice-Boltzmann solvers) and PDE solvers, including multi-grid methods, can break new performance grounds by improving temporal locality. We use a Jacobi solver in two and three dimensions to analyze two variants of temporal blocking, and show how to leverage shared caches in multi-core environments. Additionally we extend our approach to a hybrid (OpenMP+MPI) programming model suitable for clusters of shared-memory nodes. Performance results are presented for a Xeon-based InfiniBand cluster and a 4-socket Intel "Dunnington" node.

LEMMing: Cluster Management through Web Services and a Rich Internet Application

Jonas Dias (COPPE-UFRJ), Albino Aveleda (COPPE-UFRJ)

The spread of high performance computing systems to process large amounts of data and the growth of the number of processors per cluster may lead to an administration task scalability problem on these systems. We present LEMMing, a tool developed to support high performance computing centers. It concentrates the cluster administration tasks on a single rich Internet application. Web services deployed on each machine provide an interface to the system working as an abstraction layer of its administration procedures. On the other side, a web interface accesses these services and uses these operations remotely. This user interface is also designed to organize a huge number of computing nodes smartly and provide great usability. With LEMMing, the administration tasks could be more pragmatic and failures would be corrected faster even in heterogeneous environments with huge machines.

MPWide: A Communication Library for Distributed Supercomputing

Derek Groen (Leiden University), Steven Rieder (Leiden University), Paola Grosso (University of Amsterdam), Simon Portegies Zwart (Leiden University), Cees de Laat (University of Amsterdam)

MPWide is a light-weight communication library which connects two applications, each of them running with the locally recommended MPI implementation. We originally developed it to manage the long-distance message passing in the CosmoGrid project, where cosmological N-body simulations run on grids of supercomputers connected by high performance optical networks. To take full advantage of the network light paths in CosmoGrid, we need a message passing library that supports the ability to use customized communication settings (e.g. custom number of streams, window sizes) for individual network links among the sites. It supports a de-centralized startup, required for CosmoGrid as it is not possible to start the simulation on all supercomputers from one site. We intend to show a scaled-down, live run of the CosmoGrid simulation in the poster session.

Virtual Microscopy and Analysis using Scientific Workflows

David Abramson (Monash University), Colin Enticott (Monash University), Stephen Firth (Monash University), Slavisa Garic (Monash University), Ian Harper (Monash University), Martin Lackmann (Monash University), Minh Ngoc Dinh (Monash University), Hoang Nguyen (Monash University), Tirath Ramdas (University of Zurich), A.B.M. Russel (University of Melbourne), Stefan Schek (Leica Microsystems), Mary Vail (Monash University), Blair Bethwaite (Monash University)

Most commercial microscopes are stand-alone instruments, controlled by dedicated computer systems. These provide limited storage and processing capabilities. Virtual microscopes, on the other hand, link the image capturing hardware and data analysis software into a wide area network of high performance computers, large storage devices and software systems. In this paper we discuss extensions to Grid workflow engines that allow them to execute scientific experiments on virtual microscopes. We demonstrate the utility of such a system in a biomedical case study concerning the imaging of cancer and antibody based therapeutics.

Many-Task Computing (MTC) Challenges and Solutions

Scott Callaghan (University of Southern California), Philip Maechling (University of Southern California), Ewa Deelman (Information Sciences Institute), Patrick Small (University of Southern California), Kevin Milner (University of Southern California), Thomas H. Jordan (University of Southern California)

Many-task computing (MTC) may be heterogeneous, including many jobs and many files, and we believe MTC computing is a widely applicable computational model for scientific computing. However, an effective MTC calculation must address many computational challenges, such as execution of heterogeneous MPI and loosely coupled tasks, resource provisioning, distributed execution, task management, data management, automation, status monitoring, publishing of results, and comparable performance metrics. In this poster we present a discussion of these issues and describe the solutions we used to perform a large-scale MTC calculation, the Southern California Earthquake Center's (SCEC) CyberShake 1.0 Map calculation. This CyberShake run used 7 million CPU hours over 54 days, executed 190 million tasks, produced 175 TB of data, and produced scientifically meaningful results, using software tools such as Pegasus, Condor, Corral, and NetLogger. Additionally, we will explore potential standardized performance metrics for comparing MTC calculations and present measurements for CyberShake.

Enhancing the Automatic Generation of Fused Linear Algebra Kernels

Erik Silkenen (University of Colorado at Boulder), Ian Karlin (University of Colorado at Boulder), Geoff Belter (University of Colorado at Boulder), Elizabeth Jessup (University of Colorado at Boulder), Jeremy Siek (University of Colorado at Boulder)

The performance of scientific applications is often limited by the cost of memory accesses inside linear algebra kernels. We developed a compiler that tunes such kernels for memory efficiency, showing significant speedups relative to vendor-tuned BLAS libraries. Our compiler accepts annotated MATLAB code and outputs C++ code with varying levels of loop fusion. In this poster, we present an analysis of sequences of memory bound matrix-vector multiplies that suggest ways to improve the compiler. In particular, this analysis will aid development of the compiler's memory model, enabling improvements that reduce the number of loop fusion operations considered. This is important because the number of possible routines quickly increases with kernel complexity and exhaustive testing becomes expensive. We also present hardware performance counter data, which shows we need to consider register allocation in the memory model, and helps identify when other optimizations such as cache blocking and array interleaving become profitable.

Mirroring Earth System Grid Datasets

Erin L. Brady (Information Sciences Institute), Ann Chervenak (Information Sciences Institute)

The Earth System Grid (ESG) provides secure access to climate simulation data located at storage sites across the U. S. With the growing importance of the Intergovernmental Panel on Climate Change (IPCC) data sets hosted in ESG, several sites around the world have expressed interest in hosting a replica or mirror of a subset of IPCC data. The goal of these mirror sites is to provide reliable access to these datasets for local scientists and to reduce wide area data access latencies. Our project provides the design and implementation of this data mirroring functionality, which integrates several components of the federated ESG architecture currently under development. Currently, our mirror tool can query for metadata about the datasets, retrieve their physical addresses from the metadata, and create a transfer request for a data movement client. Our poster will describe the design and implementation of the ESG data mirroring tool.

Scientific Computing on GPUs using Jacket for MATLAB

James Malcolm (AccelerEyes), Gallagher Pryor (AccelerEyes), Tauseef Rehman (AccelerEyes), John Melonakos (AccelerEyes)

In this poster, we present new features of Jacket: The GPU Engine for MATLAB that enable scientific computing on GPUs. These include enhanced support for data and task parallel algorithms using 'gfor' and multi-GPU programming constructs as well as sparse support for BLAS functions including point-wise matrix arithmetic and matrix-vector multiplications. Jacket is not another GPU API nor is it another collection of GPU functions. Rather, it is simply an extension of the MATLAB language to new GPU data types, 'gsingle' and 'gdouble'. Jacket brings the speed and visual computing capability of the GPU to MATLAB programs by providing transparent overloading of MATLAB's CPU-based functions with CUDA-based functions. Jacket includes automated and optimized memory transfers and kernel configurations and uses a compile on-the-fly system that allows GPU functions to run in MATLAB's interpretive style. The poster and demos will present Jacket enhancements that are important for the scientific computing community.

A Parallel Power Flow Solver based on the Gauss-Seidel method on the IBM Cell/B.E.

Jong-Ho Byun (University of North Carolina at Charlotte), Kushal Datta (University of North Carolina at Charlotte), Arun Ravindran (University of North Carolina at Charlotte), Arindam Mukherjee (University of North Carolina at Charlotte), Bharat Joshi (University of North Carolina at Charlotte), David Chassin (Pacific Northwest National Laboratory)

In this paper, we report a parallel implementation of Power Flow Solver using Gauss-Seidel (GS) method on heterogeneous multi-core IBM Cell Broadband Engine (Cell/B.E.). GS-based power flow solver is part of the transmission module of the GridLAB-D power distribution simulator and analysis tool. Our implementation is based on PPE-centric Parallel Stages programming model, where large dataset is partitioned and simultaneously processed in the SPE computing stages. The core of our implementation is a vectorized unified-bus-computation module which employs three techniques â (1) computation reordering, (2) eliminating misprediction, (3) integration of four different bus computations using SIMDized vector intrinsic of the SPE and (4) I/O double-buffering which overlaps computation and DMA data transfers. As a result, we achieve 15 times speedup compared to sequential implementation of the power flow solver algorithm. In addition, we analyze scalability and the effect of SIMDized vectorization and double-buffering on application performance.

High-Performance Non-blocking Collective Communications in MPI

Akihiro Nomura (University of Tokyo), Hiroya Matsuba (University of Tokyo), Yutaka Ishikawa (University of Tokyo)

The introduction of non-blocking collective communications into the MPI standard has been discussed in the MPI forum. Collective communications are normally constructed using set of point-to-point communications. These point-to-point communications must be issued not at once but at appropriate timing, because there are some dependencies among them. We call these process progression. In blocking-style collective communications, we can process progression in MPI library's context. However, in non-blocking collective communications, progression cannot be handled in this context. Communication threads are generally used to solve this problem, but making new communication thread has critical disadvantages in performance. In this poster, we introduce a method to process progression in interrupt context instead of process context in operating system. The dependencies among procedures in progression are modeled and stored into the memory which is accessible from interrupt context. In this method, we no longer use communication threads and the performance will be improved.

High Resolution Numerical Approach to Turbulent Liquid-Fuel Spray Atomization using the Fujitsu FX1 Multicore Based Massively Parallel Cluster

Junji Shinjo (JAXA), Ryoji Takaki (JAXA), Yuichi Matsuo (JAXA), Setsuo Honma (Fujitsu), Yasutoshi Hashimoto (Fujitsu Kyushu Systems)

Liquid fuel spray is widely used in aerospace and automotive engines and combustion performance such as energy efficiency or exhaust gas cleanness is dependent on spray performance. However, the physical mechanism of liquid atomization has not been understood well because the phenomenon is highly multi-scale and turbulent. A direct numerical simulation of liquid fuel jet with 6 billion grid points is conducted to elucidate its physics. Fujitsu's 120TFLOPS FX1 cluster newly installed at Japan Aerospace Exploration Agency (JAXA) is utilized for this simulation using 1440 CPUs (5760 cores) for 450 hours/CPU. The obtained result indicates that the physical processes of surface instability development, ligament (liquid thread) formation and droplet formation could be directly captured by this simulation. The information about the local flow field and droplet distribution is critical to designing an actual spray system. The effectiveness of massive numerical simulation in combustion research has been demonstrated by this study.

MOON: MapReduce On Opportunistic eNvironments

Heshan Lin (Virginia Tech), Jeremy Archuleta (Virginia Tech), Xiaosong Ma (North Carolina State University / Oak Ridge National Laboratory), Wuchun Feng (Virginia Tech)

Existing MapReduce systems run on dedicated resources, where only a small fraction of such resources are ever unavailable. In contrast, we tackle the challenges of hosting MapReduce services on volunteer computing systems, which opportunistically harness idle desktop computers, e.g., Condor. Such systems, when running MapReduce, result in poor performance due to the volatility of resources. Specifically, the data and task replication scheme adopted by existing MapReduce implementations is woefully inadequate for resources with high unavailability. To address this, we extend Hadoop with adaptive task and data scheduling algorithms to offer reliable MapReduce services in volunteer computing systems. Leveraging a hybrid resource architecture, our algorithms distinguish between (1) different types of MapReduce data and (2) different types of node outages in order to strategically place tasks and data on both volatile and dedicated nodes. Our tests demonstrate that MOON can deliver a 3-fold performance improvement to Hadoop in volatile environments.

Autotuning and Specialization: Speeding Up Nek5000 with Compiler Technology

Jaewook Shin (Argonne National Laboratory), Mary W. Hall (University of Utah), Jacqueline Chame (Information Sciences Institute), Chun Chen (University of Utah), Paul F. Fischer (Argonne National Laboratory), Paul D. Hovland (Argonne National Laboratory)

Autotuning has emerged as a systematic process for evaluating alternative implementations of a computation to select the best-performing solution for a particular architecture. Specialization customizes code to a particular input data set. This poster presents a compiler approach that combines novel autotuning technology with specialization for expected data sizes of key computations. We apply this approach to nek5000, a spectral element code. Nek5000 makes heavy use of matrix multiply for small matrices. We describe compiler techniques developed, including the interface to a polyhedral transformation system for generating specialized code and the heuristics used to prune the search space of alternative implementations. We demonstrate more than 2.3X performance gains on an Opteron over the original implementation, and significantly better performance than even the hand-coded Goto, ATLAS and ACML BLAS. The entire application using the tuned code demonstrates 36%, 8% and 10% performance improvements on single, 32 and 64 processors, respectively.

FTB-Enabled Failure Prediction for Blue Gene/P Systems

Ziming Zheng (Illinois Institute of Technology), Rinku Gupta (Argonne National Laboratory), Zhiling Lan (Illinois Institute of Technology), Susan Coghlan (Argonne National Laboratory)

On large-scale systems, many software provide some degree of fault tolerance. However, most of them tend to handle faults independently and agnostically from other software. In contrast, the Coordinated Infrastructure for Fault Tolerant Systems (CIFTS) initiative takes a holistic approach for providing system-wide fault tolerance. The Fault Tolerance Backplane (FTB) serves as a backbone for CIFTS and provides a shared infrastructure for all levels of system software to coordinate fault information through a uniform interface. In this poster, we present a FTB-enabled failure prediction software for Blue Gene/P systems. It dynamically generates failure rules for failure prediction by actively monitoring RAS events in the central CMCS databases on Blue Gene/P. The prediction published through FTB can be subscribed by other FTB-enabled software such as job schedulers, checkpointing software, etc., for proactive actions. We also present preliminary results of our software on the Intrepid system at Argonne National Laboratory.

ALPS: A Framework for Parallel Adaptive PDE Solution

Carsten Burstedde (University of Texas at Austin), Omar Ghattas (University of Texas at Austin), Georg Stadler (University of Texas at Austin), Tiankai Tu (University of Texas at Austin), Lucas C Wilcox (University of Texas at Austin)

Adaptive mesh refinement and coarsening (AMR) is essential for the numerical solution of partial differential equations (PDEs) that exhibit behavior over a wide range of length and time scales. Because of the complex dynamic data structures and communication patterns and frequent data exchange and redistribution, scaling dynamic AMR to tens of thousands of processors has long been considered a challenge. We are developing ALPS, a library for dynamic mesh adaptation of PDEs that is designed to scale to hundreds of thousands of compute cores. Our approach uses parallel forest-of-octree-based hexahedral finite element meshes and dynamic load balancing based on space-filling curves. ALPS supports arbitrary-order accurate continuous and discontinuous finite element/spectral element discretizations on general geometries. We present scalability and performance results for two applications from geophysics: mantle convection and seismic wave propagation.

Cellule: Lightweight Execution Environment for Virtualized Accelerators

Vishakha Gupta (Georgia Institute of Technology), Jimi Xenidis (IBM Austin Research Laboratory), Priyanka Tembey (Georgia Institute of Technology), Ada Gavrilovska (Georgia Institute of Technology), Karsten Schwan (Georgia Institute of Technology)

The increasing prevalence of accelerators is changing the HPC landscape to one in which future platforms will consist of heterogeneous multi-core chips comprised of general purpose and specialized cores. Coupled with this trend is increased support for virtualization, which can abstract underlying hardware to aid in dynamically managing its use by HPC applications. Virtualization can also help hide architectural differences across cores, by offering specialized, efficient, and self-contained execution environments (SEE) for accelerator codes. Cellule uses virtualization to create a high performance, low noise SEE for the Cell processor. Cellule attains these goals with compute-intensive workloads, demonstrating improved performance compared to the current Linux-based runtime. Furthermore, a key principle, coordinated resource management for accelerator and general purpose resources, is shown to extend beyond Cell, using experimental results attained on another accelerator.

Thalweg: A System For Developing, Deploying, and Sharing Data-Parallel Libraries

Adam L. Beberg (Stanford University), Vijay S. Pande (Stanford University)

Developing and deploying applications that retain a high degree of portability yet can take advantage of the underlying architecture poses many real world problems. From detecting the proper hardware and verifying its functioning correctly at run time, to development issues when optimizing for different low level APIs each with different setup, memory management, and invocation schemes. Thalweg addresses this series of issues with an API for functions and portable runtime library for simplifying development and deployment of applications that can retain their high level portability while taking full advantage of any low level optimizations for the hardware detected at run time, such as SSE/Altivec, GPUs, Cell, and SMP. All code is compiled to dynamic libraries, which can then be used from languages like C, Python, Java, or Mathematica. OpenSIMD.com is for sharing code/libraries written for Thalweg. The poster will give an overview of the Thalweg architecture and its benefits.

Irregularly Portioned Lagrangian Monte-Carlo for Turbulent Reacting Flow Simulations

S. Levent Yilmaz (University of Pittsburgh), Patrick Piscuneri (University of Pittsburgh), Mehdi B. Nik (University of Pittsburgh)

High fidelity, feasible simulation methodologies are indispensable to modern gas-turbine design, and here we take on a novel methodology, Filtered Density Function (FDF) for large eddy simulation of turbulent reacting flow. FDF is a robust methodology which can provide very accurate predictions for a wide range of flow conditions. However, it involves an expensive particle/mesh algorithm where stiff chemical reaction computations cause quite interesting, problem specific, and in most cases extremely imbalanced (a couple of orders of magnitude) computational load distribution. While FDF is established as an indispensable tool in fundamental combustion research, these computational issues prevent it to get its deserved place at level of the industrial applications. We introduce an advanced implementation which combines robust parallelization libraries, such as Zoltan, and other optimized solvers (ODEPACK), with a flexible parallelization strategy to tackle the load-imbalance barrier. We demonstrate scalability with application to a large scale, high Reynolds number combustor.

ACM Student Research Competition

Tuesday, November 17

Time: 5:15-7pm

Room: Oregon Ballroom Lobby



On the Efficacy of Haskell for High-Performance Computational Biology

Author: Jacqueline R. Addesa (Virginia Tech)

While Haskell is an advanced functional programming language that is increasingly being used for commercial applications, e.g., web services, it is rarely considered for high-performance computing despite its ability to express algorithms succinctly. As such, we compare the computational and expressive power of Haskell to a more traditional imperative language, namely C, in the context of multiple sequence alignment, an NP-hard problem in computational biology. Although the C implementation mirrored Haskell's, the C implementation did not account for problems with the run-time stack, such as stack overflow. Once addressed, the C code ballooned to over 1000 lines of code, more than 37 times longer than the Haskell implementation. Not only is the Haskell implementation more succinct, but also its execution time on large genetic sequences was 2.68 times better than C, as there is less bookkeeping overhead in the Haskell code.

A Policy Based Data Placement Service

Author: Muhammad Ali Amer (University of Southern California)

Large Scientific collaborations or virtual organizations (VOs) often disseminate data sets based on VO-level policies. Such policies include ones that simply replicate data for fault tolerance to more complicated ones that include tiered dissemination of data. We present the design and implementation of a policy based data placement service (PDPS) that makes data placement decisions based on VO level policies and enforces them by initiating a set of data transfer jobs. Our service is built on top of an open-source policy engine (Drools). We present results for enforcing a tiered distribution policy over the wide area using the Planet-Lab testbed. In our tests, large data sets are generated at one experimental site. A large subset of this data set is disseminated to each of four sites in Tier-1, and smaller subsets are distributed to eight Tier-2 sites. Results show that the PDPS successfully enforces policy with low overall overhead.

Hiding Communication and Tuning Scientific Applications using Graph-Based Execution

Author: Pietro Cicotti (University of California, San Diego)

We have developed the Tarragon library to facilitate the development of latency tolerant algorithms in order to hide ever increasing communication delays. Tarragon supports a data-flow execution model, and its run-time system automatically schedules tasks, obviating the need for complicated split-phase coding that doesn't always yield an optimal schedule. In this poster we demonstrate the benefits of Tarragon's data-flow model and its dynamic scheduling. We discuss the tradeoffs of employing over-decomposition to enable communication overlap, as well as a novel feature for tuning performance, performance meta-data. We present experimental results on a stencil computation, matrix multiplication, sparse LU factorization, and DNA sequence alignment. We compare our results on thousands of processors to state-of-the-art MPI encodings on a Cray XT-4 (AMD Opteron 1356 Budapest), a Cray XT-5 (AMD Opteron 2356 Barcelona), an Intel 64 Cluster (Intel Xeon E5345 Clovertown), and a Sun Constellation Linux Cluster (AMD Opteron 8356 Barcelona).

An Automated Air Temperature Analysis and Prediction System for the Blue Gene/P

Author: Neal R. Conrad (Argonne National Laboratory)

As HPC systems grow larger and consume more power, questions of environmental impact and energy costs become more critical. One way to reduce power consumption is to decrease machine room cooling. However, this course of action will reduce the margin of safety for reacting to failures in the cooling system. In this work, we present "The Script of Hope", which is progress toward a reliable method of detecting temperature emergencies and shutting down machine components accordingly. In addition to the computation of backward difference quotients, more advanced mathematical models involving adaptive nonlinear regressions, along with other time series analysis techniques are used to determine the severity of a failing air handler component.

CUSUMMA: Scalable Matrix-Matrix Multiplication on GPUs with CUDA

Author: Byron V. Galbraith (Marquette University)

Simulating neural population signal decoding becomes computationally intractable as the population size increases due to large-scale matrix-matrix multiplications. While an MPI-based version of the simulation overcomes this problem, it introduces other complications (e.g. communication overhead, cluster availability).

We implemented a single-machine, GPU-accelerated version in order to maintain parallel efficiency while reducing communication and resource costs. However, at higher neuron counts the available GPU memory is often insufficient to accommodate the matrices involved, thus precluding the use of a simple CUBLAS SGEMM function call. To circumvent this barrier, we created CUSUMMA (CUDA SUMMA), a self-tuning, GPU-based version of the SUMMA algorithm that determines optimal submatrix partitions to pass to CUBLAS SGEMM based on available device memory. Using a GeForce GTX 260, CUSUMMA achieved 56x speedup vs. BLAS SGEMM when multiplying a 1.6GB neural signal matrix. By publicly releasing CUSUMMA, we hope to enable and enhance a wide spectrum of research applications.

BGPStatusView Overview

Authors: Matthew H. Kemp (Argonne National Laboratory)

Abstract: BGPStatusView is a client/server dynamic web application for IBM BlueGene/P machines running the Cobalt job scheduler. BGPStatusView is strongly influenced by the functionality of LLView, a legacy tcl/tk application designed to monitor job scheduling for the IBM BlueGene/L, but re-architected for the web. In addition, it will be made more accessible by harnessing standards compliant web technologies on the client side to drive the user interface. BGPStatusView is designed for multiple devices, including cell phones, computers and high definition TVs, with an emphasis on large format devices for optimum clarity. The end goal of this project is to enable quick, at-a-glance snapshots depicting the current state of a BlueGene/P machine as well as providing up to the minute information for administrators so as to reveal any complications that may arise during scheduling.

Communication Optimizations of SCEC CME AWP-Olsen Application for Petascale Computing

Author: Kwangyoon Lee (University of California, San Diego)

The Southern California Earthquake Center AWP-Olsen code has been considered to be one of the most extensive earthquake applications in supercomputing community. The code is highly scalable up to 128k cores and portable across wide variety of platforms. However, as the problem size grows, more processors are required to effectively parallelize workloads. The scalability of the code is highly limited by the communication overhead in the NUMA-based systems such as TACC Ranger and NICS Kraken which exhibit highly unbalanced memory access latency based on the location of the processors. The asynchronous communication model effectively removes interdependence between nodes

which does not show any temporal dependence. Consequently, the efficiency of utilizing system communication bandwidth is significantly improved by efficiently parallelizing all the communication paths between cores. Furthermore, this optimized code performed large scale scientific earthquake simulation and reduced the total simulation time to 1/3 of the time of the synchronous code.

Title: IO Optimizations of SCEC AWP-Olsen Application for Petascale Earthquake Computing

Author: Kwangyoon Lee (University of California, San Diego)

The Southern California Earthquake Center (SCEC) AWP-Olsen code has been considered to be one of the most extensive earthquake applications in supercomputing community. Currently, the code has been used to run large-scale earthquake simulations, for example, the SCEC ShakeOut-D wave propagation simulations using 14.4 billion mesh points with 100-m spacing on a regular grid for southern California region, and benchmarked 50-m resolution simulations with challenging 115.2 billion mesh points. Consequently, along with the computational challenges in these scales, IO emerges the most critical component to make the code highly scalable. A new MPIIO scheme in preserving high IO performance and incorporating data redistribution enhances the IO performance by providing contiguous data with optimal size, and effectively redistributing sub-data through highly efficient asynchronous point-to-point communication with minimal travel time. This scheme has been used for SCEC large-scale data-intensive seismic simulations, with reduced IO initialization time to 1/7 for SCEC ShakeOut-D run.

A Hierarchical Approach for Scalability Enhancement in Distributed Network Simulations

Author: Sowmya Myneni (New Mexico State University)

In spite of more than two decades of research in distributed simulation, developers of network simulators have been facing major challenges to incorporate parallelism due to complex dependencies inherent in the network environment. We present experiments involving a hierarchical simulation approach that is reminiscent of the simpler functional simulation at the macroscopic level, but invokes transactional network simulations as detail as any popular commercial or open-source network simulation platform at a microscopic level. Dependencies that would otherwise cause synchronization or rollback overheads among parallel threads are more easily eliminated by reflecting the spatio-temporal behavior of the physical system at a functional level. An additional advantage is the smooth migration path in the simulation

based design cycle, as demonstrated by our gradual shift from ordinary library blocks to popular “ns” family or OMNet++ simulators enclosed in S-functions, through the modified TrueTime block sets originally designed for simulating control systems.



Large-Scale Wavefront Parallelization on Multiple Cores for Sequence Alignment

Author: Shannon I. Steinfadt (Kent State University)

The goal is to investigate the performance possible from multiple cores for the Smith-Waterman sequence alignment algorithm with extremely large alignments, on the order of a quarter of a million characters for both sequences. Whereas other optimization techniques have focused on throughput and optimization for a single core or single accelerator with Cell processors and GPGPUs, we push the boundaries of what can be aligned with Smith-Waterman while returning an alignment, not just the maximum score. By using the streaming SIMD extensions (SSE) on a single core, POSIX Threads (Pthreads) to communicate between multiple cores, and the JumboMem software package to use multiple machines' memory, we can perform an extremely large pairwise alignment that is beyond the capacity of a single machine. The initial results show good scaling and a promising scalable performance as more cores and cluster nodes are tested.

A Feature Reduction Scheme for Obtaining Cost-Effective High-Accuracy Classifiers for Linear Solver Selection

Author: Brice A. Toth (Pennsylvania State University)

Performance of linear solvers depends significantly on the underlying linear system. It is almost impossible to predict apriori the most suitable solver for a given large sparse linear system. Therefore, researchers are exploring the use of supervised classification for solver selection. The classifier is trained on a dataset consisting of sparse matrix features, linear solvers and a solution criteria. However, the cost of computing features is expensive and varies from being proportional to the matrix size, to being as expensive as solving the linear system. Moreover, not all features contribute equally to the classification. In this poster, we present a selection scheme that generates cost-effective feature sets which produces classifiers of high accuracy. We filter out low information features and order the remaining features to decrease the total computation cost. Our results prove that features selected through our strategy produces highly accurate classifiers with low computational costs.

Parallelization of Tau-Leaping Coarse-Grained Monte Carlo Method for Efficient and Accurate Simulations on GPUs

Author: Lifan Xu (University of Delaware)

Recent efforts have outlined the potential of GPUs for Monte Carlo scientific applications. In this poster, we contribute to this effort by exploring the GPU potential for the tau-leaping Coarse-Grained Monte Carlo (CGMC) method. CGMC simulations are important tools for studying phenomena such as catalysis and crystal growth. Existing parallelizations of other CGMC method do not support very large molecular system simulations. Our new algorithm on GPUs provides scientists with a much faster way to study very large molecular systems (faster than on traditional HPC clusters) with the same accuracy. The efficient parallelization of the tau-leaping method for GPUs includes the redesign of both the algorithm and its data structures to address the significantly different GPU memory organization and the GPU multi-thread programming paradigm. The poster describes the parallelization process and the algorithm performance, scalability, and accuracy. To our knowledge, this is the first implementation of this algorithm for GPUs.



Computing for a Changing World

SC09 Masterworks

Masterworks consists of invited presentations that highlight innovative ways of applying high performance computing, networking and storage technologies to the world's most challenging problems. This year a star-studded lineup is in store for SC09 attendees.

Masterworks

Tuesday, November 17

Session: Future Energy Enabled by HPC

10:30am-Noon

Room: PB253-254

Session Chair: Brent Gorda (Lawrence Livermore National Laboratory)



HPC and the Challenge of Achieving a Twenty-fold Increase in Wind Energy

Steve M. Legensky (Intelligent Light)



Wind power generation has made tremendous strides over the past twenty years. Progress in turbine geometries, electrical machinery and gearboxes have enabled machine size to grow from 50kW to 5mW. Site selection has also advanced, permitting wind farms to be

located where strong, reliable winds can be expected. However, wind farms today are underperforming on predicted cost of energy by an average of 10% and operating expenses remain high. The government's Advanced Energy Initiative set a goal of meeting 20% of US electricity needs by 2030, a twenty-fold scale up from today's capability. Meeting this goal requires that performance problems are solved, requiring new tools to design machines and to locate them. Analysis and system-level optimization of the machines, wind farm location and configuration, coupled with accurate meso-micro scale weather modeling will need to be developed and validated. This unsteady, turbulent, multi-scale modeling will only be possible through the use of large scale HPC resources.

Biography: Steve M. Legensky is the founder and general manager of Intelligent Light, a company that has delivered products and services based on visualization technology since 1984. He attended Stevens Institute of Technology (Hoboken, NJ) and received a BE in electrical engineering (1977) and an MS in mathematics (1979). While at Stevens, he helped to establish and then manage the NSF-funded Undergraduate Computer Graphics Facility, an innovative effort that incorporated interactive computer graphics into the engineering curriculum. Legensky then entered industry, working in real-time image generation for flight simulation. In 1984, he founded Intelligent Light, which has evolved from producing award-winning 3D animations to launching a set of 3D rendering and animation solutions, to the current FieldView™ product line of post-processing and visualization tools for computational fluid dynamics (CFD). Steve is an Associate Fellow of the American Institute of Aeronautics and Astronautics (AIAA) and has published and presented for AIAA, IEEE, ACM/SIGGRAPH and IDC.



The Outlook for Energy: Enabled with Supercomputing

John Kuzan (ExxonMobil)



The presentation reviews ExxonMobil's global energy outlook through 2030. The projections indicate that, at that time, the world's population will be ~8 billion, roughly 25% higher than today. Along with this population rise will be continuing economic growth. This combination of population and economic growth will increase energy demand by over 50% versus 2000. As demand rises, the pace of technology improvement is likely to accelerate, reflecting the development and deployment of new technologies for obtaining energy---to include finding and producing oil and natural gas. Effective technology solutions to the energy challenges before us will naturally rely on modeling complicated processes and that in turn will lead to a strong need for supercomputing. Two examples of the supercomputing need in the oil business, seismic approaches for finding petroleum and petroleum reservoir fluid-flow modeling (also known as "reservoir simulation"), will be discussed in the presentation.

Biography: John began his career with ExxonMobil in 1990 as a reservoir engineer working at the Upstream Research Center in special core analysis, which means making fluid mechanics measurements on rock samples from petroleum reservoirs. He has held a variety of positions within ExxonMobil that include leading one of the teams that developed ExxonMobil's next-generation reservoir simulator -- known as EMpower -- and supervising various reservoir engineering sections. John also served as the Transition Manager for ExxonMobil's partnership with the Abu Dhabi National Oil Company in Zakum field.

John is currently the Research Manager for reservoir modeling and has a portfolio that includes merging geoscience and engineering approaches to modeling.

John is a Chemical Engineer by training and his PhD is from the University of Illinois, where his primary work was in turbulent fluid mechanics. He retired from the U.S. Army in 2002 with 17 years in the Reserve and 4 years of active duty. He held Company and Battalion command positions and spent a significant period at the Ballistic Research Laboratory working in supersonic flow and rocket dynamics.

Session: Data Challenges in Genome Analysis

1:30pm-3pm

Room: PB253-254

Session Chair: Susan Gregurick (DOE)



Big Data and Biology: The Implications of Petascale Science

Deepak Singh (Amazon)



The past fifteen years have seen a rapid change in how we practice biology. High-throughput instruments and assays that give us access to new kinds of data, and allow us to probe for answers to a variety of questions, are revealing new insights into biological systems and a variety of diseases. These instruments also pump out very large volumes of data at very high throughputs making biology an increasingly data intensive science, fundamentally challenging our traditional approaches to storing, managing, sharing and analyzing data while maintaining a meaningful biological context. This talk will discuss some of the core needs and challenges of big data as genome centers and individual labs churn out data at ever increasing rates. We will also discuss how we can leverage new paradigms and trends in distributed computing infrastructure and utility models that allow us to manage and analyze big data at scale.

Biography: Deepak Singh is a business development manager at Amazon Web Services (AWS) where he works with customers interested in carrying out large scale computing, scientific research, and data analytics on Amazon EC2, which provides resizable compute capacity in the Amazon cloud. He is also actively involved in Public Data Sets on AWS, a program that provides analysts and researchers easy access to public data sets and computational resources. Prior to AWS, Deepak was a strategist at Rosetta Biosoftware, a business unit of Rosetta Inpharmatics, a subsidiary of Merck & Co. Deepak came to Rosetta Biosoftware from Accelrys where he was first the product manager for the life science simulation portfolio and subsequently Director of the Accelrys NanoBiology Initiative, an effort to investigate multiscale modeling technologies to model biological applications of nanosystems. He started his career as a scientific programmer at GeneFormatics, where he was responsible for the maintenance and implementation of algorithms for protein structure and function prediction. He has a Ph.D. in Physical Chemistry from Syracuse University, where he used electronic structure theory and molecular dynamics simulations to study the structure and photophysics of retinal opsins.



The Supercomputing Challenge to Decode the Evolution and Diversity of Our Genomes

David Haussler (University of California, Santa Cruz)



Extraordinary advances in DNA sequencing technologies are producing improvements at a growth rate faster than the Moore's law exponential. The amount of data is likely to outstrip our ability to share and process it with standard computing infrastructure soon. By comparing genomes of present-day species we can computationally reconstruct most of the genome of the common ancestor of placental mammals from 100 million years ago. We can then deduce the genetic changes on the evolutionary path from that ancient species to humans and discover how natural selection shaped us at the molecular level. About 5% of our genome has remained surprisingly unchanged across millions of years, suggesting important function. There are also short segments that have undergone unusually rapid change in humans, such as a gene linked to brain development. It will be many years before we fully understand the biology of such examples but we relish the opportunity to peek at the molecular tinkering that transformed our animal ancestors into humans.

Biography: David Haussler develops new statistical and algorithmic methods to explore the molecular evolution of the human genome, integrating cross-species comparative and high-throughput genomics data to study gene structure, function, and regulation. He focuses on computational analysis and classification of DNA, RNA, and protein sequences.

He leads the Genome Bioinformatics Group, which participates in the public consortium efforts to produce, assemble, and annotate the first mammalian genomes. His group designed and built the program that assembled the first working draft of the human genome sequence from information produced by sequencing centers worldwide and participated in the informatics associated with the finishing effort.

Session: Finite Elements and Your Body

3:30pm-5pm

Room: PB253-254

Session Chair: Mark Adams (Columbia University)



μ -Finite Element Analysis of Human Bone Structures

Peter Arbenz (ETH, Zurich)



The investigation of the mechanical properties of trabecular bone presents a major challenge due to its high porosity and complex architecture, both of

which vary substantially between anatomic sites and across individuals. A promising technique that takes bone microarchitecture into account is microstructural finite element (μ -FE) analysis. Hundreds of millions of finite elements are needed to accurately represent a human bone with its intricate microarchitecture; hence, the resulting μ -FE models possess a very large number of degrees of freedom, sometimes more than a billion. Detailed μ -FE models are obtained through high-resolution micro-computed tomography of trabecular bone specimens allowing nondestructive imaging of the trabecular microstructure with resolutions on the order of 80 micrometer in living patients. The discrete formulation is based on a standard voxel discretization for linear and nonlinear elasticity. We present highly adapted solvers that efficiently deal with the huge systems of equations on Cray XT5 as well as IBM Blue Gene.

Biography: Peter Arbenz has been a professor at the Institute of Computational Science / Institut für Computational Science at ETH, the Swiss Federal Institute of Technology, in Zurich since 2003. He has a Msc. in Mathematics and a PhD in Applied Mathematics from University of Zurich. From 1983 to 1987 he was a software engineer with BBC Brown, Boveri & Cie (now ABB) in Switzerland and in 1988 joined ETHZ as a senior researcher. His research interests are in Numerical Linear Algebra, High Performance Computing, Parallel and Distributed Computing and Computational Science & Engineering. He is a co-author of the book *Introduction to Parallel Computing-A practical guide with examples in C* (Oxford University Press, 2004).



Virtual Humans: Computer Models for Vehicle Crash Safety

Jesse Ruan (Ford Motor Company)



With computational power growing exponentially, virtual human body models have become routine tools in scientific research and product development. The Ford Motor Company full human body Finite Element (FE) model is one of a few such virtual humans for vehicle crash safety research and injury analysis. This model simulates a 50th percentile male from head to toe, including skull, brain, a detailed skeleton, internal organs, soft tissues, flesh, and muscles. It has shown great promise in helping to understand vehicle impact injury problems and could help reduce dependence on cadaveric test studies. Human body finite element models can also be used to extrapolate results from experimental cadaver investigations to better understand injury mechanisms, validate injury tolerance levels, and to establish injury criteria. Furthermore, it is currently believed that assessing the effectiveness of injury mitigation technologies (such as safety

belts, airbags, and pretensioners, etc.) would be done more efficiently with finite element models of the human body.

Biography: Jesse Ruan graduated from Wayne State University in 1994 with a Ph. D. in Biomedical Engineering. He has worked at Ford Motor Company for 19 years and published more than 60 scientific research papers in peer reviewed journal and conferences. He has been the speaker of numerous international conferences and forums in CAE simulation, impact biomechanics, and vehicular safety. He has completed the development of the Ford Full Human Body Finite Element Model. Recently, he was elected as a Fellow of American Institute for Medical and Biological Engineering.

Jesse is the founding Editor-in-Chief of International Journal of Vehicle Safety. He is a member of American Society of Mechanical Engineers (ASME) and Society of Automotive Engineers (SAE). In addition to his employment with Ford Motor Company, he is a Visiting Professor of Tianjin University of Science and Technology and South China University of Technology in China.

Wednesday, November 18

Session: HPC in Modern Medicine

1:30pm-3pm

Room: PB253-254

Session Chair: Peg Folta (Lawrence Livermore National Laboratory)



Grid Technology Transforming Healthcare

Jonathan Silverstein (University of Chicago)



Healthcare in the United States is a complex adaptive system. Individual rewards and aspirations drive behavior as each stakeholder interacts, self-organizes, learns, reacts, and adapts to one another. Having such a system for health is not inherently bad if incentives are appropriately aligned. However, clinical care, public health, education and research practices have evolved into such a state that we are failing to systematically deliver measurable quality at acceptable cost. From a systems level perspective, integration, interoperability, and secured access to biomedical data on a national scale and its creative re-use to drive better understanding and decision-making are promising paths to transformation of healthcare from a broken system into a high performing one. This session will survey HealthGrid issues and projects across clinical care, public health, education and research with particular focus on transformative efforts enabled by high performance computing.

Biography: Jonathan C. Silverstein, associate director of the Computation Institute of the University of Chicago and Argonne National Laboratory is associate professor of Surgery, Radiology, and The College, scientific director of the Chicago Biomedical Consortium, and president of the HealthGrid.US Alliance. He focuses on the integration of advanced computing and communication technologies into biomedicine, particularly applying Grid computing, and on the design, implementation, and evaluation of high-performance collaboration environments for anatomic education and surgery. He holds an M.D. from Washington University (St. Louis) and an M.S. from Harvard School of Public Health. He is a Fellow of the American College of Surgeons and a Fellow of the American College of Medical Informatics. Dr. Silverstein provides leadership in information technology initiatives intended to transform operations at the University of Chicago Medical Center and is informatics director for the University of Chicago's Clinical and Translational Science Award (CTSA) program. He has served on various national advisory panels and currently serves on the Board of Scientific Counselors for the Lister Hill Center of the NIH National Library of Medicine.



Patient-specific Finite Element Modeling of Blood Flow and Vessel Wall Dynamics

Charles A. Taylor, *Departments of Bioengineering and Surgery, Stanford University*



Cardiovascular imaging methods, no matter how advanced, can provide data only about the present state and do not provide a means to predict the outcome of an intervention or evaluate alternate prospective therapies. We have developed a computational framework for computing blood flow in anatomically relevant vascular anatomies with unprecedented realism. This framework includes methods for (i) creating subject-specific models of the vascular system from medical imaging data, (ii) specifying boundary conditions to account for vasculature beyond the limits of imaging resolution, (iii) generating finite element meshes, (iv) assigning blood rheological and tissue mechanical properties, (v) simulating blood flow and vessel wall dynamics on parallel computers, and (vi) visualizing simulation results and extracting hemodynamic data. Such computational solutions offer an entirely new era in medicine whereby doctors utilize simulation-based methods, initialized with patient-specific data, to design improved treatments for individuals based on optimizing predicted outcomes. Supercomputers play an essential role in this process and new opportunities for high-performance computing in clinical medicine will be discussed.

Biography: Charles A. Taylor received his B.S. degree in Mechanical Engineering in 1987 from Rensselaer Polytechnic

Institute. He has an M.S. in Mechanical Engineering (1991) and in Mathematics (1992) from RPI. His Ph.D. in Applied Mechanics from Stanford (1996) was for finite element modeling of blood flow, co-advised by Professors Thomas Hughes (Mechanical Engineering) and Christopher Zarins (Surgery). He is currently Associate Professor in the Stanford Departments of Bioengineering and Surgery with courtesy appointments in the Departments of Mechanical Engineering and Radiology. Internationally recognized for development of computer modeling and imaging techniques for cardiovascular disease research, device design and surgery planning, his contributions include the first 3D simulations of blood flow in the abdominal aorta and the first simulations of blood flow in models created from medical images. He started the field of simulation-based medicine using CFD to predict outcomes of cardiovascular interventions and developed techniques to model blood flow in large, patient-specific models, with applications ranging in congenital heart malformations, hypertension and aneurysms. He received Young Investigator Awards in Computational Mechanics from the International Association for Computational Mechanics and from the U.S. Association for Computational Mechanics. He is a Fellow of the American Institute for Medical and Biological Engineering.

Session: Multi-Scale Simulations in Bioscience

Session Chair: Christine Chalk (DOE Office of Science)

3:30pm-5pm

Room: PB253-254



Big Science and Computing Opportunities: Molecular Theory, Models and Simulation

Teresa Head-Gordon (*University of California, Berkeley*)



Molecular simulation is now an accepted and integral part of contemporary chemistry and chemical biology. The allure of molecular simulation is that most if not all relevant structural, kinetic, and thermodynamic observables of a (bio)chemical system can be calculated at one time, in the context of a molecular model that can provide insight, predictions and hypotheses that can stimulate the formulation of new experiments. This talk will describe the exciting opportunities for "big" science questions in chemistry and biology that can be answered with an appropriate theoretical framework and with the aid of high end capability computing*. *The Potential Impact of High-End Capability Computing on Four Illustrative Fields of Science and Engineering, National Research Council ISBN: 978-0-309-12485-0 (2008).

Biography: Teresa Head-Gordon received a B.S. in chemistry from Case Western Reserve Institute of Technology in 1983. After a year of full-time waitressing she decided to expand her employment opportunities by obtaining a Ph.D. in theoretical chemistry from Carnegie Mellon University in 1989, developing simulation methods for macromolecules. In 1990 she became a Postdoctoral Member of Technical Staff at AT&T Bell Laboratories in Murray Hill, NJ working with chemical physicists (Frank Stillinger) and mathematical optimization experts (David Gay and Margaret Wright) on perturbation theories of liquids and protein structure prediction and folding. She joined Lawrence Berkeley National Laboratory as a staff scientist in 1992, and then became assistant professor in Bioengineering at UC Berkeley in 2001, rising to the rank of full professor in 2007. She has served as Editorial Advisory Board Member for the Journal of Physical Chemistry B (2009-present); for the Journal of Computational Chemistry (2004-present); and for the SIAM book series on Computational Science and Engineering (2004-2009), and panel member of the U.S. National Academies Study on Potential Impact of Advances in High-End Computing in Science and Engineering (2006-2008). She spent a most enjoyable year in the Chemistry department at Cambridge University as Schlumberger Professor in 2005-2006 interacting with scientists in the UK and the European continent, biking 13 miles a day to and from the Theory Center on Lensfield Road and sharpening her croquet skills as often as possible in her backyard in a local Cambridge village. She remains passionate about the centrality of "applied theory," i.e. simulation, in the advancement of all physical sciences discoveries and insights in biology, chemistry, and physics.



Fighting Swine Flu through Computational Medicine

Klaus Schulten (University of Illinois at Urbana-Champaign)



The swine flu virus, spreading more and more rapidly, threatens to also become resistant against present forms of treatment. This lecture illustrates how a look through the "computational microscope"

(CM) contributes to shaping a pharmacological strategy against a drug resistant swine flu. The CM is based on simulation software running efficiently on many thousands of processors, analyzes terabytes of data made available through GPU acceleration, and is adopted in the form of robust software by thousands of biomedical researchers worldwide. The swine flu case shows the CM at work: what type of computing demands arise, how software is actually deployed, and what insight emerges from atomic resolution CM images. In viewing, in chemical detail, the recent evolution of the swine flu virus against the binding of modern drugs, computing technology responds to a global 911 call.

Biography: Klaus Schulten received his Ph.D. from Harvard University in 1974. He is Swanlund Professor of Physics and is also affiliated with the Department of Chemistry as well as with the Center for Biophysics and Computational Biology. Professor Schulten is a full-time faculty member in the Beckman Institute and directs the Theoretical and Computational Biophysics Group. His professional interests are theoretical physics and theoretical biology. His current research focuses on the structure and function of supramolecular systems in the living cell, and on the development of non-equilibrium statistical mechanical descriptions and efficient computing tools for structural biology. Honors and awards: Award in Computational Biology 2008; Humboldt Award of the German Humboldt Foundation (2004); University of Illinois Scholar (1996); Fellow of the American Physical Society (1993); Nernst Prize of the Physical Chemistry Society of Germany (1981).

Thursday, November 19

Session: Toward Exascale Climate Modeling

10:30am-Noon

Room: PB252

Session Chair: Anjuli Bamzai (DOE Office of Science)



Toward Climate Modeling in the ExaFlop Era

David Randall (Colorado State University)



Since its beginnings in the 1960s, climate modeling has always made use of the fastest machines available. The equations of fluid motion are solved from the global scale, 40,000 km, down to a truncation scale which has been on the order of a few hundred kilometers. The effects of smaller-scale processes have "parameterized" using less-than-rigorous statistical theories. Recently, two radically new types of model have been demonstrated, both of which replace major elements of these parameterizations with methods based more closely on the basic physics. These new models have been made possible by advances in computing power. The current status of these models will be outlined, and a path to exaflop climate modeling will be sketched. Some of the physical and computational issues will be briefly summarized. **Biography:** David Randall is Professor of Atmospheric Science at Colorado State University. He received his Ph.D. in Atmospheric Sciences from UCLA in 1976. He has been developing global atmospheric models since 1972. He is currently the Director of a National Science Foundation Science and Technology Center on Multiscale Modeling of Atmospheric Processes, and also the Principal Investigator on a SciDAC project with the U.S.

Department of Energy. He has been the Chief Editor of two different journals that deal with climate-modeling. He was a Coordinating Lead Author for the Intergovernmental Panel on Climate Change. He has chaired several science teams. He received NASA's Medal for Distinguished Public Service, NASA's Medal for Exceptional Scientific Achievement, and the Meisinger Award of the American Meteorological Society. He is a Fellow of the American Geophysical Union, the American Association for the Advancement of Science, and the American Meteorological Society.



Green Flash: Exascale Computing for Ultra-High Resolution Climate Modeling

Michael Wehner (Lawrence Berkeley National Laboratory)



Since the first numerical weather experiments in the late 1940s by John Von Neumann, machine limitations have dictated to scientists how nature could be studied by computer simulation. As the processors used in traditional parallel computers have become ever more complex, electrical power demands are approaching hard limits dictated by cooling and power considerations. It is clear that the growth in scientific computing capabilities of the last few decades is not sustainable unless fundamentally new ideas are brought to bear. In this talk we discuss a radically new approach to high-performance computing (HPC) design via an application-driven hardware and software co-design which leverages design principles from the consumer electronics marketplace. The methodology we propose has the potential of significantly improving energy efficiency, reducing cost, and accelerating the development cycle of exascale systems for targeted applications. Our approach is motivated by a desire to simulate the Earth's atmosphere at kilometer scales.

Biography: Michael Wehner is a staff scientist in the Scientific Computing Group in the Computational Research Division at Lawrence Berkeley National Laboratory. His current research focuses on a variety of aspects in the study of climate change. In particular, Wehner is interested in the behavior of extreme weather events in a changing climate, especially heat waves and tropical cyclones. He is also interested in detection and attribution of climate change and the improvement of climate models through better use of high performance computing. Before joining LBNL, Wehner worked as a physicist at Lawrence Livermore National Laboratory, where he was a member of the Program for Climate Modeling and Intercomparison, the Climate System Modeling Group and B Division. He is the author or co-author of 78 papers. Wehner earned his master's degree and Ph.D. in nuclear engineering from the University of Wisconsin-Madison, and his bachelor's degree in Physics from the University of Delaware.

Session: High Performance at Massive Scale

1:30pm-3pm

Room: PB252

Session Chair: Horst Simon (Lawrence Berkeley National Laboratory)



Warehouse-Scale Computers

Urs Hoelzle (Google)



The popularity of Internet-based applications is bringing increased attention to server-side computing, and in particular to the design of the large-scale computing systems that support successful service-based products. In this talk I will cover the issues involved in designing and programming this emerging class of very large (warehouse-scale) computing systems, including physical scale, software and hardware architecture, and energy efficiency.

Biography: Urs Hölzle served as Google's first vice president of engineering and led the development of Google's technical infrastructure. His current responsibilities include the design and operation of the servers, networks, and datacenters that power Google.

High Performance at Massive Scale: Lessons Learned at Facebook

Robert Johnson (Facebook)



Data at facebook grows at an incredible rate with more than 500,000 people registering daily and 200,000,000 existing users increasingly adding new information. The rate of reads against the data also is growing, with more than 50,000,000 low-latency random accesses a second. Building a system to handle this is quite challenging, and I'll be sharing some of our lessons learned. One critical component of the system is the memcached cluster. I'll discuss ways we've maintained performance and reliability as the cluster has grown and techniques to accommodate massive amounts of network traffic. Some techniques are at the application layer, such as data clustering and access pattern analysis. Others are at the system level, such as dynamic network traffic throttling and modifications to the kernel's network stack. I'll also discuss PHP interpreter optimizations to reduce initialization cost with a large codebase and lessons learned scaling MySQL to tens of terabytes. I'll include measured data and discuss the impact of various optimizations on scaling and performance.

Biography: Robert Johnson is Director of Engineering at Facebook, where he leads the software development efforts to cost-effectively scale Facebook's infrastructure and optimize performance for its many millions of users. During his time with the company, the number of users has expanded by more than thirty-fold and Facebook now handles billions of page views a day.

Robert was previously at ActiveVideo Networks where he led the distributed systems and set-top software development teams. He has worked in a wide variety of engineering roles from robotics to embedded systems to web software. He received a B.S. in Engineering and Applied Science from Caltech.

Session: Scalable Algorithms and Applications

3:30pm-5pm

Room: PB252

Session Chair: Douglas Kothe (Oak Ridge National Laboratory)

Scalable Parallel Solvers in Computational Electrophysiology

Luca Pavarino (Universita di Milano)



Research on electrophysiology of the heart has progressed greatly in the last decades, producing a vast body of knowledge ranging from microscopic description of cellular membrane ion channels to macroscopic anisotropic propagation of excitation and repolarization fronts in the whole heart. Multiscale models have proven very useful in the study of these complex phenomena, progressively including more detailed features of each component in more models that couple parabolic systems of nonlinear reaction-diffusion equations with stiff systems of several ordinary differential equations. Numerical integration is, therefore, a challenging large-scale computation, requiring parallel solvers and distributed architectures. We review some recent advances in numerical parallel solution of these cardiac reaction-diffusion models, focusing, in particular, on scalability of domain decomposition iterative solvers that belong to the family of multilevel additive Schwarz preconditioners. Numerical results obtained with the PETSc library on Linux Clusters confirm the scalability and optimality of the proposed solvers, for large-scale simulations of a complete cardiac cycle.

Biography: Luca F. Pavarino is a professor of Numerical Analysis at the University of Milano, Italy. He graduated from the Courant Institute of Mathematical Sciences, New York University, USA, in 1992. After spending two postdoctoral years

at the Department of Computational and Applied Mathematics of Rice University, Houston, USA, he became assistant professor at the University of Pavia, Italy in 1994 and then professor at the University of Milano in 1998. His research activity has focused on domain decomposition methods for elliptic and parabolic partial differential equations discretized with finite or spectral elements, in particular on their construction, analysis and parallel implementation on distributed memory parallel computers. He has applied these parallel numerical methods to problems in computational fluid dynamics, structural mechanics, computational electrocardiology.

Simulation and Animation of Complex Flows on 10,000 Processor Cores

Ulrich Rüde (University of Erlangen-Nuremberg)



The Lattice Boltzmann Method (LBM) is based on a discretization of the Boltzmann equation and results in a cellular automaton that is relatively easy to extend and is well suited for parallelization. In the past few years, the LBM method has been established as an alternative method in computational fluid dynamics. Here, we will discuss extensions of the LBM to compute flows in complex geometries such as blood vessels or porous media, fluid structure interaction problems with moving obstacles, particle laden flows, and flows with free surface. We will outline the implementation in the waLBerla software framework and will present speedup results for current parallel hardware. This includes heterogeneous multicore CPUs, such as the IBM Cell processor, and the implementation on massively parallel systems with thousands of processor cores.

Biography: Ulrich Ruede studied Mathematics and Computer Science at Technische Universität München (TUM) and The Florida State University. He holds a Ph.D. and Habilitation degrees from TUM. After graduation, he spent a year as Post Doc at the University of Colorado to work on parallel, adaptive finite element multigrid methods. He joined the Department of Mathematics at University of Augsburg in 1996 and since 1998 he has been heading the Chair for Simulation at the University of Erlangen-Nuremberg. His research interests are in computational science and engineering, including mathematical modeling, numerical analysis, multigrid methods, architecture-aware algorithms, visualization, and highly scalable methods for complex simulation tasks. He received the ISC Award 2006 for solving the largest finite element system and has been named a Fellow of the Society of Industrial and Applied Mathematics. Currently he also serves as the Editor-in-Chief for the SIAM J. Scientific Computing.



Computing for a Changing World

SC09 Panels

The SC09 panels program aims to bring out the most interesting and challenging ideas at the conference, covering topics that are relevant, timely, and thought provoking. Designed to promote discussion on large and small topics, panels bring together great thinkers, practitioners, and the

occasional gadfly to present and debate various topics relevant to high performance computing, networking, storage and analysis. Audience participation is always encouraged. Drop by a panel and enjoy the show.

Panels

Tuesday, November 17

Special Events

10:30am- Noon; 1:30pm- 3pm

Room: D135-136



Building 3D Internet Experiences

In this half-day special event, technology practitioners representing both HPC and the Virtual Worlds will share their perspectives and experiences creating compelling scientific simulations, visualizations and collaborative educational environments. Through a series of short talks and discussions, these technologists will educate attendees on development processes and programming techniques for architecting 3D Internet experiences for a variety of applications.

3D Internet Panel

Chair: John Hengeveld (Intel Corporation)

3:30pm- 5pm

Room: PB251



3D Cross HPC: Technology and Business Implications

Moderators: John Hengeveld (Intel Corporation)

Panelists: Cristina Videira Lopes (Univ. of California, Irvine), Nicholas Polys (Virginia Tech), Tony O'Driscoll (Duke University), Richard Boyd (Lockheed Martin Corporation)

In this event, two panels will explore the business and technology implications in emerging 3D internet applications which are utilizing HPC hardware and Software solutions. The technology discussion will focus on the technical impact of HPC in improving scalability, parallelism and real time performance of emerging gaming, 3D content, education and scientific visualization experiences delivered via the internet. The business discussion will focus on emerging business models to bring mass adoption of 3D internet technology.

Wednesday, November 18th

Cyberinfrastructure in Healthcare Management

Chair: Arun Datta (National University)

1:30pm- 3pm

Room: PB252



Cyberinfrastructure in Healthcare Management

Moderators: Arun Datta (National University)

Panelists: Terry Boyd (Centers for Disease Control and Prevention), Stanley J. Watowich (University of Texas Medical Branch), Wilfred Li (San Diego Supercomputer Center), Fang-Pang Lin (National Center for High-Performance Computing Taiwan)

This panel will discuss the impact of cyberinfrastructure in healthcare management. The recent outbreaks of H1N1 influenza in US and other parts of the Globe will be discussed as a case study. Swine influenza viruses cause disease in pigs and can mutate to be efficiently transmitted between humans as demonstrated with the current pandemic of swine-derived influenza A(H1N1) virus. The influenza A(H1N1) virus has sequence similarity with circulating avian and human influenza viruses. The panel will share how informatics can take an important role in identifying and controlling this pandemic flu. Discovery research on H1N1 antivirals using World Community Grid will also be discussed. In this context, this panel will also share the role of PRAGMA in controlling SARS outbreak that occurred in 2003 and discuss the possible impact of sharing genomic information in Electronic Health Record of an individual.

Thursday, November 19

Energy Efficient Data Centers for HPC: How Lean and Green do we need to be?

Chair: Michael K Patterson (Intel Corporation)

10:30am- Noon

Room: PB256



Energy Efficient Data Centers for HPC: How Lean and Green do we need to be?

Moderators: Michael K. Patterson (Intel Corporation)

Panelists: William Tschudi (Lawrence Berkeley National Laboratory), Phil Reese (Stanford University), David Seger (IDC Architects), Steve Elbert (Pacific Northwest National Laboratory)

The performance gains of HPC machines on the Top500 list is actually exceeding a Moore's Law growth rate, providing more capability for less energy. The challenge is now the data center design; the power and cooling required to support these machines. The energy cost becomes an increasingly important factor. We will look at best practices and on-going developments in data center and server design that support improvements in TCO and energy use. The panel will also explore future growth and the very large data centers. Microsoft has a 48 MW data center, Google has one at 85 MW. Google also operates one with a PUE of 1.15. How efficient can an exascale system get? Or must be? Is the answer tightly-packed containers or spread-out warehouses? Will we need to sub-cool for performance or run warmer to use less energy? Join us for a lively discussion of these critical issues.

Friday, November 20

Applications Architecture Power Puzzle

Chair: Mark Stalzer (California Institute of Technology)

8:30am- 10am

Room: PB252

Applications Architecture Power Puzzle

Moderators: Mark Stalzer (California Institute of Technology)

Panelists: Thomas Sterling (Louisiana State University), Allan Snively (San Diego Supercomputer Center), Stephen Poole (Oak Ridge National Laboratory), William Camp (Intel Corporation)

Over the next few years there are two boundary conditions that should constrain computer systems architecture: commodity components and applications performance. Yet, these two seem strangely disconnected. Perhaps we need some human optimization, as opposed to repeated use of Moore's Law. Our panelists

have been given a set of standard components that are on announced vendor roadmaps. They also each get to make one mystery component of no more complexity than a commercially available FPGA. The applications are HPL for linear algebra, Map-Reduce for databases, and a sequence matching algorithm for biology. The panelists have 10 minutes to disclose their systems architecture and mystery component, and estimate performance for the three applications at 1MW of power.

Flash Technology in HPC: Let the Revolution Begin

Chair: Robert Murphy (Sun Microsystems)

8:30am- 10 a.m.

Room: PB251

Flash Technology in HPC: Let the Revolution Begin

Moderators: Robert Murphy (Sun Microsystems)

Panelists: Andreas Bechtolsheim (Sun Microsystems), David Flynn (Fusion-io), Paresh Pattani (Intel Corporation), Jan Silverman (Spansion)

The inexorable march of ever more CPU cores, and even more GFLOPS per core, means HPC application performance will no longer be bounded by the CPU, but by getting data into and out of these fast processors. I/O has always lagged behind computation, being ultimately dependent on disk drives throttled by rotational rates limited by mechanical physics. With an exponential growth spurt of peak GFLOPs available to HPC system designers and users imminent, the CPU performance I/O gap will reach increasingly gaping proportions. To bridge this gap, Flash is suddenly being deployed in HPC as a revolutionary technology that delivers faster time to solution for HPC applications at significantly lower costs and lower power consumption than traditional disk based approaches. This panel, consisting of experts representing all points of the Flash technology spectrum, will examine how Flash can be deployed and the effect it will have on HPC workloads.

Preparing the World for Ubiquitous Parallelism

Chair: Paul Steinberg (Intel Corporation)

10:30am- Noon

Room: PB252

Preparing the World for Ubiquitous Parallelism

Moderators: Matthew Wolf (Georgia Institute of Technology)

Panelists: Benedict Gaster (AMD), Kevin Goldsmith (Adobe Systems Inc.), Tom Murphy (Contra Costa College), Steven Parker (NVIDIA), Michael Wrinn (Intel Corporation)

Multicore platforms are transforming the nature of computation. Increasingly, FLOPs are free, but the people who know how to program these petascale, manycore platforms are not. This panel, composed of a diverse set of industry and academic representatives, is focused on presenting and discussing the abstractions, models, and (re-)training necessary to move parallel programming into a broad audience. The panelists, often vigorous competitors, will present the situation from their respective viewpoints, likely fueling a lively discussion of the entire room as has been seen recently in other venues.

The Road to Exascale: Hardware and Software Challenges

Chair: Bill Camp (Sandia National Laboratories)

10:30am- Noon

Room: PB251

The Road to Exascale: Hardware and Software Challenges

Moderators: William J. Camp (Intel Corporation)

Panelists: Jack Dongarra (Oak Ridge National Laboratory / University of Tennessee, Knoxville), Peter Kogge (University of Notre Dame), Marc Snir (University of Illinois at Urbana-Champaign), Steve Scott (Cray Inc.)

For three decades we have increased the performance of highest-end computing about 1000-fold each decade. In doing so, we have had to increase the degree of parallelism, cost, footprint and power dissipation of the top systems in each generation. Now we are challenged to move from petascale in 2008-9 to exascale in 2018-19. The obstacles we face are qualitatively harder than any we have dealt with until now and are unprecedented even by the history we build on. To achieve reasonable power, reliability, programmability, cost, and size targets will take major breakthroughs in hardware and software. For example we will have to create and program systems with hundreds of millions of processor cores.

This panel will review the hardware and software advances needed for exascale computing and will attempt to lay out research, development and engineering roadmaps to successfully overcome the obstacles in our path and create the needed advances.



Computing for a Changing World

SC09 Awards

ACM Gordon Bell Prize

The Gordon Bell Prize is awarded each year to recognize outstanding achievement in HPC. Now administered by the ACM, financial support of the \$10,000 award is provided by Gordon Bell, a pioneer in high performance and parallel computing. The purpose of the award is to track the progress over time of parallel computing, with particular emphasis on rewarding innovation in applying HPC to applications in science. Gordon Bell prizes have been awarded every year since 1987. Prizes may be awarded for peak performance, as well as special achievements in scalability, time-to-solution on important science and engineering problems, and low price/performance.

Student Awards

George Michael Memorial HPC Fellowship Program

The Association of Computing Machinery (ACM), IEEE Computer Society and the SC Conference series established the High Performance Computing (HPC) Ph.D. Fellowship Program to honor exceptional Ph.D. students throughout the world with the focus areas of high performance computing, networking, storage and analysis. Fellowship recipients are selected based on: overall potential for research excellence; the degree to which their technical

interests align with those of the HPC community; their academic progress to date, as evidenced by publications and endorsements from their faculty advisor and department head as well as a plan of study to enhance HPC-related skills; and the demonstration of their anticipated use of HPC resources. This year there are two events for the fellowship; presentations of the work of the three students awarded fellowships in SC08 and the announcement of the SC09 awardees.

ACM Student Research Competition

Since 2006, students have been invited to submit posters as part of the internationally recognized ACM Student Research Competition (SRC), sponsored by Microsoft Research. The SRC venue allows students the opportunity to experience the research world, share research results with other students and conference attendees, and rub shoulders with leaders from academia and industry. Winners will receive cash prizes and will be eligible to enter the annual Grand Finals, the culmination of the ACM SRCs.

Doctoral Research Showcase

The Doctoral Showcase invites Ph.D. students in HPC, networking, storage, and analytics, who anticipate graduating within 12 months to submit a short summary of their research. Those selected will be able to present a 15-minute summary of their best research to experts in academia, industry and research laboratories.

Tuesday, November 17

ACM Gordon Bell Finalists 1

Chair: Douglas Kothe (Oak Ridge National Laboratory)

3:30pm-5pm

Room: E145-146

Beyond Homogenous Decomposition: Scaling Long-Range Forces on Massively Parallel Architectures

David F. Richards (Lawrence Livermore National Laboratory), James N. Glosli (Lawrence Livermore National Laboratory), Bor Chan (Lawrence Livermore National Laboratory), Milo R. Dorr (Lawrence Livermore National Laboratory), Erik W. Draeger (Lawrence Livermore National Laboratory), Jean-luc Fattebert (Lawrence Livermore National Laboratory), William D. Krauss (Lawrence Livermore National Laboratory), Michael P. Surh (Lawrence Livermore National Laboratory), John A. Gunnels (IBM Corporation), Frederick H. Streitz (Lawrence Livermore National Laboratory)

With supercomputers anticipated to expand from thousands to millions of cores, one of the challenges facing scientists is how to effectively utilize the ever-increasing number of tasks. We report here an approach that creates a heterogeneous decomposition by partitioning effort according to the scaling properties of the component algorithms. We have demonstrated our strategy by performing a 140 million-particle MD simulation of stopping power in a hot, dense, hydrogen plasma and benchmarked calculations with over 10 billion particles. With this unprecedented simulation capability we are beginning an investigation of plasma properties under conditions where both theory and experiment are lacking- in the strongly-coupled regime as the plasma begins to burn. Our strategy is applicable to other problems involving long-range forces (i.e., biological or astrophysical simulations). More broadly we believe that the general approach of heterogeneous decomposition will allow many problems to scale across current and next-generation machines.

A Scalable Method for Ab Initio Computation of Free Energies in Nanoscale Systems

Markus Eisenbach (Oak Ridge National Laboratory), Chenggang Zhou (J.P. Morgan Chase & Co), Donald M. Nicholson (Oak Ridge National Laboratory), Gregory Brown (Florida State University), Jeff Larkin (Cray Inc.), Thomas C. Schulthess (ETH Zürich)

Calculating the thermodynamics of nanoscale systems presents challenges in the simultaneous treatment of the electronic structure, which determines the interactions between atoms, and the

statistical fluctuations that become ever more important at shorter length scales. Here we present a highly scalable method that combines ab initio electronic structure techniques, Locally Self-Consistent Multiple Scattering (LSMS) with the Wang-Landau (WL) algorithm to compute free energies and other thermodynamic properties of nanoscale systems. The combined WL-LSMS code is targeted to the study of nanomagnetic systems that have anywhere from about one hundred to a few thousand atoms. The code scales very well on the Cray XT5 system at ORNL, sustaining 1.03 Petaflop/s in double precision on 147,464 cores.

Liquid Water: Obtaining the Right Answer for the Right Reasons

Edoardo Apra (Oak Ridge National Laboratory), Robert J. Harrison (Oak Ridge National Laboratory), Wibe A. de Jong (Pacific Northwest National Laboratory), Alistair Rendell (Australian National University), Vinod Tipparaju (Oak Ridge National Laboratory), Sotiris Xantheas (Pacific Northwest National Laboratory)

Water is ubiquitous on our planet and plays an essential role in several key chemical and biological processes. Accurate models for water are crucial in understanding, controlling and predicting the physical and chemical properties of complex aqueous systems. Over the last few years we have been developing a molecular-level based approach for a macroscopic model for water that is based on the explicit description of the underlying intermolecular interactions between molecules in water clusters. As an example of the benchmarks needed for the development of accurate models for the interaction between water molecules, for the most stable structure of (H₂O)₂₀ we ran a coupled-cluster calculation on the ORNL's Jaguar petaflop computer that used over 100 TB of memory for a sustained performance of 487 TFLOP/s (double precision) on 96,000 processors, lasting for 2 hours. By this summer we will have studied multiple structures of both (H₂O)₂₀ and (H₂O)₂₄ and completed basis set and other convergence studies.

Wednesday, November 18

ACM Gordon Bell Finalists 2

Chair: Jack Dongarra (University of Tennessee, Knoxville)

1:30pm- 3pm

Room: D135-136

Enabling High-Fidelity Neutron Transport Simulations on Petascale Architectures

Dinesh Kaushik (Argonne National Laboratory), Micheal Smith (Argonne National Laboratory), Allan Wollaber (Argonne National Laboratory), Barry Smith (Argonne National Laboratory), Andrew Siegel (Argonne National Laboratory), Won Sik Yang (Argonne National Laboratory)

The UNIC code is being developed as part of the DOE's Nuclear Energy Advanced Modeling and Simulation (NEAMS) program. UNIC is an unstructured, deterministic neutron transport code that allows a highly detailed description of a nuclear reactor core. The goal of this simulation effort is to reduce the uncertainties and biases in reactor design calculations by progressively replacing existing multi-level averaging (homogenization) techniques with more direct solution methods. Since the neutron transport equation is seven dimensional (three in space, two in angle, one in energy, and one in time), these simulations are among the most memory and computationally intensive in all of computational science. In this paper, we present UNIC simulation results for the sodium-cooled fast reactor PHENIX and the critical experiment facility ZPR-6. In each case, UNIC shows excellent weak scalability on up to 163,840 cores of BlueGene/P (Argonne) and 122,800 cores of XT5 (ORNL).

Scalable Implicit Finite Element Solver for Massively Parallel Processing with Demonstration to 160K cores

Onkar Sahni (Rensselaer Polytechnic University), Min Zhou (Rensselaer Polytechnic University), Mark S. Shephard (Rensselaer Polytechnic University), Kenneth E. Jansen (Rensselaer Polytechnic University)

Implicit methods for partial differential equations using unstructured meshes allow for an efficient solution strategy for many real-world problems (e.g., simulation-based virtual surgical planning). Scalable solvers employing these methods not only enable solution of extremely-large practical problems but also lead to dramatic compression in time-to-solution. We present a parallelization paradigm and associated procedures that enable our implicit, unstructured flow-solver to achieve strong scalability. We consider fluid-flow examples in two application areas to show

the effectiveness of our procedures that yield near-perfect strong-scaling on various (including near-petascale) systems. The first area includes a double-throat nozzle (DTN) whereas the second considers a patient-specific abdominal aortic aneurysm (AAA) model. We present excellent strong-scaling on three cases ranging from relatively small to large; a DTN model with $O(10^6)$ elements up to 8,192 cores (9 core-doublings), an AAA model with $O(10^8)$ elements up to 32,768 cores (6 core-doublings) and $O(10^9)$ elements up to 163,840 cores.

42 TFlops Hierarchical N-body Simulations on GPUs with Applications in both Astrophysics and Turbulence

Tsuyoshi Hamada (Nagasaki University), Rio Yokota (University of Bristol), Keigo Nitadori (RIKEN), Tetsu Narumi (University of Electro-Communications), Kenji Yasuoka (Keio University), Makoto Tajiri (RIKEN), Kiyoshi Oguri (Nagasaki University)

We have performed a hierarchical N-body simulation on a cluster of 256 GPUs. Unlike many previous N-body simulations on GPUs that scale as $O(N^2)$, the present method calculates the $O(N \log N)$ treecode and $O(N)$ fast multipole method on the GPUs with unprecedented efficiency. We demonstrate the performance of our method by choosing one standard application -- a gravitational N-body simulation -- and one non-standard application -- the simulation of turbulence using vortex particles. The gravitational simulation using the treecode with 1.6 billion particles showed a sustained performance of 42 TFlops (28 TFlops when corrected). The vortex particle simulation of homogeneous isotropic turbulence using the FMM with 17 million particles showed a sustained performance of 20.2 TFlops. The overall cost of the hardware was 228,912 dollars, which results in a cost performance of $28,000,000/228,912 = 124$ MFlops/\$. The good scalability up to 256 GPUs demonstrates the good compatibility between our computational algorithm and the GPU architecture.

ACM Gordon Bell Finalists 3

Chair: Mateo Valero (Barcelona Supercomputing Center)

3:30pm-5pm

Room: PB255



Indexing Genomic Sequences on the IBM Blue Gene

Amol Ghoting (IBM T.J. Watson Research Center), Konstantin Makarychev (IBM T.J. Watson Research Center)

With advances in sequencing technology and through aggressive sequencing efforts, DNA sequence data sets have been growing at a rapid pace. To gain from these advances, it is important to provide life science researchers with the ability to process and

query large sequence data sets. For the past three decades, the suffix tree has served as a fundamental data structure in processing sequential data sets. However, tree construction times on large data sets have been excessive. While parallel suffix tree construction is an obvious solution to reduce execution times, poor locality of reference has limited parallel performance. In this paper, we show that through careful parallel algorithm design, this limitation can be removed, allowing tree construction to scale to massively parallel systems like the IBM Blue Gene. We demonstrate that the entire Human genome can be indexed on 1024 processors in under 15 minutes.

The Cat is Out of the Bag: Cortical Simulations with 10^9 Neurons, 10^{13} Synapses

Rajagopal Ananthanarayanan (IBM Almaden Research Center), Steven K. Esser (IBM Almaden Research Center), Horst D. Simon (Lawrence Berkeley National Laboratory), Dharmendra S. Modha (IBM Almaden Research Center)

In the quest for cognitive computing, we have built a massively parallel cortical simulator, C2, that incorporates a number of innovations in computation, memory, and communication. Using C2 on LLNL's Dawn Blue Gene/P supercomputer with 147,456 CPUs and 144 TB of main memory, we report two cortical simulations -- at unprecedented scale -- that effectively saturate the entire memory capacity and refresh it at least every simulated second. The first simulation consists of 1.6 billion neurons and 8.87 trillion synapses with experimentally-measured gray matter thalamocortical connectivity. The second simulation has 900 million neurons and 9 trillion synapses with probabilistic connectivity. We demonstrate nearly perfect weak scaling and attractive strong scaling. The simulations, which incorporate phenomenological spiking neurons, individual learning synapses, axonal delays, and dynamic synaptic channels, exceed the scale of the cat cortex, marking the dawn of a new era in the scale of cortical simulations.



Millisecond-Scale Molecular Dynamics Simulations on Anton*

David E. Shaw (D.E. Shaw Research), Ron O. Dror (D.E. Shaw Research), John K. Salmon (D.E. Shaw Research), J.P. Grossman (D.E. Shaw Research), Kenneth M. Mackenzie (D.E. Shaw Research), Joseph A. Bank (D.E. Shaw Research), Cliff Young (D.E. Shaw Research), Martin M. Deneroff (D.E. Shaw Research), Brannon Batson (D.E. Shaw Research), Kevin J. Bowers (D.E. Shaw Research), Edmond Chow (D.E. Shaw Research), Michael P. Eastwood (D.E. Shaw Research), Douglas J. Ierardi (D.E. Shaw Research), John L. Klepeis (D.E. Shaw Research), Jeffrey S. Kuskin (D.E. Shaw Research), Richard H. Larson (D.E. Shaw Research), Kresten Lindorff-Larsen (D.E. Shaw

Research), Paul Maragakis (D.E. Shaw Research), Mark A. Moraes (D.E. Shaw Research), Stefano Piana (D.E. Shaw Research), Yibing Shan (D.E. Shaw Research), Brian Towles (D.E. Shaw Research)

Anton is a recently completed special-purpose supercomputer designed for molecular dynamics (MD) simulations of biomolecular systems. The machine's specialized hardware dramatically increases the speed of MD calculations, making possible for the first time the simulation of biological molecules at an atomic level of detail for periods on the order of a millisecond -- about two orders of magnitude beyond the previous state of the art. Anton is now running simulations on a timescale at which many critically important, but poorly understood phenomena are known to occur, allowing the observation of aspects of protein dynamics that were previously inaccessible to both computational and experimental study. Here, we report Anton's performance when executing actual MD simulations whose accuracy has been validated against both existing MD software and experimental observations. We also discuss the manner in which novel algorithms have been coordinated with Anton's co-designed, application-specific hardware to achieve these results.

*This paper was also accepted in the Technical Papers program and is also a finalist for the Best Paper award.

Wednesday, November 18

Doctoral Showcase I

Chair: X. Sherry Li (Lawrence Berkeley National Laboratory)

3:30pm-5pm

Room: PB252

Scalable Automatic Topology Aware Mapping for Large Supercomputers

Presenter: Abhinav Bhatele (University of Illinois at Urbana-Champaign)

This dissertation will demonstrate the effect of network contention on message latencies and propose and evaluate techniques to minimize communication traffic and hence, bandwidth congestion on the network. This would be achieved by topology-aware mapping of tasks in an application. By placing communication tasks on processors which are in physical proximity on the network, communication can be restricted to near neighbors. Our aim is to minimize hop-bytes, which is a weighted sum of the number of hops between the source and destination for all messages, the weights being the message sizes. This can minimize the communication time and hence, lead to significant speed-ups for parallel applications and also remove scaling bottlenecks in certain cases. The dissertation will involve developing a general

automatic topology-aware mapping framework which takes the task graph and processor graph as input, and outputs near-optimal mapping solutions.

Performance Analysis of Parallel Programs: From Multicore to Petascale

Presenter: Nathan Tallent (Rice University)

The challenge of developing scalable parallel applications is only partially aided by existing languages and compilers. As a result, manual performance tuning is often necessary to identify and resolve poor parallel efficiency. To help resolve performance problems, I have developed novel techniques for pinpointing performance problems in multithreaded and petascale programs. These techniques are based on attributing very precise (instruction-level) measurements to their full static and dynamic context --- all for a run time overhead of less than 5%. First, I present binary analysis for performance measurement and attribution of fully-optimized programs. Second, I describe techniques for measuring and analyzing the performance of multithreaded executions. Third, I show how to diagnose scalability bottlenecks in petascale applications. This work is implemented in Rice University's HPCToolkit performance tools.



Energy Efficiency Optimizations using Helper Threads in Chip Multiprocessors

Presenter: Yang Ding (Pennsylvania State University)

Chip multiprocessors (CMPs) are expected to dominate the landscape of computer architecture in the near future. As the number of processor cores is expected to keep increasing, how to fully utilize the abundant computing resources on CMPs becomes a critical issue. My thesis work focuses on the optimizations in resource management for CMPs and explores the trade-offs between performance and energy consumption. I investigate the benefits by exposing application-specific characteristics to helper threads in the OS and runtime system in an attempt to enable software to make better use of the new architecture features in hardware. Specifically, I propose using helper threads in (1) adapting application executions to the core availability change; (2) adapt application executions to the core-to-core variations; and (3) dynamic management of shared resources across multiple applications. These studies indicate that helper threads can be very useful for utilizing CMPs in an energy-efficient manner.

Consistency Aware, Collaborative Workflow Developer Environments

Presenter: Gergely Sipos (MTA SZTAKI)

Real-time collaborative editing systems allow a group of users to view and edit the same item at the same time from geographically dispersed sites. Consistency maintenance in the face of concurrent accesses to shared entities is one of the core issues in the design of these systems. In my research I provide solution for the problem of consistency maintenance of acyclic grid workflows in collaborative editing environments. I develop locking based algorithms to assure that under no circumstances can application developers break the acyclic criteria of workflow graphs or add invalid edges to them. I prove that the algorithms result consistent graphs, moreover, do not result the cancellation of any user's editing transaction. This is highly important because grid workflow development is time consuming and knowledge intensive process. I develop a formal evaluation method to compare the efficiency of the algorithms and to find the most optimal locking method.

Architecture and Performance of Runtime Environments for Data Intensive Scalable Computing

Presenter: Jaliya Ekanayake (Indiana University Bloomington)

Large scale data/compute intensive applications found in many domains such as particle physics, biology, chemistry, and information retrieval are composed from multiple "filters" for which many parallelization techniques can be applied. Filters that perform "pleasingly parallel" operations can be built using cloud technologies such as Google MapReduce, Hadoop, and Dryad. Others require frameworks such as ROOT and R or MPI style functionalities. My research identifies MapReduce enhancements that can be applied to large classes of filters. I have developed an architecture and a prototype implementation of a new programming model based on MapReduce, which support; faster intermediate data transfers, long-running map/reduce tasks, and iterative MapReduce computations, by combining the MapReduce programming model and the data streaming techniques. The research also identifies the basic classes of filter pipelines, and discusses their mapping to the different cloud technologies along with a detailed analysis of their performances, on both direct and virtualized hardware platforms.

Providing Access to Large Scientific Datasets on Clustered Databases

Presenter: Eric Perlman (Johns Hopkins University)

My research enables scientists to perform new types of analyses over large datasets. In one example, I helped build a new environment for computational turbulence based on storing the complete space-time histories of the solution in a cluster of databases. We use a regular indexing and addressing scheme for both partitioning data and workload scheduling across the cluster. This environment provides good support for data exploration, particularly for examining large time-spans or iterating back and forth through time. In another area, estuarine research, I worked on new techniques to improve cross-dataset analysis. I used computational geometry techniques to automatically characterize the region of space from which spatial data are drawn, partition the region based on that characterization, and create an index from the partitions. This can significantly reduce the number of I/Os required in query processing, particularly on queries focusing on individual estuaries and rivers in large water networks.

ACM Student Research Competition

Wednesday, November 18

ACM Student Research Competition

1:30pm-3:00pm

Room: PB251

On the Efficacy of Haskell for High-Performance Computational Biology

Author: Jacqueline R. Addesa (Virginia Tech)

A Policy Based Data Placement Service

Author: Muhammad Ali Amer (University of Southern California)

Hiding Communication and Tuning Scientific Applications using Graph-Based Execution

Author: Pietro Cicotti (University of California, San Diego)

An Automated Air Temperature Analysis and Prediction System for the Blue Gene/P

Author: Neal R. Conrad (Argonne National Laboratory)

CUSUMMA: Scalable Matrix-Matrix Multiplication on GPUs with CUDA

Author: Byron V. Galbraith (Marquette University)

BGPStatusView Overview

Author: Matthew H. Kemp (Argonne National Laboratory)

Communication Optimizations of SCEC CME AWP-Olsen Application for Petascale Computing

Author: Kwangyoon Lee (University of California, San Diego)

IO Optimizations of SCEC AWP-Olsen Application for Petascale Earthquake Computing

Author: Kwangyoon Lee (University of California, San Diego)

A Hierarchical Approach for Scalability Enhancement in Distributed Network Simulations

Author: Sowmya Myneni (New Mexico State University)

Large-Scale Wavefront Parallelization on Multiple Cores for Sequence Alignment

Author: Shannon I. Steinfadt (Kent State University)

A Feature Reduction Scheme for Obtaining Cost-Effective High-Accuracy Classifiers for Linear Solver Selection

Author: Brice A. Toth (Pennsylvania State University)

strategy produces highly accurate classifiers with low computational costs

Parallelization of Tau-Leaping Coarse-Grained Monte Carlo Method for Efficient and Accurate Simulations on GPUs

Author: Lifan Xu (University of Delaware)

See Poster Session, Tuesday, 5:15pm - 7pm, for a listing and abstracts of the finalist.

Thursday, November 19

Doctoral Showcase II

Chair: Umit Catalyurek (Ohio State University)

3:30pm-5pm

Room: PB252

Processing Data Intensive Queries in Global-Scale Scientific Database Federations

Presenter: Xiaodan Wang (Johns Hopkins University)

Global scale database federations are an attractive solution for managing petabyte scale scientific repositories. Increasingly important are queries that comb through and extract features from data that are distributed across multiple nodes. These queries are I/O intensive (scans of multi-terabyte tables) and generate considerable network traffic. We develop scheduling techniques that ensure high throughput when concurrently executing queries compete for shared resources. First, we put forth LifeRaft, a batch processing system that eliminates redundant disk accesses by identifying partial overlap in the data accessed by incoming queries and reordering existing queries to maximize data sharing. Instrumenting LifeRaft for Astronomy workloads demonstrates a two-fold improvement in query throughput. We also develop algorithms that incorporate network structure when scheduling join queries across multiple nodes. We exploit excess network

capacity and minimize the utilization of network resources over multiple queries. This provides an order of magnitude reduction in network cost for Astronomy federations.



GPGPU and Cloud Computing for DNA Sequence Analysis

Presenter: Michael C. Schatz (University of Maryland)

Recent advances in DNA sequencing technology have dramatically increased in the scale and scope of DNA sequence analysis, but these analyses are complicated by the volume and complexity of the data involved. For example, genome assembly computes the complete sequence of a genome from billions of short fragments read by a DNA sequencer, and read mapping computes how short sequences match a reference genome to discover conserved and polymorphic regions. Both of these important computations benefit from recent advances in high performance computing, such as in MUMmerGPU (<http://mummergpu.sourceforge.net>) which uses highly parallel graphics processing units as high performance parallel processors, and in CloudBurst (<http://cloudburst-bio.sourceforge.net>) and

CrossBow (<http://bowtie-bio.sourceforge.net/crossbow>) which use the MapReduce framework coupled with cloud computing to parallelize read mapping across large remote compute grids. These techniques have demonstrated orders of magnitude improvements in computation time for these problems, and have the potential to make otherwise infeasible studies practical.

Providing QoS for Heterogeneous Workloads in Large, Volatile, and Non-Dedicated Distributed Systems

Presenter: Trilce Estrada (University of Delaware)

It is widely accepted that heterogeneous, distributed systems can provide QoS only if resources are guaranteed or reserved in advance. Contrary to this belief, in this work I show that QoS can also be provided by heterogeneous, distributed systems even if they are highly volatile, non-dedicated, and whose resources cannot be reserved a-priori. In particular, I focus on Volunteer Computing (VC) as a representative of the latter environment. In this context, QoS means maximizing throughput and speed while producing a robust and accurate solution within time constraints. VC resources are heterogeneous and may not always suit job requirements resulting in higher latencies and lower throughput for the VC application. Using statistical modeling and machine learning, I show how adaptive mechanisms can balance tradeoffs between resource availability, job requirements, and application needs providing the QoS required for heterogeneous workloads in large, volatile, and non-dedicated distributed environments such as the VC environment.

Computer Generation of FFT Libraries for Distributed Memory Computing Platforms

Presenter: Srinivas Chellappa (Carnegie Mellon University)

Existing approaches for realizing high performance libraries for the discrete Fourier transform (DFT) on supercomputers and other distributed memory platforms (including many-core processors), have largely been based on manual implementation and tuning. This reflects the complex nature of mapping algorithms to the platform to produce code that is vectorized, parallelized and takes into account the target's interconnection architecture. Our goal is to enable computer generation of high-performance DFT libraries for distributed memory parallel processing systems, given only a high-level description of a set of DFT algorithms and some platform parameters. We base our approach on Spiral, a program generation system that uses a domain-specific language to rewrite algorithms at a high abstraction level to fit the target platform. Initial results on the Cell processor have successfully generated code that performs up to 4.5x faster than existing cell-specific DFT implementations. Current work is focused on cluster platforms.

Adaptive Runtime Optimization of MPI Binaries

Presenter: Thorvald Natvig (Norwegian University of Science & Technology)

When moving applications from traditional supercomputers to Ethernet-connected clusters, scalability of existing programs tend to suffer since the communication/computation ratios increase. Ethernets have acceptable bandwidth, but their much higher latencies imply many programs have to be rewritten to use asynchronous operations to hide the latency. Unfortunately, this can be a time-consuming affair. Our method injects itself at runtime into the application, intercepting all calls to MPI communication functions. The call is started in an asynchronous manner, but the memory associated with the request is protected by our method. Hence, if the application accesses the same memory, we make it wait for communication function to finish before allowing the application to continue. We have extended our method to also work on file I/O using the same techniques. Our methods also work on any application, without changing the source code or even requiring recompilation.

An Integrated Framework for Parameter-Based Optimization of Scientific Workflows

Presenter: Vijay S. Kumar (Ohio State University)

Scientific applications that analyze terabyte-scale, multidimensional datasets are expressed as dataflow-graphs that are configured to accommodate a range of queries. End-users who cannot afford high execution times (due to various resource constraints) trade accuracy of analysis for performance gains, and supplement their queries with certain minimum quality requirements for the analysis to meet. My thesis focuses on performance optimizations for such applications, with emphasis on supporting accuracy-related trade-offs. The proposed framework allows users to express high-level constraints and quality requirements, and transforms these into appropriate low-level execution strategies for distributed computing environments. Using semantic representations of a domain's data and application characteristics, the framework allows explicit definition of various performance and accuracy-related parameters that influence application execution. Performance optimization is viewed as tuning of these parameters (by runtime systems) to determine their optimal values. Our



Computing for a Changing World

SC09 Challenges

The SC09 Conference Challenges provide a way to showcase both expertise and HPC resources in friendly yet spirited competitions with other participants. This year SC09 features the Storage and Bandwidth challenges, along with the Student Cluster Competition, formerly known as the Cluster Challenge. This year's challenges complement the conference theme and technical program agenda.

The Bandwidth Challenge

The Bandwidth Challenge is an annual competition for leading-edge network applications developed by teams of researchers from around the globe. Past competitions have showcased multi-10gigabit-per-second demonstrations. This year's Bandwidth Challenge will focus on real-world applications and data movement issues and will include live demonstrations across the SCinet infrastructure.

Student Cluster Competition

The Student Cluster Competition will showcase the amazing power of clusters and the ability to utilize open source software to solve interesting and important problems. Only teams comprising students (both high-school and college) are eligible to participate. Teams will compete in

real-time on the exhibit floor, where they will run a workload of real-world applications on clusters of their own design. The winning team will be chosen based on workload completed, benchmark performance, and overall knowledge of the applications. Another award will be given for the most power-efficient design.

Storage Challenge

The Storage Challenge is a competition showcasing applications and environments that effectively use the storage subsystem in high performance computing, which is often the limiting system component. Submissions can be based upon tried and true production systems as well as research or proof of concept projects not yet in production. Participants will describe their implementations and present measurements of performance, scalability and storage subsystem utilization. Judging will be based on these measurements as well as innovation and effectiveness; maximum size and peak performance are not the sole criteria. Finalists will be chosen on the basis of submissions that are in the form of proposals; they will present their completed results in a technical session on Tuesday, November 17, 10:30 pm, at which the winners will be selected.

Storage Challenge Finalists

Tuesday, November 17

Chair: Raymond L. Paden (IBM Corporation)

10:30am-Noon

Room: PB251



Low Power Amdahl-Balanced Blades for Data Intensive Computing

Team Members: Alexander Szalay (Johns Hopkins University), Andreas Terzis (Johns Hopkins University), Alainna White (Johns Hopkins University), Howie Huang (George Washington University), Jan Vandenberg (Johns Hopkins University), Tamas Budavari (Johns Hopkins University), Sam Carliles (Johns Hopkins University), Alireza Mortazavi (Johns Hopkins University), Gordon Bell (Microsoft Research), Jose Blakeley (Microsoft Corporation), David Luebke (NVIDIA), Michael Schuette (OCZ Technology), Ani Thakar (Johns Hopkins University)

Enterprise and scientific data sets double every year, forcing similar growths in storage size and power consumption. As a consequence, current system architectures used to build data warehouses are hitting a power consumption wall. We propose a novel alternative architecture comprising a large number of so-called "Amdahl blades" that combine energy-efficient CPUs and GPUs with solid state disks to increase sequential I/O throughput by an order of magnitude while keeping power consumption constant. Offloading throughput-intensive analysis to integrated GPUs in the chipset increases the relative performance per Watt. We also show that while keeping the total cost of ownership constant, Amdahl blades offer five times the throughput of a state-of-the-art computing cluster for data-intensive applications. Finally, using the scaling laws originally postulated by Amdahl, we show that systems for data-intensive computing must maintain a balance between low power consumption and per-server throughput to optimize performance per Watt.

Accelerating Supercomputer Storage I/O Performance

Team Members: I-Hsin Chung (IBM Research), Seetharami Seelam (IBM Research), Guojing Cong (IBM Research), Hui-Fang Wen (IBM Research), David Klepacki (IBM Research)

We present an application level I/O caching and prefetching system to hide I/O access latency experienced by parallel applications. Our solution of user controllable caching and prefetching system maintains a file-IO cache in the user space of the application, analyzes the I/O access patterns, prefetches requests, and performs write-back of dirty data to storage asynchronously. So

each time the application does not have to pay the full I/O latency penalty in going to the storage and getting the required data. This extra layer will handle the I/O for the application and drives the storage to hardware limits. We have implemented this caching and asynchronous prefetching on the Blue Gene/P system. We present experimental results with ocean and weather related benchmarks: POP, WRF. The initial results on a two-rack BG/P system demonstrate that our method hides access latency and improves application I/O access time by as much as 65%.

Data Intensive Science: Solving Scientific Unknowns by Solving Storage Problems

Team Members: Arun Jagatheesan (San Diego Supercomputer Center), Jiahua He (San Diego Supercomputer Center), Allan Snively (San Diego Supercomputer Center), Michael Norman (San Diego Supercomputer Center), Eva Hocks (San Diego Supercomputer Center), Jeffrey Bennett (San Diego Supercomputer Center), Larry Diegel (San Diego Supercomputer Center)

We want to promote data intensive science by alleviating problems faced by our scientific user community with storage latency. We demonstrate how we could achieve scientific successes by mitigating the storage latency problem through a combination of changes in hardware and storage system architecture. We perform several experiments on our storage infrastructure test bed consisting of 4 TB of flash disks and 40 TB of disk storage, in a system called "SDSC DASH." We use real-world scientific applications in our experiments to show how existing applications could benefit from our proposed new technologies. Our experiments on DASH would guide our scientific community in design, deployment and use of data intensive scientific applications. In our final report, we will contribute to the best practices for effective use of new storage technologies and provide techniques such as memory scalability and memory scheduling that can be used with these new technologies.

An Efficient and Flexible Parallel I/O Implementation for the CFD General Notation System

Team Members: Kyle Horne (Utah State University), Nate Benson (Utah State University), Thomas Hauser (Utah State University)

One important, often overlooked, issue for large, three dimensional time-dependent computational fluid dynamics (CFD) simulations is the input and output performance of the CFD solver, especially for large time-dependent simulations. The development of the CFD General Notation System (CGNS) has brought a standardized and robust data format to the CFD community, enabling the exchange of information between the various stages of numerical simulations. Application of this standard

data format to large parallel simulations is hindered by the reliance of most applications on the CGNS Mid-Level library. This library has only supported serialized I/O. By moving to HDF5 as the recommended low level data storage format, the CGNS standards committee has created the opportunity to support parallel I/O. In our work, we present the parallel implementation of the CGNS Mid-Level library and detailed benchmarks on parallel file systems for typical structured and unstructured CFD applications.



Computing for a Changing World

SC09 Exhibitor Forum

The Exhibitor Forum showcases the latest advances of industry exhibitors, including new products and upgrades, recent research and development initiatives, and future plans and roadmaps. Thought-leaders from industry will give insight into the technology trends driving exhibitors' strategies, the potential of emerging technologies in their product lines, or the impact

of adopting academic research into their development cycle. Topics cover a wide range of products and technologies, including supercomputer architectures, scalable 4GL language, parallel programming and debugging tools, and storage systems. Whatever the topic, Exhibitor Forum is the place to be to learn what is available today and tomorrow.

Exhibitor Forum

Tuesday, November 17

Software Tools and Libraries for C, C++ and C#

Chair: Lori Diachin (Lawrence Livermore National Laboratory)

10:30am-Noon

Room: E147-148

Vector C++: C++ and Vector Code Generation by Transformation

Presenter: Ira Baxter (Semantic Designs)

C++ is becoming popular for scientific computing, but does not address vector- and multi-core CPUs well. Subroutine libraries or C++ templates can provide some access to such facilities, but cannot generate high performance code because they do not account for synergistic interactions among sequences of parallel code, and do not generate native code that uses the vector facilities of modern CPUs. This talk will discuss VectorC++, a C++ dialect extended with a Vector datatype that relaxes the C++ memory model while preserving C++ semantics for the rest of the language. Vectors are declared explicitly, and can be processed by a wide variety of vector operators built directly into the VectorC++ language. The language is translated into conventional C, or into manufacturer-specific C++ compilers by a commercial transformation tool, DMS. Custom transformation rules generate target-specific vector machine instructions. The result provides high-level, portable C++ code that can generate low level, high-performance code.

A Methodology to Parallelize Code without Parallelization Obstacles

Presenter: Per Stenstrom (Nema Labs)

Parallelization of C/C++ legacy code is in general difficult, labor-intensive and error-prone. In particular, one must first identify the hot-spots in the algorithm, then identify what parts in the hot code that can run in parallel and finally, validate that the parallel code runs correctly and fast on multi-core platforms. Methodologies built on top of low-level parallelization APIs, such as OpenMP, make this process truly challenging. Nema Labs has developed a methodology allowing programmers to parallelize ordinary C/C++ code without reasoning about parallelism. The methodology embodied in our parallelization tools quickly guides the programmer to the hot code segments. Through a powerful auto-parallelization framework, the programmer is guided to make trivial changes to the sequential code to make it

parallelizable even in the presence of pointers. This talk will present the methodology along with use cases that show that the parallelization effort is reduced dramatically.

Parallelizing C# Numerical Algorithms for Improved Performance

Presenter: Edward Stewart (Visual Numerics, Inc.)

Parallel algorithms enable developers to create applications that take advantage of multi-core systems. Different technologies are available for platforms and languages to support parallelization efforts. For .NET developers, Microsoft provides parallel APIs and tools as part of Visual Studio 2010 and .NET Framework 4*. This presentation will discuss using these APIs and tools to enable parallelism in numerical algorithms in the IMSL C# Library. Adding parallel directives in IMSL C# Library algorithms had minimal impact on existing user code as the same algorithms were recently parallelized in the IMSL C Library and the technology approach was similar across languages. Benchmark results show that scalability results are also similar across native and managed code, a positive sign for developers who want the ease and flexibility of developing .NET applications and want to take advantage of multi-core performance. * These versions are in beta as of the July 2009 submission of this abstract.

Storage Solutions I

Chair: Jeffery A Kuehn (Oak Ridge National Laboratory)

10:30am-Noon

Room: E143-144

Panasas: pNFS, Solid State Disks and RoadRunner

Presenter: Garth Gibson (Panasas), Brent Welch (Panasas)

The most significant component of NFS v4.1 is the inclusion of Parallel NFS (pNFS) which represents the first major performance upgrade to NFS in over a decade. pNFS represents the standardization of parallel IO and allows clients to access storage directly and in parallel, eliminating the scalability and performance issues associated with typical NFS servers today. Solid State Disks (SSD) offer the potential to radically change the performance characteristics of Storage systems and accelerate application performance. However there are important trade-offs in reliability, management overhead and cost when implementing SSD technology. Panasas will detail the application benefits while improving reliability, eliminating management issues in the latest Series 9 ActiveStor systems. The RoadRunner SuperComputer at LANL was the first to achieve a sustained PetaFLOP and

continues to be the fastest system in production. Panasas will detail the ActiveStor system that provides the storage infrastructure for this system.

Solving the HPC I/O Bottleneck: Sun Lustre Storage System

Presenter: Sean Cochrane (Sun Microsystems)

With ongoing increases in CPU performance and the availability of multiple cores per socket, many clusters can now generate I/O loads that are challenging to deliver. Traditional shared file systems, such as NFS and CIFS, were not originally designed to scale to the performance levels required. Building a high performance cluster without addressing the need for sustainable high I/O can lead to sub optimal results. This presentation explores how you can use the Lustre File System and Sun Open Storage to address this challenge with a proven reference architecture.

Benefits of an Appliance Approach to Parallel File Systems

Presenter: Rick Friedman (Terascale)

Terascale has developed a unique appliance approach to delivering parallel file systems based storage solutions that are delivering up to 800% of the performance of NFS based solutions. Our appliances combine purpose-built hardware with Lustre™, an open-source, very high throughput parallel file system, and proprietary enhancements in performance, management, systems efficiency, reliability, and usability. We deliver true appliances that are simple to deploy, simple to manage and expand, and deliver consistent performance as your environment grows. This presentation will share our experiences of implementing an appliance architecture and how that has impacted our customers total system performance, uptime, and end user satisfaction. We will discuss the trade-offs and benefits of the appliance approach compared to end user built parallel storage installations and to on-site customized solutions.

HPC Architectures: Toward Exascale Computing

Chair: Charles Nietubicz (Army Research Laboratory)

1:30pm-3pm

Room: E143-144

Cray: Impelling Exascale Computing

Presenter: Steve Scott (Cray Inc.)

Having led the industry to petascale achievements, Cray continues its mission to introduce new innovations to high performance computing, and is now setting its sights on exascale computing. This talk will address the need for and issues involved in delivering exascale computing systems, and what that means for HPC customers.

Cray: Impelling Exascale Computing

Presenter: Chris Maher (IBM Corporation)

How will a million trillion calculations per second change the world? That's 1000 times the power of today's most advanced supercomputer. While the magnitude of an exascale supercomputer's power is difficult to fathom, the uses for such a system are already clear to IBM scientists as well as global leaders of businesses and governments. In this presentation, we will discuss the journey to exascale, from today's supercomputing and cloud systems, through partnerships and research projects, to the achievement of the next 'moon-shot' in the evolution of the world's most powerful computers - an 'exaflop' system with nearly unimaginable power.

Scalable Architecture for the Many-Core and Exascale Era

Presenter: Eng Lim Goh (SGI)

Our industry has moved from the era of regular clock speed improvements to the new many-core era. This is a fundamental enough change that even consumer PC users and applications will need to adapt. This talk focuses on the implications for productive high-performance and clustered computing architectures as we reach Exascale.

Tuesday, November 17

Software Tools for Multi-core, GPUs and FPGAs

Chair: Doug Post (DOD HPC Modernization Program)

1:30pm - 3pm

Room: E147-148

Acumem: Getting Multicore Efficiency

Presenter: Erik Hagersten (Acumem)

Efficient multicore applications are fundamental to green computing and high performance. This requires tuning for deep memory hierarchies and thread interaction in addition to parallelization. Actually, sometimes a well-tuned sequential program can outperform a parallelized program. This process requires expert knowledge and many weeks of data gathering and human analysis using conventional performance tools. Acumem's tools automatically analyze the efficiency of your applications and suggest necessary fixes to unleash multicore performance. What used to take experts weeks is now done in minutes at the click of a button. These language-agnostic tools do not require new languages or proprietary extensions. The resulting code increases performance across a wide range of architectures. In this session we will demonstrate the analysis and fix of popular open-source applications in minutes. We show repeated examples where parallelization alone completely misses the point. Performance improvements or energy savings of up to 30 times are demonstrated.

A Programming Language for a Heterogeneous Many-Core World

Presenter: Stefan Möhl (Mitrionics)

As the era of Multi-Cores moves into the era of Many-Cores, with early 2010 bringing 12 cores from AMD and 32 cores from Intel Larrabee, scalability issues that used to be the sole realm of HPC experts will now spread to the general programming public. Many of these programmers will not even know of Amdahl's Law. At the same time, accelerators like GPGPUs and FPGAs compound the complexity of the programming task further with Heterogeneous systems. When used to program the Mitrion Virtual Processor (MVP) in an FPGA, the inherently fine-grain MIMD parallel programming language Mitrion-C has shown to scale to 10's of thousands of parallel MVP processing elements. A new prototype compiler of Mitrion-C for Multi-Cores and Clusters has been developed, with a clear path for support of GPGPUs and Vector processors. Early scaling results show that

the same source code will scale efficiently on FPGAs, Multi-Cores and Clusters.

PGI Compilers for Heterogeneous Many-Core HPC Systems

Presenter: Michael Wolfe (Portland Group)

For the past decade, the top HPC systems were massively parallel clusters of commodity processors; current and future HPC systems increase performance and strong scaling by adding one or more highly parallel accelerators to each multicore node. We present PGI's roadmap to address the programming challenges for such systems. PGI Accelerator Fortran and C99 compilers offer a proven method to program host+accelerator (such as x64+GPU) platforms, using techniques employed by classical vectorizing and OpenMP compilers. Automatic compiler analysis, assisted and augmented by directives, enables users to generate efficient accelerator code without resorting to new languages or rewriting in new C++ class libraries. Compiler feedback allows the programmer to drive the optimization process and tune performance incrementally. These same techniques can be used to program massively multicore CPUs, supporting high degrees of parallelism, preserving portability across a wide range of system architectures, and increasing programmer productivity on heterogeneous manycore systems.

Networking I

Chair: Becky Verastegui (Oak Ridge National Laboratory)

3:30pm - 5pm

Room: E147-148

Managing the Data Stampede: Securing High Speed, High Volume Research Networks

Presenter: Joel Ebrahimi (Bivio Networks)

Academic and research organizations with immense, high-performance data processing needs, such as the National Labs, often find that traditional data security and management tools leave much to be desired when it comes to protecting and ensuring the integrity of their networks. Cutting edge scientific applications require robust data streaming to multiple organizations in real-time without sacrificing data security or network performance. In this presentation, Joel Ebrahimi of Bivio Networks will discuss the ever growing pool of bandwidth-intensive network traffic and the challenge of monitoring unauthorized activities in high-speed, high-volume network links and those within protocols intended to obscure the details of the information carried. By enabling applications like deep packet inspection, the next-gener-

ation of network platforms will deliver unprecedented security and optimal performance in the most demanding network environments for scientific and research projects.

Juniper Networks Showcases Breakthrough 100 Gigabit Ethernet Interface for T Series Routers

Presenter: Debbie Montano (Juniper Networks)

Juniper Networks, Inc., the leader in high-performance networking, introduced the industry's first 100 Gigabit Ethernet (100 GE) router interface card, which will be delivered on the T1600 Core Router. This presentation will discuss the capabilities, design objectives and system architecture of the T1600 Router, including the packet forwarding architecture, single-chassis switch fabric architecture and multi-chassis switch fabric architecture. The T1600 architectural components which will be covered include the physical interface card (PIC), which performs both physical and link-layer packet processing; the packet forwarding engine (PFE), which implements 100Gbps forwarding per slot; the switch fabric, which provides connectivity between the packet forwarding engines; and the routing engine, which executes JUNOS software and creates the routing tables which are downloaded in the lookup ASICs of each PFE.

Update on the Delivery of 100G Wavelength Connectivity

Presenter: Dimple Amin (Ciena)

While there is wide anticipation for 100G connectivity in support of research networks, achieving true 100G connectivity poses unique technical challenges. However, emerging standards are finally providing a clear framework for industry-wide execution, and we have recently observed initial implementations of 100G in public demonstrations. In this presentation, Dimple Amin, Ciena's Vice President, Products & Technology, will provide an overview and update on the development of 100G WAN transmission, as well as an update on the development and deployment of this technology in the WAN. Mr. Amin will showcase Ciena's latest product advances and market accomplishments with 100G, and give insight into the technology's future possibilities for researchers.

Storage Solutions II

Chair: John Grosh (Lawrence Livermore National Laboratory)

3:30pm - 5pm

Room: E143-144

Storage and Cloud Challenges

Presenter: Henry Newman (Instrumental Inc)

Our industry is facing many challenges, but one of the greatest is the lack of progress in advancing storage technology, and how they will impact data distribution needed for cloud computing and cloud archives. Some of the challenges the community is facing are disk and tape hard error rates, network limitations, security, and undetectable errors, just to name a few. This talk will explore some of these and other challenges and attempt to bound the problem. The goal of the talk is to provide the audience with an understanding of some of the challenges and what must be considered when designing large complex storage systems in the HPC environments.

Tape: Looking Ahead

Presenter: Molly Rector (Spectra Logic)

Almost behind the scenes during the past decade tape has been unobtrusively evolving into a robust, reliable, and capacious component of data protection, perfect for storing data in the quantities HPC sites must manage. Molly Rector, VP of Product Management, reviews how far tape has come, and what is in store for tape and its automation. As data backups increase in size, information about backups becomes available for harvest in tracking data growth, tape health, library functioning, and more. HPC administrators can look to increasingly sophisticated data sets from their tape automation systems about data backup and recovery, including tape health markers and tape use tracking.

Dynamic Storage Tiering: Increase Performance without Penalty

Presenter: Michael Kazar (Avere Systems)

In today's HPC environments, storage systems are difficult to configure for optimal performance, and the addition of Flash storage to the available media mix only makes the job more challenging. This session discusses how adding a self-managing tier of storage with dynamically chosen content to your architecture can provide significant performance improvements while simultaneously reducing both purchase and operating costs.

Top 500 Supercomputers

Chair: Hans Meuer (Prometeus GmbH)

5:30pm - 7pm

Room: PB253-254

Top 500 Supercomputers

Presenter: Hans Meuer (Prometeus GmbH)

Now in its 17th year, the TOP500 list of supercomputers serves as a "Who's Who" in the field of High Performance Computing (HPC). The TOP500 list was started in 1993 as a project to compile a list of the most powerful supercomputers in the world. It has evolved from a simple ranking system to a major source of information to analyze trends in HPC. The 34th TOP500 list will be published in November 2009 just in time for SC09. A detailed analysis of the TOP500 will be presented, with a discussion of the changes in the HPC marketplace during the past years. This special event is meant as an open forum for discussion and feedback between the TOP500 authors and the user community.

Wednesday, November 18

Software Tools: Scalable 4GL Environments

Chair: Stanley Ahalt (Ohio Supercomputer Center)

10:30am-Noon

Room: E143-144

MATLAB: The Parallel Technical Computing Environment

Presenter: Jos Martin (MathWorks)

MATLAB together with the Parallel Computing Toolbox provides a rich environment for parallel computing that is suitable for beginners and experienced alike. This environment encompasses many different areas such as parallel language constructs, parallel numeric algorithms, interaction with scheduling environments and low level MPI-like functionality. This presentation will introduce some of the features that enable simple parallelization of simulation, modelling and analysis codes.

Supercomputing Engine for Mathematica

Presenter: Dean E. Dauger (Dauger Research, Inc.)

The Supercomputing Engine for Mathematica enables Wolfram Research's Mathematica to be combined with the programming

paradigm of today's supercomputers. Our patent-pending approach takes advantage of Mathematica's Kernel-Front End infrastructure to parallelize it without its code being "aware" it is running in parallel. This adaptation of Mathematica to parallel computing closely follows the industry-standard Message-Passing Interface (MPI), as in modern supercomputers. After creating an "all-to-all" communication topology connecting Mathematica kernels on a cluster, this new technology supports low-level and collective MPI calls and a suite of high-level calls, all within the Mathematica programming environment. We present our technology's API, structure, and supporting technologies including new ways we leverage Linux. We also discuss the possibility of applying this parallelization technique to other applications that have a Kernel-Front End formalism.

<http://daugerresearch.com/pooch/mathematica/>

Solving Large Graph-Analytic Problems from Productivity Languages with Many Hardware Accelerators

Presenter: Steven P. Reinhardt (Interactive Supercomputing Inc.), John R. Gilbert (University of California, Santa Barbara), Viral B. Shah (Interactive Supercomputing Inc.)

Numerous problems in homeland security, biology, and social network analysis are well represented by graphs. The meteoric rise in the amount of data available often makes these graphs extremely large by traditional standards. Analysis of these graphs needs to be interactive and exploratory, as the best methods are not known in advance. We have built a suite of algorithms in the M language of MATLAB useful for very large graphs, including simple graph operations (e.g., union), graph traversal (minimum spanning tree), and dimensionality reduction (e.g., SVD and non-negative matrix factorization). By using the Star-P parallel platform, we have scaled these algorithms to graphs of nearly a terabyte, running on >100 cores to achieve interactivity, and accelerated them with graph kernels running on GPUs. This presentation will cover details of the Knowledge Discovery Suite, future plans, and several recent applications.

Storage Systems, Networking and Supercomputing Applications

Chair: Larry P Davis (DOD HPC Modernization Program)

10:30am-Noon

Room: E147-148

InfiniStor: Most Feature-Rich Cluster Storage System

Presenter: KyungSoo Kim (PSPACE Inc.)

InfiniStor is PSPACE Technologies new distributed cluster storage system for large scale storage service needs. In this talk, we will discuss what Cloud Scale Storage should be and needs. We identified five fundamental issues that must be considered and provide answers: Performance, Scalability, Fault Resilience & Self Healing, Advanced Features, and Cost Effectiveness. Then, we will briefly introduce the InfiniStor product: Its Architecture, Its Features, and Experiences from the initial deployment for the last 18 Months. In this introduction, we will try to answer how we approach to the five issues. Advanced Features that Cloud Scale Storage should have will be introduced. Current listing includes DeDuplications, Snapshot, Information Lifecycle Management, Hot-Contents Optimizations, Hybrid Disk Optimizations, Tiered Storage Optimizations, Remote Mirroring & Sync and more. Finally, we will briefly discuss difficulties our experiences during the development and introduce our roadmap for the next 18 Months.

Ethernet Data Center: Evolving to a Flat Network and a Single Switch

Presenter: Ori Aruj (Dune Networks)

The growth in the data center market coupled with the consolidation into large, sometimes mega, data centers, have promoted a large array of development in the market. The transition into large data centers and introduction of new Reliable Ethernet standards are driving forces in the network flattening. In this presentation we would like to explore the driving forces that promote the data center to be flat, as well as explore a number of ways to reduce layers and implement a flat network.



Smith Waterman Implementation for the SX2000 Reconfigurable Compute Platform

Presenter: Joe Hee (siXis Inc.)

The SX2000 Reconfigurable Computing Platform integrates 4 fully-interconnected compute nodes, each featuring an Altera Stratix IV EP4SGX230K FPGA with 2 Gbytes DDR3 memory,

four QSFP high-speed serial ports, and a 1.5 GHz Freescale MPC8536E PowerQuicc III processor with 512 Mbytes DDR3 memory. Available bandwidth between compute node pairs is 20 Gbps. siXis and Stone Ridge Technology are working together to implement the Smith Waterman algorithm to the SX2000. Smith Waterman is a key algorithm used in the field of Bioinformatics to find exact optimal local alignments of two DNA or protein sequences. Projected performance using linear extrapolation of resource ratios is on the order 400 billion cell updates per second.

HPC Architectures: Microprocessor and Cluster Technology

Chair: John Shalf (Lawrence Berkeley National Laboratory)

1:30pm - 3pm

Room: E143-144

Scaling Performance Forward with Intel Architecture Platforms in HPC

Presenter: William Magro (Intel Corporation)

With clusters now in the mainstream, HPC users are concerned not only about performance, but also about their productivity. Ease of development, ease of deployment, and longevity of parallel applications are all top of mind. Simplified deployment and maintenance of clusters is also a key factor in administrator and user productivity. Intel is investing in both hardware and software innovations to enable its customers to more easily create high performance parallel applications and to seamlessly scale those applications forward across multiple generations of multi-core platforms and clusters. In this session, I will review Intel's vision for scalable architecture, platforms, and applications and introduce Intel's latest technology, products, and programs designed to simplify not only the development of parallel applications and systems, but also their successful deployment.

AMD: Enabling the Path to Cost Effective Petaflop Systems

Presenter: Jeff Underhill (AMD)

Government, research and Industry are looking for affordable ways to build their Petaflop infrastructures moving forward. AMD's HPC strategy will offer the required computing density to make your Petaflop system affordable. In this session, AMD will describe the key components allowing you to build and protect your Petaflop systems.



Aurora Highlights: Green Petascale Performance

Presenter: Walter Rinaldi (Eurotech), Pierfrancesco Zuccato (Eurotech)

In this presentation we will describe Aurora, Eurotech's HPC offering: an innovative system, featuring Intel's latest generation processors and chipsets, liquid cooled in order to ensure high energy efficiency, at the highest possible packaging density. The current Aurora architecture is based on a dual socket Intel Xeon5500 processor/chipset based processing node, resulting in one 19-inch chassis capable of providing 3Tflops@12.5kW @0.5m3 or an 8-chassis rack 24Tflops@100kW@3.6m3. Such a high density is achieved with extensive usage of liquid cooling, carried out with metal plates coupled to all Aurora boards, having coolant flow inside them. Aurora modules have no moving parts, no attached modules, and are hot-replaceable, being connected with miniaturized couplings to the heat removal hydraulic infrastructure. Aurora interconnect consists of both QDR Infiniband for its switched network and a 60Gbs fully customizable, robust, low latency 3D Torus, implemented via a high density FPGA and short cabling.

Networking II

Chair: Jeff Graham (Air Force Research Laboratory)

1:30pm - 5pm

Room: E147-148

Network Automation: Advances in ROADM and GMPLS Control Plane Technology

Presenter: Jim Theodoras (ADVA Optical Networking)

Service providers, enterprises and R&E institutions are challenged to deliver more bandwidth with reduced effort and cost. Using ROADMs (Reconfigurable Optical Add/Drop Multiplexers), bandwidth providers can launch new services, alter networks, protect revenue streams and reduce truck rolls through remote management. Advanced ROADM technologies create non-blocking network nodes to send any wavelength to any port, any time. Known as colorless and directionless ROADMs, they enable networks that are scalable, robust and easy to operate. Meanwhile, PCE (Path Computation Element)-based control plane architectures allow networks enabled with GMPLS (Generalized Multiprotocol Label Switching) technology to efficiently operate at greater richness and scale than previously possible. Providing standards-based methods for controlling and distributing path computation functions allows operators to tailor control plane deployments to the emerging capabilities of next-gen optical networks. This presentation will explore in greater detail how these technologies are empowering organizations to do more for less within their existing networks.

Ethernet: *The* Converged Network

Presenter: Thomas Scheibe (Cisco Systems)

The Ethernet Alliance is showcasing a converged Ethernet Fabric at SC09. The demo highlights the ability to converge LAN, SAN and IPC data traffic over a single 10 Gigabit Ethernet fabric. Multiple companies are working together to show the interoperability of innovative technologies such as FCoE, DCB, iSCSI, and iWARP on a converged Ethernet Fabric. In this presentation, the Ethernet Alliance will discuss how the demonstration utilizes DCB features such as PFC and ETS for converged network management; FCoE and iSCSI traffic performance; and iWARP low latency performance.

More Performance with Less Hardware through Fabric Optimization

Presenter: Yaron Haviv (Voltaire)

To maximize application performance in a scale-out computing environment, it's not enough to simply deploy the latest, fastest hardware. With fabric optimization software, we now have the ability to minimize congestion, automatically allocate additional bandwidth and resources to high-demand applications, and ensure high throughput throughout the fabric. The result: higher levels of utilization and reduced overall latency means more performance can be achieved using less hardware and consuming less power. In this session you'll learn how to get the highest levels of performance possible from your hardware investment. See how fabric optimization software has enabled organizations to cut application latency in half and double the volume of critical jobs by addressing inefficiencies that were previously undetectable.

Infiniband, Memory and Cluster Technology

Chair: John Shalf (Lawrence Berkeley National Laboratory)

3:30pm - 5pm

Room: E143-144

Driving InfiniBand Technology to Petascale Computing and Beyond

Presenter: Michael Kagan (Mellanox Technologies), Gilad Shainer (Mellanox Technologies)

PetaScale and Exascale systems will span tens-of-thousands of nodes, all connected together via high-speed connectivity solutions. With the growing size of clusters and CPU cores per cluster node, the interconnect needs to provide all of the following features: highest bandwidth, lowest latency, multi-core linear scal-

ing, flexible communications capabilities, autonomic handling of data traffics, high reliability and advanced offload capabilities. InfiniBand has emerged to be the native choice for PetaScale clusters, and was chosen to be the connectivity solution for the first Petaflop system, and is being used for 40% of the world Top10 supercomputers and nearly 60% of the Top100 supercomputers (according to the TOP500 list). With the capabilities of QDR (40Gb/s) InfiniBand, including adaptive routing, congestion control, RDMA and quality of service, InfiniBand shows a strong roadmap towards ExaScale computing and beyond. The presentation will cover the latest InfiniBand technology, advanced offloading capabilities, and the plans for InfiniBand EDR solutions.

Meeting the Growing Demands for Memory Capacity and Available Bandwidth in Server and HPC Applications

Presenter: Sameer Kuppahalli (Inphi Corporation)

Enterprise High Performance Computing (HPC) servers are challenged to meet the growing demands of memory capacity and available bandwidth requirements of the intensive server and HPC applications. Memory technology is currently the weakest link. IT and Data-center administrators are forced to make trade-offs in capacity or bandwidth, compromising one for the other. Inphi will reveal a new technology that will significantly increase the server capacities while operating at higher frequencies. Virtualized servers that require massive amounts of memory and bandwidth will benefit, while the data centers benefit by increasing the utilization of their installed server based. This reduces the total cost of ownership (TCO) by reducing infrastructure procurement cost, as well as powering and cooling cost.

Open High Performance and High Availability Supercomputer

Presenter: John Lee (Appro International Inc.)

This presentation will focus on the value of energy efficient and scalable supercomputing clusters using dual quad data rate (QDR) Infiniband to combine high performance capacity computing with superior fault-tolerant capability computing. It will also cover the benefits of latest a redundant dual port 40Gb/s (QDR) InfiniBand technology and go over how an open and robust cluster management system that supports diskless configuration and network failover is important for the overall system to achieve maximum reliability, performance and high availability.

Parallel Programming and Visualization

Chair: Alice E. Koniges (Lawrence Berkeley National Laboratory)

Time: 3:30pm - 5pm

Room: E147-148

HPC and Parallel Computing at Microsoft

Presenter: Ryan Waite (Microsoft Corporation), Steve Teixeira (Microsoft Corporation), Alex Sutton (Microsoft Corporation), Keith Yedlin (Microsoft Corporation)

The importance of high performance computing and parallel computing is increasing across industries as organizations realize value through accurate and timely modeling, simulation, and collaboration. However, taking advantage of high performance and parallel computing is inherently complex and labor intensive. To address these issues, Microsoft is focusing on providing common development and management models that make parallel computing and HPC easier for developers, IT professionals, and end users while also integrating with existing systems. In this presentation we will discuss how solutions across Windows HPC Server 2008 R2, Visual Studio 2010, and .NET Framework 4.0 enable performance, ease of use, and efficiency in today's HPC and parallel computing environment.



VizSchema: A Unified Interface for Visualization of Scientific Data

Presenter: Svetlana Shasharina (Tech-X Corporation)

Different scientific applications use many different formats to store their data. Even if common, self-described data formats, such as HDF5 or NetCDF, are used, data organization (e.g. the structure and names of groups, datasets and attributes) differs between applications and experiments, which makes development of uniform visualization tools problematic. VizSchema is an effort to standardize metadata of common self-described data formats so that the entities needed to visualize the data can be identified and interpreted by visualization tools. These standards are expressed both as human-readable text, programmatically, and as an XML description. An HDF5 data reader and a plugin to the visualization tool VisIt implementing this standard has been developed. This plugin allows visualization of data from multiple applications that use notions of fields, particles, and geometries. The data that has been visualized comes from multiple domains: fusion and plasma physics simulations, accelerator physics, climate modeling, and nanotechnology.

Thursday, November 19

Grid Computing, Cyber Infrastructures and Benchmarking

Chair: Barbara Horner-Miller (Arctic Region Supercomputing Center)

10:30am-Noon

Room: E147-148

Bright Cluster Manager: Advanced Cluster Management Made Easy

Presenter: Martijn de Vries (Bright Computing Inc)

Setting up and managing a large cluster can be a challenging task without the right tools at hand. Bright Cluster Manager allows clusters to be deployed, used and managed with minimal effort without compromising on flexibility.

Bright Cluster Manager gives cluster administrators access to advanced features and extreme scalability while hiding the complexity that is typical for large clusters. In addition, Bright Cluster Manager provides complete and consistently integrated user environment that enables end-users to get their computations running as quickly and easily as possible.

In this session, various aspects of Bright Cluster Manager, as well as its technical foundations, will be discussed.

Contribution of Cyberinfrastructure to Economic Development in South Africa: Update on Developments

Presenter: Happy Sithole (Centre for High Performance Computing)

High Performance Computing (HPC) has lead to success in innovations around the world. It is also becoming evident that these tools have advanced to a stage where they complement the traditional scientific investigations. South Africa had realized this opportunity and assessed how its challenges could be addressed through HPC. A co-ordinated cyberinfrastructure development approach by South Africa will be presented. Furthermore, scientific applications and technological developments that have benefited from this intervention will be discussed.

Common Application Benchmarks on Current Hardware Platforms

Presenter: Bart Willems (Atipa Technologies)

Agile businesses want to be able to quickly adopt new technologies, whether it 's clustering or symmetric multiprocessor servers, to help them stay ahead of the competition. With today's fast changing commodity hardware platforms, adopting new tech-

nologies is associated with risks of instability or certain unknowns. Atipa Technologies, as a HPC vendor, constantly helps researchers benchmark their applications before and/or after their decision of procuring and delivering new clustering systems. Our engineers help clients test the changes against real-life workloads and then fine-tune the changes before putting them into production. We will be sharing our benchmarking experiences on some commonly used applications such as CHARMM, NamD2, Gaussian, Amber, LAMMPS, BLAST, VASP, GULP, GAMESS, HYCOM, AVUS. We will elaborate on our experience of how minor tweaking on the hardware can enhance performance, how you can test changes against real-life workloads, and then fine-tune the changes before putting them into production.

Virtualization and Cloud Computing

Chair: Allan Snively (San Diego Supercomputer Center)

10:30am-Noon

Room: E143-144

High-End Virtualization as a Key Enabler for the HPC Cloud or HPC as a Service

Presenter: Shai Fultheim (ScaleMP)

The IT industry's latest hot topic has been the emergence of cloud computing. The primary activity to date has been in the consumer (e.g. Google) and the enterprise space (Amazon EC2). The focus of this presentation is the potential of high-end virtualization and aggregation as a key enabler to deliver High Performance Computing services in the cloud. HPC used to be the domain of scientists and researchers, but it has increasingly gone mainstream, moving deeper into the enterprise and commercial marketplace. Leveraging cluster architectures and aggregation, multiple users and multiple applications can be dynamically served versus the current model of serving one application at a time. The key advantage of integrating aggregation into the cloud architecture is the flexibility and versatility gained. Cluster nodes can be dynamically reconfigured for jobs ranging from large memory jobs to multithreaded applications and a variety of programming models can be deployed (OPENMP, MPI, etc).

Managing HPC Clouds

Presenter: William Lu (Platform Computing)

Cloud computing is expected to transform the way IT departments operate, but the relevance of private clouds for HPC is not yet understood, much less how to apply the dynamic resource management, self-service, utility control and interoperability of

cloud infrastructures to HPC environments. For example, most computing capacity provided by public cloud vendors is delivered via VM, which may not be suitable for HPC applications. By exploring the evolution of clusters and grids, it's apparent that private and hybrid clouds can be a next generation solution for delivering more HPC capacity more cost effectively. This presentation will focus on how to build and manage private HPC cloud infrastructures, as well as examine potential hurdles. Attendees will learn about HPC cloud software components and capabilities and how to extend the capacity of a private HPC cloud by leveraging the public cloud.

Maximizing the Potential of Virtualization and the Cloud: How to Unlock the Traditional Storage Bottleneck

Presenter: Sam Grocott (Isilon Systems)

Gartner predicts virtualization will continue to be the most change-driving catalyst for IT infrastructure through 2013. Meanwhile, the next evolution of virtualization - cloud computing - continues to gain traction in HPC and enterprise environments for its ability to maximize resources with minimal investment. However, virtualization and cloud computing face a dire challenge - the bottleneck of traditional, "scale-up" SAN and NAS storage. In virtualized or cloud environments traditional storage yields a complicated mish-mash of storage volumes, which only grows in complexity as virtualized workloads increase, requiring additional costs for hardware and staff. Simply put, the full potential of virtualization and cloud computing cannot be realized until new, "scale-out" storage technologies are deployed to unlock the traditional storage bottleneck and deliver the promise of virtualization and cloud computing. Mr. Grocott will detail the evolution of virtualization and cloud computing, demonstrating how "scale-out" storage can maximize the potential virtualization and cloud computing.

GPUs and Software Tools for Heterogeneous Architectures

Chair: Bronis R. de Supinski (Lawrence Livermore National Laboratory)

1:30pm-3pm

Room: E147-148

Tesla: Fastest Processor Adoption in HPC History

Presenter: Sumit Gupta (NVIDIA)

Science continues to demand an increasing amount of computing performance, whether it be to study biochemical processes or to simulate advanced engineering systems. GPUs are an exciting

new development in the high-performance computing domain that offer a big jump in gigaflop performance and memory bandwidth. GPUs not only enable higher performance clusters, but also enable desktop computers to become much more powerful "supercomputers" that enable a new level of productivity at the desktop. In this talk, we will talk about the application domains that are seeing the big benefits with GPUs, talk about the latest developments using CUDA and OpenCL, the power efficiency benefits of GPUs, present insight into new tools such as a Fortran compiler, libraries, and products that are going to be announced soon, and also discuss some of the challenges of GPU computing.

Debugging the Future: GPUs and Petascale

Presenter: David Lecomber (Allinea Software)

To achieve results at Petascale and beyond, HPC developers are exploring a range of technologies - from hybrid GPU systems to homogeneous multi-core systems at unprecedented scale. This is leading to the most complex of software development challenges that can only be solved with help from developer tools such as debuggers. This talk presents Allinea's DDT debugging tool as a solution - revealing how it is bringing simplicity and performance to the task of debugging GPU and Petascale systems.

Developing Software for Heterogeneous and Accelerated Systems

Presenter: Chris Gottbrath (TotalView Technologies)

Heterogeneous clusters, with nodes powered by both conventional CPUs and computational accelerators (Cell and GPU processors), are an increasingly important part of the HPC environment. This trend introduces some significant challenges for scientists and engineers who choose to take advantage of these new architectures. Developers who are troubleshooting need tools that can show what is happening across their entire parallel application. This talk will review TotalView for Cell and introduce the new port of TotalView for Cuda. TotalView clearly represents the hierarchical memory model used in hybrid clusters and provides a coherent view of the many threads and many processes that make up such applications. The talk will also introduce a new TotalView feature to catch memory bounds errors, called Red Zones. It will conclude with a brief review of ReplayEngine, the productivity-enhancing add-on to TotalView. The latest version of ReplayEngine has greatly improved support for parallel programs.

HPC Architectures: Future Technologies and Systems

Chair: F. Ron Bailey (NASA Ames Research Center)

1:30pm-3pm

Room: E143-144

Convey's Hybrid-Core Computing: Breaking Through the Power/Performance Wall

Presenter: Tony Brewer (Convey Computer), Kirby Collins (Convey Computer), Steve Wallach (Convey Computer)

The need to feed compute-hungry HPC applications is colliding with the laws of physics and the flattening of commodity processor clock speeds. Adding more cores to processors and more racks to computer rooms is costly and inefficient. The solution: hybrid-core computing. Hybrid-core computing, pioneered by Convey Computer Corporation, combines a hardware-based coprocessor with application-specific instruction sets to deliver greater performance than a standard x86 server. The coprocessor instructions - supported by a general-purpose, ANSI-standard programming model - appear as extensions to the x86 instruction set, and share the same physical and virtual address space with the host x86 processor. This approach reduces or eliminates the programming hurdles that make GPUs and FPGA-based accelerators so difficult to use. Learn the latest on hybrid-core computing as Convey presents an update on the Convey products, some real-world customer success stories, and a look at industries that can benefit from hybrid-core computing.

Fujitsu's Technologies for Sustained Petascale Computing

Presenter: Motoi Okuda (Fujitsu)

Fujitsu is driving its R&D for targeting Petascale computing and is also deeply involved in Japan's Next-Generation Supercomputer Project with the technologies for Petascale computing. Effective use of multi/many-core CPUs and efficient massive parallel processing between nodes are key issues to realize Petascale computing. Fujitsu's new architecture called Integrated Multi-core Parallel ArChiTecture and new interconnect are solutions for these issues. Some of these technologies have been already implemented in FX1 and the presentation will cover their effect including real application performance results. The presentation will also give you solutions for other key issues of low-power consumption and effective cooling.

Next Generation High Performance Computer

Presenter: Rudolf Fischer (NEC Corporation)

For more than 20 years the NEC SX series of vector computers combines a consistent strategy with innovation. The balance between peak performance and memory bandwidth is the key differentiator of this product line. As NEC moves forward for next generation high performance computer, the vector technology will continue to be a key technology component for the HPC solutions. The presentation will focus on next generation HPC solutions from NEC. As recent successes have proven, efficiency is not only reflected in price/performance, but increasingly more in terms of floor space and power consumption per sustained application performance, resulting in most favorable total cost of ownership. NEC is committed to continue to innovate and develop leading edge HPC systems and to provide superior tools to enable unprecedented breakthroughs in science and engineering.

Technologies for Data and Computer Centers

Chair: John Grosh (Lawrence Livermore National Laboratory)

3:30pm-5pm

Room: E147-148

Effective Data Center Physical Infrastructure Management

Presenter: Jason Dudek (APC by Schneider Electric)

The rise in energy costs is a key imperative for implementing comprehensive and intelligent physical infrastructure management, to provide data center operators with critical information needed to design, monitor and operate their physical infrastructure. This Exhibitor Forum presentation will examine APC InfraStruXure® Central, Capacity Manager and Change Manager, the next evolution of APC's Data Center Physical Infrastructure (DCPI) Management Suite of applications, featuring an open and flexible architecture to integrate with existing IT and building management systems. The APC Management Pack for System Center Operations Manager 2007 enables integration with Microsoft's System Center Operations Manager 2007, giving data center managers the ability to view and manage physical infrastructure, including power, cooling, physical space and security, as well as providing a comprehensive view of the health of servers, applications, and clients within the data center.

Using Air and Water Cooled Miniature Loop Heat Pipes to Save Up to 50% in Cluster and Data Center Cooling Costs

Presenter: Stephen S. Fried (Microway)

Loop Heat Pipes (LHPs) are very low thermal resistance devices currently used to cool electronics in space vehicles. Using LHP technology, Microway eliminates the need for water chillers, air blowers and almost all of the unreliable, noisy, energy inefficient 1U fans found in rack mount chassis. Miniature LHPs can cool electric devices more efficiently than any other technology and work with loads up to 1000 Watts/cm² (a world record). This presentation describes their use to efficiently cool electronics in servers housed in data centers, other enclosures and even laboratory closets. This revolutionary design completely changes the approach to data center cooling, from thinking about it as a cooling method, to one whose servers produce water hot enough to go directly to a cooling tower or be used for regeneration.

**48V VR12 Solution for High Efficiency Data Centers**

Presenter: Stephen J. Oliver (VI Chip Inc.)

For system architects, there are traditional conversion approaches for power delivery. One example is converting from AC to 48 V, 48 V to 12 V and 12 V to 1 V; another converts AC to 12 V and 12 V to 1 V. The 48 V bus represents a 'safe' choice but with previous generation voltage regulator (VR) specifications, the extra conversion step to 12 V is required to feed VRx solutions. New architectures enable different approaches that can improve the efficiency of the power delivery. For example, direct 48 V to 1.xV conversion, coupled with new control techniques, allows an increase in system efficiency while meeting all the requirements of the new VR12 specification. This presentation will compare traditional approaches and new architecture examples of AC and 380 V DC input systems for processor, memory, and legacy loads, using efficiency as a benchmark.



Computing for a Changing World

SC09 Birds-of-a-Feather

Birds of a Feather (BOF) sessions have been a staple of SC conferences since they were still called Supercomputing conferences. SC09 continues this tradition with a full schedule of exciting, informal, interactive sessions devoted to special topics, including

- Future of HPC (e.g. exascale software, sustainability)
- Lists and Competitions (e.g. Top500, Green 500, HPC Challenge)
- Software Packages (e.g. MPICH, OSCAR, PBS, eclipse)
- Languages and Programming Models (e.g. OpenMP, MPI, PGAS)

- Government Initiatives and Groups (e.g. NSF HECURA, INCITE, European initiatives)

In addition, there are many other specialized topics, from fault tolerance to parallel I/O.

With 60 BOFs (selected from over 90 submission) over the course of three days, there is sure to be something of interest for everyone.

BOFs meet Tuesday and Wednesday at 12:15pm - 1:15pm and 5:30pm - 7pm, and Thursday at 12:15pm - 1:15pm.

Tuesday, November 17

2009 HPC Challenge Awards

12:15pm-1:15pm

Room: E145-146

2009 HPC Challenge Awards

Primary Session Leader: Jeremy Kepner (MIT Lincoln Laboratory)

Secondary Session Leader: Piotr Luszczek (University of Tennessee, Knoxville)

The 2009 HPC Challenge Awards is a competition with awards in two classes: Class 1 and 2. Class 1: Best Performance awards best run submitted to the HPC Challenge website. Since there are multiple tests, the term “best” is subjective. It has been decided by the committee that winners in four categories will be announced: HPL, Global-RandomAccess, EP-STREAM-Triad per system, and Global-FFT. Class 2: Most Elegant awards implementation of three or more of the HPC Challenge benchmarks with special emphasis being placed on: HPL, Global-RandomAccess, STREAM-Triad and Global-FFT. This award would be weighted 50% on performance and 50% on code elegance/clarity/size. Competition in Class 1 offers a rich view of contemporary supercomputers as they compete for supremacy not just in one category but in four. Class 2, on the other hand, offers a glimpse into high end programming technologies and the effectiveness of their implementation.

Blue Gene/P User Forum

12:15pm-1:15pm

Room: A103-104

Blue Gene/P User Forum

Primary Session Leader: Raymond Loy (Argonne National Laboratory)

Secondary Session Leader: Scott Parker (Argonne National Laboratory)

The Argonne Leadership Computing Facility (ALCF) is part of the U.S. Department of Energy's (DOE) effort to provide leadership-class computing resources to the scientific community to enable breakthrough science and engineering. ALCF is home to Intrepid, a Blue Gene/P system with 163,840 processors and a peak performance of 557 teraflops, as well as Surveyor, a smaller Blue Gene/P with 4,096 processors and peak of 13.9 teraflops. For this BoF, ALCF staff members with in-depth applications expertise on the Blue Gene/P platform will host an open forum discussion. We will share our experiences with various key appli-

cations. Anyone with interest in the platform is welcome, and users of Blue Gene, whether from ALCF or other sites, are encouraged to join us and share their experiences.

Breaking the Barriers to Parallelization at Mach Speed BoF

12:15pm-1:15pm

Room: D139-140

Breaking the Barriers to Parallelization at Mach Speed BoF

Primary Session Leader: Ronald W. Green (Intel Corporation)

Secondary Session Leader: Niraj Srivastava (Interactive Supercomputing Inc.)

With the proliferation of multicore processors and commodity clusters EVERYONE should be writing parallelized applications by now, shouldn't they? Outside of the HPC community there is a growing awareness that in the future all programs will need to be designed and developed to exploit parallelism. So why is there any hesitance? What are the barriers? Is it a lack of parallel programming education and training, a lack of software tools, operating system support, or is it just inertia from a juggernaut of serial algorithms and code dating back to the days of punched cards and paper tape? This session seeks to explore the diverse reasons behind decisions to not parallelize. Participants are invited to share their successful methods for reaching outside the HPC community to bring parallelization to a broader audience.

CIFTS: A Coordinated Infrastructure for Fault Tolerant Systems

12:15pm-1:15pm

Room: D137-138

CIFTS: A Coordinated Infrastructure for Fault Tolerant Systems

Primary Session Leader: Pete Beckman (Argonne National Laboratory)

Secondary Session Leader: Al Geist (Oak Ridge National Laboratory), Rinku Gupta (Argonne National Laboratory)

The Coordinated Infrastructure for Fault Tolerant Systems (CIFTS) initiative provides a standard framework, through the Fault Tolerance Backplane (FTB), where any component of the software stack can report or be notified of faults through a common interface - thus enabling coordinated fault tolerance and recovery. SC07 and SC08 saw an enthusiastic audience of

industry leaders, academia, and research participate in the CIFTS BOF. Expanding on our previous success, the objectives of this BOF are: 1. Discuss the experiences gained, challenges faced in comprehensive fault management on petascale leadership machines, and the impact of the CIFTS framework in this environment. Teams developing FTB-enabled software such as MPICH2, Open MPI, MVAPICH2, BLCR etc., will share their experiences. 2. Discuss the recent enhancements and planned developments for CIFTS and solicit audience feedback. 3. Bring together individuals responsible for exascale computing infrastructures, who have an interest in developing fault tolerance specifically for these environments

Developing and Teaching Courses in Computational Science

12:15pm-1:15pm

Room: D133-134

Developing and Teaching Courses in Computational Science

Primary Session Leader: Brad J. Armosky (Texas Advanced Computing Center)

The computational power and science potential of the new NSF-funded HPC resources are growing faster than the pool of talent who can use them. The Grand Challenges facing our world demand scientific discovery that will only be possible with new talent to apply these resources in industry and academia. During this 90-minute BoF, an expert panel will discuss their experiences and lead a dialogue focused on how colleges and universities can develop courses for students to apply computational science and advanced computing. Invited panelists include Jay Boisseau (TACC), Robert Panoff (Shodor and Blue Waters), and Steven Gordon (Ralph Regula School of Computational Science). The panel will share their lessons-learned, motivations, and future plans. We invite attendees from all disciplines and organizations to ask questions, make recommendations, and learn how to support their colleges and universities to implement courses. The panel will produce recommendations to share with the community.

Next Generation Scalable Adaptive Graphics Environment (SAGE) for Global Collaboration

12:15pm-1:15pm

Room: A107-108



Next Generation Scalable Adaptive Graphics Environment (SAGE) for Global Collaboration

Primary Session Leader: Jason Leigh (University of Illinois at Chicago)

Secondary Session Leader: Sachin Deshpande (Sharp Laboratories of America), Luc Renambot (University of Illinois at Chicago)

This session will discuss next generation SAGE (<http://www.evl.uic.edu/cavern/sage/index.php>) development by its global community of users. As globally distributed research teams work more closely to solve complex problems utilizing high-performance cyber-infrastructure, the need for high-resolution visualization is becoming more critical. An OptIPortal is an ultra-resolution visualization display instrument interconnected by optical networks that enables the creation of “cyber-mashups,” juxtapositions of data visualizations. OptIPortals have been identified as a crucial cyberinfrastructure technology. Early adopters embraced SAGE as a common middleware that works in concert with many visualization applications, enabling users to simultaneously stream and juxtapose multiple visualizations on OptIPortals leveraging high-speed networks. SAGE is currently a research prototype that must become a hardened technology. In this BoF, we will present technical details of SAGE middleware, including recent extensions and algorithm improvements. We will present and solicit feedback about the SAGE roadmap and hear about community-developed use cases and applications.

NSF Strategic Plan for a Comprehensive National CyberInfrastructure

12:15pm-1:15pm

Room: E141-142

NSF Strategic Plan for a Comprehensive National CyberInfrastructure

Primary Session Leader: Edward Seidel (National Science Foundation)

Secondary Session Leader: Jose Munoz (National Science Foundation),

Abani Patra (National Science Foundation), Manish Parashar

(National Science Foundation), Robert Pennington (National Science Foundation)

It is widely accepted that CI is playing an important role in transforming the nature of science and will be critical to addressing the grand challenges of the 21st century. However, it is also critical for CI to evolve beyond HPC into a holistic and comprehensive ecosystem that incorporates all aspects of research, education, workforce development, outreach, etc., and is both balanced and integrated. Such a CI should have a broad and transformative impact on both science and on society in general. To realize this bold CI vision, the NSF Advisory Committee on Cyberinfrastructure has constituted six task forces focusing on (1) Grand Challenge Communities, (2) HPC (Clouds, Grids), (3) Data and Visualization, (4) Software, (5) Campus Bridging, and (6) Education Workforce. The goal of this BoF is to present this CI vision and the task forces to the community, and to get community inputs on requirements and challenges.

Scalable Fault-Tolerant HPC Supercomputers

12:15pm-1:15pm

Room: D135-136

Scalable Fault-Tolerant HPC Supercomputers

Primary Session Leader: Maria McLaughlin (Appro International Inc.)

Secondary Session Leader: Gilad Shainer (Mellanox Technologies)

High-performance computing systems consist of thousand of nodes and ten of thousand of cores, all connected via high speed networking such as InfiniBand 40Gb/s. Future systems will include a higher number of nodes and cores, and the challenge to have them all available for long scientific simulation run time will increase. One of the solutions for this challenge is to add scalable fault-tolerance capability as an essential part of the HPC system architecture. The session will review scalable fault-tolerant archi-

tectures and examples of energy efficient and scalable supercomputing clusters using dual quad data rate (QDR) InfiniBand to combine capacity computing with network failover capabilities with the help of Programming languages such as MPI and a robust Linux cluster management package. The session will also discuss how fault-tolerance plays in the multi core systems and what are the required modification to sustain long scientific and engineering simulation on those systems.

Accelerating Discovery in Science and Engineering through Petascale Simulations and Analysis: The NSF PetaApps Program

5:30pm-7pm

Room: D133-134

Accelerating Discovery in Science and Engineering through Petascale Simulations and Analysis: The NSF PetaApps Program

Primary Session Leader: Abani Patra (National Science Foundation)

Secondary Session Leader: Manish Parashar (National Science Foundation)

It is anticipated that in the near future researchers will have access to HPC systems capable of delivering sustained performance in excess of one petaflop/s. Furthermore, highest end production systems are expected to consist of tens of thousands to a few hundred thousand processors, with each processor containing multiple cores, each core capable of executing multiple threads, and arithmetic units that support small vector instructions. Developing applications for these machines is a large challenge that is being addressed by the NSF PetaApps program. The NSF PetaApps program has grown to support over 40 projects that will enable researchers to capitalize on these emerging petascale-computing architectures and catalyze progress in science and engineering beyond the current state-of-the-art. In this BOF we intend to provide a forum for all engaged or interested in this activity to share successes and challenges so that the community can move forward to efficiently using these resources.

Art of Performance Tuning for CUDA and Manycore Architectures

5:30pm-7pm

Room: E141-142

Art of Performance Tuning for CUDA and Manycore Architectures

Primary Session Leader: Kevin Skadron (University of Virginia)

Secondary Session Leader: Paulius Micikevicius (NVIDIA), David Tarjan (NVIDIA Research)

High throughput architectures for HPC seem likely to emphasize many cores with deep multithreading, wide SIMD, and sophisticated memory hierarchies. GPUs present one example, and their high throughput has led a number of researchers to port computationally intensive applications to NVIDIA's CUDA architecture. This session will explore the art of performance tuning for CUDA. Topics will include profiling to identify bottlenecks, effective use of the GPU's memory hierarchy and DRAM interface to maximize bandwidth, data versus task parallelism, avoiding branch divergence, and effective use of native hardware functionality such as transcendentals and synchronization primitives to optimize CPU utilization. Many of the lessons learned in the context of CUDA are likely to apply to other manycore architectures used in HPC applications. About half the time will be spent in an organized presentation by experienced CUDA programmers and the other half in open discussion.

Data Curation

5:30pm-7pm

Room: D137-138

Data Curation

Primary Session Leader: Robert H. McDonald (Indiana University)

Secondary Session Leader: Stephen C. Simms (Indiana University), Chris Jordan (Texas Advanced Computing Center)

Management, utilization, and preservation of large datasets is a topic of growing importance to the scientific and HPC community. Increasingly more HPC centers are planning or implementing data curation services for certain classes of data collections. Many in the HPC community are interested in forming partnerships such as those within TeraGrid or through NSF funded DataNet partnerships, DataOne and Data Conservancy, in order to offer shared data curation responsibilities, replication services and improved data collection and management tools. How will these data curation partnerships interact? How will researchers

work across collaborative organizations to enable HPC access for their long-term curated data? What can the researcher expect in terms of support for data curation both short-term and long-term within these data curation partnerships? This BOF will discuss these issues along with topics related to sustainable data curation arising from expanded HPC partnerships working with humanities, social sciences, and cultural heritage collections.

European HPC and Grid Infrastructures

5:30pm-7pm

Room: D135-136



European HPC and Grid Infrastructures

Primary Session Leader: Hermann Lederer (Max Planck Society)

Secondary Session Leader: Dietmar Erwin (Forschungszentrum Juelich), Ludek Matyska (Masaryk University)

EGI, the European Grid Initiative, represents an effort to establish a sustainable European grid infrastructure. Its foundations are the National Grid Initiatives (NGIs) and the EGI Organization (EGI.eu). The first phase of EGI implementation is prepared to start in May 2010. DEISA is a consortium of the most powerful supercomputer centres in Europe, operating supercomputers in a distributed but integrated HPC infrastructure. It regularly hosts the most challenging European supercomputing projects and prepares a turnkey operational solution for a future integrated European HPC infrastructure. PRACE, the Partnership for Advanced Computing in Europe¹, prepared the creation of a persistent pan-European HPC infrastructure to support world-class science on world-class systems. PRACE will become operational in 2010 and deploy up to five leadership systems at renowned partner sites. For EEF, the recently formed European E-Infrastructure Forum will present its ideas on a future European compute ecosystem.

Low Latency Ethernet through New Concept of RDMA over Ethernet

5:30pm-7pm

Room: PB252

Low Latency Ethernet through New Concept of RDMA over Ethernet

Primary Session Leader: Gilad Shainer (Mellanox Technologies)

Latency is among the most critical capabilities the influence on application performance in high performance and beyond. For

years, Ethernet lagged behind other networking solution in providing real low latency solution. While many organization claim to achieve low latency Ethernet, in most cases it is still an order of magnitude higher than what is needed, or what other solution were able to provide. For enabling the true low latency Ethernet, new proposals now raise the idea of using different transport protocols on top of Ethernet, as such transports were proven to provide low latency and other important networking elements. The Session will review the various proposals that are being discussed in the different standardization bodies, and provide example of usage cases and simulations results.

Lustre, ZFS, and End-to-End Data Integrity

5:30pm-7pm

Room: E145-146

Lustre, ZFS, and End-to-End Data Integrity

Primary Session Leader: Hua Huang (Sun Microsystems)

Lustre kDMU uses Solaris ZFS as a new storage backend to gradually replace ldiskfs. The goal is to take advantage of the many advanced features ZFS provides including capacity and performance enhancements coupled with a data integrity architecture and simplified administrative interface. In 2008, an LLNL engineer successfully ported kernel-resident ZFS for Redhat Linux. Since then, the Lustre group has leveraged LLNL's work and is building Lustre on ZFS/DMU within a production environment. The project is refereed to as Lustre kDMU since the DMU (Data Management Unit) is the core ZFS component being introduced into Linux. In this BOF session, we will present the architecture for Lustre/ZFS integration and will demonstrate the data and metadata functionalities as well as end-to-end data integrity through an interactive user session. We will explore the performance comparison between ZFS and ldiskfs based Lustre solutions.

Micro-Threads and Exascale Systems

5:30pm-7pm

Room: PB251

Micro-Threads and Exascale Systems

Primary Session Leader: Loring Craymer (University of Southern California)

Secondary Session Leader: Andrew Lumsdaine (Indiana University)

As high-end computing systems head towards exascale, the old ways of dealing with parallelism--heroic programmers and super-human system architects--have become less and less effective. Locality management adds complexity; system and operational costs bar architectural innovation. Aided by DoD's ACS program, we have formed a community of interest focused on solving these problems. Our research covers new models of computation that support dynamic movement of data and computation, compilation strategies to reduce the burden on the programmer, and disciplined analysis approaches that reduce innovation risks for high-end computer architectures. We would like to invite other researchers to join our community, to raise awareness of our efforts, and to broaden our base of support in the applications and sponsor communities. If your research seems to fit, if your applications no longer fit your machine, or if you just feel frustrated at the state of the art, please join us.

MPI Acceleration in Hardware

5:30pm-7pm

Room: D139-140

MPI Acceleration in Hardware

Primary Session Leader: Karl Feind (SGI)

In this session, we bring together customers and users interested in MPI performance and scalability, to discuss and comment on hardware features in large scale systems aimed at accelerating MPI. MPI acceleration in hardware takes many forms: Bandwidth acceleration, Latency acceleration, MPI-2 one-sided fence acceleration and MPI Collective communication acceleration.

NSF High End Computing University Research Activity (HECURA)

5:30pm-7pm

Room: PB255

NSF High End Computing University Research Activity (HECURA)

Primary Session Leader: Almadena Chitchelkanova (National Science Foundation)

NSF program High-End Computing University Research Activity (HECURA) 2009 invited research and education proposals in the areas of I/O, file and storage systems design for efficient, high-throughput data storage, retrieval and management in cases where HEC systems comprise hundreds-of-thousands to millions of processors. This solicitation generated a lot of interest within HEC community. This BOF brings together Awardees of HECURA 2009 solicitation, general public, industry and national laboratories representatives. The goal of this BOF to discuss possible synergy between the projects, identify existing gaps and foster collaborations between the Awardees and the industry and National Laboratories. NSF program Directors will discuss the future of the HECURA Program to alert HEC community of this opportunity and to encourage a wider participation in the future HECURA competitions.

PGAS: The Partitioned Global Address Space Programming Model

5:30pm-7pm

Room: PB256

PGAS: The Partitioned Global Address Space Programming Model

Primary Session Leader: Tarek El-Ghazawi (George Washington University)

Secondary Session Leader: Lauren Smith (DOD), Michael Lewis (DOD)

PGAS, the Partitioned Global Address Space, programming model provides ease-of-use through a global shared address space while emphasizing performance through locality awareness. Over the past several years, the PGAS model has been gaining rising attention. A number of PGAS languages such as UPC, CAF and Titanium are widely available on high-performance computers. The DARPA HPCS program has also resulted in new promising PGAS languages, such as X10 and Chapel. This BoF will bring together developers from the various aforementioned PGAS lan-

guage groups and the research and development community for exchange of ideas and information. The BoF will be conducted in a panel format with one representative from each PGAS language development community as a panelist. The panelists will share the most recent developments in each one of those language paradigms (from specifications to tools and applications) and respond to questions of prospective as well as seasoned users.

pNFS: Parallel Storage Client and Server Development Panel Update

5:30pm-7pm

Room: E143-144

pNFS: Parallel Storage Client and Server Development Panel Update

Primary Session Leader: Joshua Konkle (NetApp)

This panel will appeal to Virtual Data Center Managers, Database Server administrators, and those that are seeking a fundamental understanding pNFS. This panel will cover the four key reasons to start working with NFSv4 today. Explain the storage layouts for parallel NFS; NFSv4.1 Files, Blocks and T10 OSD Objects. We'll engage the panel with a series of questions related to client technology, storage layouts, data management, virtualization, databases, geoscience, video production, finite element analysis, MPI-IO client/file locking integration, data intensive searching, etc. You'll have an opportunity to ask detailed questions about technology and panel participant plans from SNIA, NFS Special Interest Group, which is part of the Ethernet Storage Forum.

Productivity Tools for Multicore and Heterogeneous Systems

5:30pm-7pm

Room: A107-108

Productivity Tools for Multicore and Heterogeneous Systems

Primary Session Leader: Felix Wolf (Juelich Supercomputing Centre)

Secondary Session Leader: Allen Malony (University of Oregon), Valerie Taylor (Texas A&M University)

To meet the challenges of mapping application codes to petascale and potentially exascale computer systems in a manner that is correct and efficient, we would like to invite HPC users and tool developers to discuss requirements and priorities for productivity-

enhancing software tools. We will focus on the following topics: Resource Contention on Multicore Systems; What data are needed to determine whether resource contention is occurring on multi-core architectures?; How can various types of contention be detected, modeled, and evaluated through the use of tools? Tool Scalability; How can tools for petascale and exascale architectures be made reasonably easy to use and be made to scale in terms of the amount of data collected and the analysis time? Tools for Heterogeneous Architectures; How can applications be mapped efficiently to heterogeneous architectures that include accelerators?; What tools are needed to ensure correctness and measure performance on these architectures?

SLURM Community Meeting

5:30pm-7pm

Room: A103-104

SLURM Community Meeting

Primary Session Leader: Morris A. Jette (Lawrence Livermore National Laboratory)

Secondary Session Leader: Danny Auble (Lawrence Livermore National Laboratory)

SLURM development was begun at Lawrence Livermore National Laboratory in 2002 to provide a highly scalable and portable open source resource management facility. Today SLURM is used on about 40 percent of the TOP500 systems and provides a rich set of features including topology aware optimized resource allocation, the ability to power down idle nodes and restart them as needed, hierarchical bank accounts with fair-share job prioritization and many resource limits. Details about the recently released SLURM version 2.1 will be presented along with future plans. This will be followed by an open discussion. Everyone interested in SLURM use and/or development is encouraged to attend.

Top500 Supercomputers

5:30pm-7:30pm

Room: PB253-254

Top500 Supercomputers

Primary Session Leader: Hans Meuer (Prometeus GmbH)

The Top500 event is now a special Exhibitor Forum event. Details are provided in that section.

Users of EnSight Visualization Software

5:30pm-7pm

Room: E147-148



Users of EnSight Visualization Software

Primary Session Leader: Darin McKinnis (CEI)

Secondary Session Leader: Anders Grimsrud (CEI)

Purpose of the EnSight BOF is to give the EnSight user community a chance to meet and discuss needs, accomplishments, and anticipated problems. Current status of EnSight on Viewmaster for solutions run on Roadrunner at Los Alamos will be a featured presentation. Topics will include EnSight software and other utilities/viewers associated with it such as EnLiten, EnVideo, Reveal and EnVE. The particular needs of HPC users will be emphasized at this meeting. Other meetings being held in Germany, US, Japan, and elsewhere are focused on other non-HPC areas.

Wednesday, November 18

Benchmark Suite Construction for Multicore and Accelerator Architectures

12:15pm-1:15pm

Room: B119

Benchmark Suite Construction for Multicore and Accelerator Architectures

Primary Session Leader: Kevin Skadron (University of Virginia)

Benchmark suites help drive research but do not yet adequately support multicore and accelerator architectures. In particular, current suites do not consider the asymmetric (intra- vs. inter-chip) nature of locality in multicore architectures, features that are supported in hardware within a chip but only in software among chips, cores on the same chip with differing capabilities, or different programming models presented by various accelerators. New benchmarks are needed that are designed to take advantage of these features. This in turn helps expose how software can use and hardware can improve such features and helps identify new requirements. The objective of this session is to bring together researchers interested in benchmark design and users of benchmark suites in order to identify and prioritize research needs. The format will be a moderated discussion. The session's findings will be summarized in an online report.

Best Practices for Deploying Parallel File Systems

12:15pm-1:15pm

Room: D137-138

Best Practices for Deploying Parallel File Systems

Primary Session Leader: Rick Friedman (Terascale)

This session will focus on the challenges and rewards of implementing a parallel file system to improve cluster performance. As more clusters grow in size, increasingly users see that adding nodes does not result in any additional performance benefits. This is often caused by bottlenecks in the I/O system. Implementing a parallel file system involves a number of key decisions including which file system to use, whether to build your own platform or purchase a complete solution, how to configure your network, and planning for system deployment and management. We will discuss best practices and ask for audience participation to try to refine those best practices to help users understand how to leverage these kinds of technologies.

Building Parallel Applications using Microsoft's Parallel Computing Models, Tools, and Platforms

12:15pm-1:15pm

Room: A107-108

Building Parallel Applications using Microsoft's Parallel Computing Models, Tools, and Platforms

Primary Session Leader: Steve Teixeira (Microsoft Corporation)

Microsoft is creating an integrated platform and tools solution-stack that enables developers to build parallel applications that take advantage of architectures from multicore clients to large clusters. This Birds-of-a-Feather session discusses the use of the new Parallel Pattern Library and Task Parallel Library programming models on clients and individual nodes while using message passing interface (MPI) and Service Oriented Architectures (SOA) to distribute work over a cluster. We will also take a look at the new capabilities of Visual Studio to help developers more easily debug and optimize for multi-core. Come discuss the unique advantages afforded by this integrated stack, architectural challenges and best practices.

Deploying HPC and Cloud Computing Services for Interactive Simulation

12:15pm-1:15pm

Room: D133-134

Deploying HPC and Cloud Computing Services for Interactive Simulation

Primary Session Leader: Roger Smith (US Army)

Secondary Session Leader: Dave Pratt (SAIC)

The community of academia, industry, and government offices that are leading the development of new interactive simulations for training and analysis have reached a point at which the application of traditional networks of computing assets are no longer able to support simulation scenarios of sufficient scope, breadth, and fidelity. Several organizations are turning to various forms of HPCs, cloud computing architectures, and service-based software to create a computing platform that is powerful enough to run realistic models of military activities and very large scenarios. This BoF will discuss the problem-space and experiments that have been conducted in applying HPCs and cloud computing architectures to this domain. The BoF continues the discussions and community building from SC07 and SC08.

Developing Bioinformatics Applications with BioHDF

12:15pm-1:15pm

Room: D139-140



Developing Bioinformatics Applications with BioHDF

Primary Session Leader: Todd Smith (Geospiza)

HDF5 is an open-source technology suite for managing diverse, complex, high-volume data in heterogeneous computing and storage environments. The BioHDF project is investigating the use of HDF5 for working with very large scientific datasets. HDF5 provides a hierarchical data model, binary file format, and collection of APIs supporting data access. BioHDF will extend HDF5 to support DNA sequencing requirements. Initial prototyping of BioHDF has demonstrated clear benefits. Data can be compressed and indexing in BioHDF to reduce storage needs and enable very rapid (typically, few millisecond) random access into these sequence and alignment datasets, essentially independent of the overall HDF5 file size. Additional prototyping activities we have identified key architectural elements and tools that will form BioHDF. The BOF session will include a presentation of the current state of BioHDF and proposed implementations to encourage discussion of future directions.

Early Access to the Blue Waters Sustained Petascale System

12:15pm-1:15pm

Room: A103-104

Early Access to the Blue Waters Sustained Petascale System

Primary Session Leader: Robert A. Fiedler (National Center for Supercomputing Applications)

Secondary Session Leader: Robert Wilhelmsen (National Center for Supercomputing Applications)

We will describe the Blue Waters system and the PRAC program, which is the primary pathway to obtain an allocation through NSF in the first 1-2 years of operation. We will discuss our plan to support PRAC winners in preparing their applications for Blue Waters, including training, access to application performance simulators and early hardware, and the level of effort to we expect to devote to each application. We will also give potential users some tips on writing a winning proposal.

HPC Centers

12:15pm-1:15pm

Room: B118

HPC Centers

Primary Session Leader: Robert M. Whitten (Oak Ridge National Laboratory)

This BOF provides an open forum for user assistance personnel to discuss topics on interest. Topics include but are not limited to ticket procedures, queue policies, organization and structure of support options, and any pressing topic. This BoF is an ongoing meeting of the HPC Centers working group and all interested parties are encouraged to attend.

Network Measurement

12:15pm-1:15pm

Room: B117

Network Measurement

Primary Session Leader: Jon Dugan (ESnet)

Secondary Session Leader: Jeff Boote (Internet2)

Networks are critical to high performance computing, they play a crucial role inside compute systems, within the data center, to connect to remote resources and for communication with collaborators and remote sites. As a result it is an imperative that these networks perform optimally. In order to understand the behavior and performance of these networks they need to be measured. This is a forum for scientists who need better performance, network engineers supporting HPC sites and anyone interested network measurement particularly as it relates to high performance networking. This BoF will include presentations from various experts in network performance measurement and analysis as well as time for questions, discussion and impromptu presentations. Potential topics to include (but not limited to) measurement tools (Iperf, nuttcp, OWAMP, etc), measurement frameworks (perfSONAR, BWCTL, etc), emerging standards (NWMG, GGF standards, etc), and current research.

Open MPI Community Meeting

12:15pm-1:15pm

Room: E145-146

Open MPI Community Meeting

Primary Session Leader: Jeff Squyres (Cisco Systems)

Secondary Session Leader: George Bosilca (University of Tennessee, Knoxville)

MPI is increasingly being called upon to handle higher scale environments: "manycore," petascale, and faster networks. Years of experience in traditional HPC environments are leading to revolutions in leading-edge hardware platforms; software must evolve to match. The challenges are significant. As we proved last year by being the first MPI implementation to achieve a petaflop, we believe that a diverse open source team representing many different backgrounds, philosophies, and biases (from both industry and academia) is uniquely equipped to meet these challenges. Come hear where Open MPI is going, and how you can (and should!) join our efforts. The meeting will consist of three parts: (1) Members of the Open MPI core development team will present our current status; (2) Discussions of ongoing and

future projects; and (3) Discuss the Open MPI roadmap and actively soliciting feedback from real-world MPI users and ISVs with MPI-based products (please bring your suggestions!).

Practical HPC Considerations for Advanced CFD

12:15pm-1:15pm

Room: E141-142

Practical HPC Considerations for Advanced CFD

Primary Session Leader: Stan Posey (Panasas)

Secondary Session Leader: James Leylek (Clemson University)

Parallel efficiency and overall simulation turn-around time continues to be a determining factor behind scientific and engineering decisions to develop CFD models at higher fidelity. While the ROI in recent years of CFD verses physical experimentation has been remarkable, the maturity of parallel CFD solvers and availability of inexpensive scalable HPC clusters has not been enough to advance most CFD practice beyond steady state modeling. Fluids are inherently unsteady, yet to model such complexity with a URANS, DES, LES or other advanced approach at a practical HPC scale for industry, requires parallel efficiency of both solvers and I/O. This BoF will examine the current state of parallel CFD for HPC cluster environments at industry-scale. BOF panel members will comprise parallel CFD experts from commercial CFD software developers ANSYS and CD-adapco, and HPC practitioners from industry who deploy both commercial and research-developed CFD application software.

Trends and Directions in Workload and Resource Management using PBS

12:15pm-1:15pm

Room: D135-136

Trends and Directions in Workload and Resource Management using PBS

Primary Session Leader: Bill Nitzberg (Altair Engineering)

Secondary Session Leader: Bill Webster (Altair Engineering)

Today, doing more with less is simply business as usual. New and complex applications, combined with dynamic and unpredictable workloads make harder than ever to balance the demands of HPC users, while taking advantage of highly scalable "computing clouds." First developed at NASA Ames Research Center nearly 20 years ago, PBS is the most widely deployed workload and resource management software used today any-

where in the world. Companies and institutions of all sizes use PBS to ensure their workloads run efficiently and reliably, on every type of HPC platform possible. This Birds of a Feather session will bring together users, administrators and developer to share their experiences and expertise, as well as having an opportunity to give candid feedback about the direction of PBS to solve tomorrow's HPC challenges.

Campus Champions: Your Road to Free HPC Resources

5:30pm-7pm

Room: E141-142

Campus Champions: Your Road to Free HPC Resources

Primary Session Leader: Kay Hunt (Purdue University)

Since its formation in early 2008, the TeraGrid Campus Champion program has been a primary conduit for the recruitment of new user communities across 43 U.S. campuses, thus providing a return on the NSF's multimillion dollar investment in national cyberinfrastructure. The goal is to build on this success and expand to campuses in every state. Continuing the discussion started at SC08, this BoF will talk about the Campus Champion program and provide a focal point for communication and discussion of best practices, challenges, and opportunities to improve services and communications for faculty and staff that are engaging in critical scientific research. Campus level knowledge of national "free" HPC resources will empower significantly larger numbers of researchers and educators to advance scientific discovery, as well as engaging traditionally under-represented communities in becoming HPC users. Campus Champions from three sites will be leading this discussion among current and potential campuses.

Can OpenCL Save HPC?

5:30pm-7pm

Room: E145-146

Can OpenCL Save HPC?

Primary Session Leader: Ben Bergen (Los Alamos National Laboratory)

With the number of exotic computing architectures on the rise---not to mention multicore and hybrid platforms, combined with the fact that these "accelerated" solutions might just offer the best bang for the buck, what chance does the average computational scientist have to write portable, sustainable code? Enter OpenCL, a framework expressly designed to address these issues. In this

BoF, we will answer the questions, “What is it?,” “How do I use it?,” “How does it work for HPC?,” and “Where can I get it?.” In addition, there will be a lighthearted and lively panel discussion/debate on the direction and potential for this emerging standard.

Communicating Virtual Science

5:30pm-7pm

Room: E147-148

Communicating Virtual Science

Primary Session Leader: Aaron Dubrow (Texas Advanced Computing Center)

The conduct of computational science is an esoteric activity, but the results of that research are of interest and importance to everyone. “Communicating Virtual Science” will address the myriad ways writers communicate complex ideas about math, computer programming and discipline-specific research to lay and scientific audiences. During the BoF, a diverse panel of writers, editors, and academics will discuss approaches to science communication and ways to make computational research more broadly appealing and relevant. The talk will also address new media trends to which writers are responding and what's at stake if we fail to make the case for science as a societal imperative. With major media outlets cutting science journalists (and whole departments), it is becoming increasingly important for the HPC community to find new ways to communicate science breakthroughs to the public. Through this BoF, experts will debate which communication models are most suitable to the changing media environment.

Eclipse Parallel Tools Platform

5:30pm-7pm

Room: D137-138

Eclipse Parallel Tools Platform

Primary Session Leader: Beth R. Tibbitts (IBM)

Secondary Session Leader: Greg Watson (IBM)

The Eclipse Parallel Tools Platform (PTP, <http://eclipse.org/ptp>) is an open-source project providing a robust, extensible platform for parallel application development. PTP includes a parallel runtime, parallel debugger, remote development tools, and static analysis tools. It supports development on C/C++ and Fortran. PTP is the basis for the NCSA Blue Waters application development workbench. The BoF will consist of brief demos and dis-

cussions about PTP features, including: - PTP support for job schedulers/resource managers - Remote services, including editing, launching, and debugging on remote targets - Performance tools and the External Tools Framework (ETFW), which integrates existing command-line tools into Eclipse - Static and dynamic analyses for MPI programs - Fortran development and refactoring support via the Eclipse Photran project - The NCSA Blue Waters workbench, based on PTP The discussion will also be designed to find possible collaborators and directions for future development.

FAST-OS

5:30pm-7pm

Room: B118

FAST-OS

Primary Session Leader: Ronald G. Minnich (Sandia National Laboratories)

Secondary Session Leader: Eric Van Hensbergen (IBM)

BoF is to allow the DOE FAST-OS community to catch up with each other and their work, and to communicate results. Demos are most welcome, i.e. demonstrations of new operating systems booting on HPC hardware, or new OS capabilities being made available.

HPC Advisory Council Initiative

5:30pm-7pm

Room: PB252

HPC Advisory Council Initiative

Primary Session Leader: Gilad Shainer (HPC Advisory Council)

Secondary Session Leader: Jeffery Layton (Dell Inc.)

The HPC Advisory Council is a distinguished body representing the high-performance computing ecosystem that includes more than 80 members worldwide from OEMs, strategic technology suppliers, ISVs and selected end-users across the entire range of the HPC segments. The Council was to accelerate HPC innovations and new technologies, and to optimize system performance, efficiency and scalability in both hardware and software. The Council also collaborates to extend the reach of HPC into new segments, to meet current and future end-user requirements and to bridge the gap between HPC usage and its TMs potential. The HPC Advisory Council operates a centralized support center providing end users with easy access to leading edge HPC systems for development and benchmarking, and a support/

advisory group for consultations and technical support. The session will introduce the new council initiatives such as HPC in cloud computing, best practices for high performance applications future plans.

International Exascale Software Program

5:30pm-7pm

Room: PB256

International Exascale Software Program

Primary Session Leader: Jack Dongarra (University of Tennessee, Knoxville)

Secondary Session Leader: Pete Beckman (Argonne National Laboratory)

Over the last twenty years, the international open source community has increasingly provided the software at the heart of the world's high-performance computing systems. The community provides everything from operating system components to compilers and advanced math libraries. As an international community, however, we have only loosely coordinated activities and plans for development. The new rapidly changing technologies in multicore, power consumption, GPGPUs, and memory architectures creates an opportunity for the community to work together and build an international program to design, build, and deliver the software so critical the science goals of our institutions. To help plan how the international community could build a partnership to provide the next generation of HPC software to support scientific discovery, we have had a series of workshops. In this BoF, we report on the effort and solicit feedback from the community.

iPlant Collaborative: Computational Scaling Challenges in Plant Biology

5:30pm-7pm

Room: B117



iPlant Collaborative: Computational Scaling Challenges in Plant Biology

Primary Session Leader: Dan Stanzione (Texas Advanced Computing Center)

Secondary Session Leader: Martha Narro (University of Arizona), Steve Goff (University of Arizona)

The Plant Science Cyberinfrastructure Collaborative (PSCIC) program is intended to create a new type of organization: a cyberinfrastructure collaborative for the plant sciences - that

would enable new conceptual advances through integrative, computational thinking. To achieve this goal, we have developed the iPlant Collaborative (iPC). The iPC will utilize new cyberinfrastructure solutions to address an evolving array of grand challenges in the plant sciences. In the past year, iPlant has undertaken two new community-driven grand challenges: The iPlant Tree of Life, an attempt to build the tree of all green plant species on Earth, and the iPlant Genotype-Phenotype project, a comprehensive attempt to understand the mapping between Genetic information and expressed characteristics. This session will provide a forum for members of the HPC community to learn more about the iPC, and how to become involved. We will particularly focus on the barriers to computational scalability in plant biology.

MPI Forum: Preview of the MPI 3 Standard (Comment Session)

5:30pm-7pm

Room: D135-136

MPI Forum: Preview of the MPI 3 Standard (Comment Session)

Primary Session Leader: Rich Graham (Oak Ridge National Laboratory)

Secondary Session Leader: George Bosilca (University of Tennessee, Knoxville)

With the release of the Message Passing Interface (MPI) standard version 2.2 (www.mpi-forum.org) the standard has been consolidated into a single, more coherent document, numerous corrections and clarifications have been made, and the standard has been updated to support the current versions of the C, C++ (C99), and Fortran (2003) standards. The work of this standards body has shifted to focusing on proposed changes needed to fill functionality and scalability gaps. This includes the addition of support for nonblocking collectives, support for fault-tolerance within the standard, improved support for remote memory operations, improved support in threaded environments, and standardized hooks for tool access to information internal to MPI implementations.

Network for Earthquake Engineering Simulation (NEES): Open Invitation to Build Bridges with Related Virtual Organizations

5:30pm-7pm

Room: E143-144

Network for Earthquake Engineering Simulation (NEES): Open Invitation to Build Bridges with Related Virtual Organizations

Primary Session Leader: Sriram Krishnan (San Diego Supercomputer Center)

Secondary Session Leader: Steven McCabe (NEES Consortium, Inc)

NEES, the Network for Earthquake Engineering Simulation includes 15 Earthquake Research sites, connected through a US-wide cyberinfrastructure. The cyberinfrastructure represents a large virtual organization. Its tasks include collecting experimental data gathered at the diverse facilities, making the data available to the larger community, facilitating simulation experiments, supporting education, outreach, and training as well as providing a host of web-enabled services in support of the NEES research facilities and the community. While the purpose of NEES is unique, other virtual organizations face similar issues in gathering and managing data, providing simulation support, and facilitating the collaboration of the members of the involved organizations. Discussing such common needs and establishing links for potential coordination could create significant synergy with other organizations and save cost. The BoF will provide the grounds for such discussions and the creation of synergy of NEES with virtual organization that face similar issues.

OpenMP: Evolving in an Age of Extreme Parallelism

5:30pm-7pm

Room: PB255

OpenMP: Evolving in an Age of Extreme Parallelism

Primary Session Leader: Larry Meadows (Intel Corporation)

Secondary Session Leader: Barbara Chapman (University of Houston), Nawal Copti (Sun Microsystems)

Several implementations of the OpenMP 3.0 specification have become available since it was released in 2008. While initial experiences indicate that the tasking extensions included in the 3.0 specification significantly expand the applicability of OpenMP, they also demonstrate that some additional extensions would be useful. In this BoF, members of the OpenMP lan-

guage committee will discuss plans for refinements to the 3.0 specification in the short term, as well as plans for OpenMP 4.0. In addition, the BoF will include a lively panel discussion of whether (and how) OpenMP should adapt to heterogeneous platforms (such as GPGPU, CPU/LRB, or CPU/Cell combinations). The BoF will offer ample opportunity for questions and audience participation.

OSCAR Community Meeting

5:30pm-7pm

Room: B119

OSCAR Community Meeting

Primary Session Leader: Stephen L. Scott (Oak Ridge National Laboratory)

Secondary Session Leader: Geoffroy Vallee (Oak Ridge National Laboratory)

Since the first public release in 2001, there have been well over 200,000 downloads of the Open Source Cluster Application Resources (OSCAR) software stack. OSCAR is a self-extracting cluster configuration, installation, maintenance, and operation suite consisting of "best known practices" for cluster computing. OSCAR has been used on highly ranking clusters in the TOP500 list and is available in both a freely downloadable version as well as commercially supported instantiations. The OSCAR team is comprised of an international group of developers from research laboratories, universities, and industry cooperating in the open source effort. As it has for the past seven years at SC, the OSCAR BoF will be a focal point for the OSCAR community at SC09, where both developers and users may gather to discuss the "current state" as well as future directions for the OSCAR software stack. New and potential users and developers are welcome.

Python for High Performance and Scientific Computing

5:30pm-7pm

Room: A103-104

Python for High Performance and Scientific Computing

Primary Session Leader: Andreas Schreiber (German Aerospace Center)

Secondary Session Leader: William R. Scullin (Argonne National Laboratory), Steven Brandt (Louisiana State University), James B. Snyder (Northwestern University), Nichols A. Romero (Argonne National Laboratory)

The Python for High Performance and Scientific Computing BOF is intended to provide current and potential Python users and tool providers in the high performance and scientific computing communities a forum to talk about their current projects; ask questions of experts; explore methodologies; delve into issues with the language, modules, tools, and libraries; build community; and discuss the path forward.

Simplify Your Data Parallelization Woes with Ct: C++ for Throughput Computing

5:30pm-7pm

Room: D133-134

Simplify Your Data Parallelization Woes with Ct: C++ for Throughput Computing

Primary Session Leader: Rita E. Turkowski (Intel Corporation)

Secondary Session Leader: Sanjay Goil (Intel Corporation)

Intel offers comprehensive software optimization development tools for Windows Visual Studio developers. Now our suite of tools is broadening to address data parallelism with a new high level CPU/GPU compute programming model. A product based on the Ct programming model, now in beta in Intel's Developer Products Division, will be the focus of this BoF. We'll introduce C++ programmers to the benefits of applying the Ct programming model to their data parallel code. Come hear how to develop highly-parallelized and scalable software that takes advantage of Intel's current multi-core (which will scale for future many-core) processors. We hope to engage in a conversation with BoF attendees on how Ct can be applied to assist programmers in creating productive and performance sensitive parallelized future-proof applications. We'll also discuss real world parallelization challenges via examples from a variety of performance sensitive applications.

Solving Interconnect Bottlenecks with Low Cost Optical Technologies

5:30pm-7pm

Room: D139-140

Solving Interconnect Bottlenecks with Low Cost Optical Technologies

Primary Session Leader: Marek Tlalka (Luxtera)

As the performance in computer systems increases, so does the demand for bandwidth intensive services. The challenge becomes bottlenecks caused by information transfers and finding a high

performance solution at low cost. Current connectivity options include copper cables, copper active cables, and optical modules but these solutions struggle with performance limitations such as reach-constraints and size. Silicon Photonics enabled active optical cables are an alternative that provides the advantages of fiber optics and copper cabling offering high performance at a cost associated with copper. This session will cover distinct approaches that address connectivity problems by displacing conventional copper with optical interconnects to meet future bandwidth demands. Recognizing that the market for bandwidth intensive services continues to grow, it will cover optical technologies that can improve datacenter connectivity, operational performance, overall capital expenditures as well as discuss benefits and trade-offs.

Update on OpenFabrics Software (OFED) for Linux and Windows Latest Releases

5:30pm-7pm

Room: PB251

Update on OpenFabrics Software (OFED) for Linux and Windows Latest Releases

Primary Session Leader: Bill Boas (OpenFabrics Alliance)

This session enables OpenFabrics users to provide feedback, learn the latest information and meet with the developers who create and release the code. We will describe the capabilities of the latest releases for Linux and Windows planned for the fall of 2009. OpenFabrics Software (OFS) is a unified, transport independent, open-source software stack for RDMA-capable fabrics including InfiniBand and the new features of Ethernet, known by the IEEE as Data Center Bridging. These features are intended to support Ethernet as lossless, flow controlled and packet prioritized. New features in OFS include support for multi-core, Data Center Bridging, NFSoRDMA, improved RDS & SDP for sockets, additional MPI options, etc. Operating systems support includes Linux kernel 2.6.29 and 2.6.30, and the latest RHEL and SLSE releases, Open SUSE, Fedora Core, Ubuntu, CentOS and Windows. OFS developers present from the following companies will include: Cisco, IBM, Intel, Mellanox, Novell, Qlogic, RedHat and Voltaire.

What Programs *Really* Work for Students Interested in Research and Computing?

5:30pm-7pm

Room: A107-108

What Programs *Really* Work for Students Interested in Research and Computing?

Primary Session Leader: Roscoe Giles (Boston University)

Let's address the disconnect between student needs for career goals, and the actual resources available as students, what do you need? As resource providers, what are your interests? What can work against quality programs that keep them from being used to their full capacity? To put a point to this, some students have limited access to resources; some may have dedicated faculty but no resources. How do we bridge the students and the resources available? The Empowering Leadership Alliance will orchestrate this session as part of its goal to determine the best approaches that make a difference for minority scholars.

Thursday, November 19

Energy Efficient High Performance Computing Working Group

12:15pm-1:15pm

Room: D139-140



Energy Efficient High Performance Computing Working Group

Primary Session Leader: William Tschudi (Lawrence Berkeley National Laboratory)

Secondary Session Leader: Michael Patterson (Intel Corporation)
This BoF will bring together HPC professionals that volunteered to join an Energy Efficiency HPC users group. This began as a BOF at SC08, with approximately 45 joining the group. The group has prioritized a number of topical areas to advance such as HPC metrics, HPC best practices, etc. with an overall goal of dramatically improving overall energy performance while maintaining high computational ability. The purpose of the BoF will be to further define the overall direction of the group and brainstorm additional areas where a unified HPC community can make significant advancement in technologies to improve energy performance. Sub committees will be formed to focus on specific topics and the meeting will provide the user group a venue to meet their counterparts and begin jointly developing an action plan.

Extending Global Arrays to Future Architectures

12:15pm-1:15pm

Room: B118

Extending Global Arrays to Future Architectures

Primary Session Leader: Bruce Palmer (Pacific Northwest National Laboratory)

Secondary Session Leader: Manojkumar Krishnan (Pacific Northwest National Laboratory), Sriram Krishnamoorthy (Pacific Northwest National Laboratory)

The purpose of this BoF is to obtain input from the Global Array user community on proposed development of the GA toolkit. This session is intended to cap a planning process that began in early summer and will provide GA users from the broader HPC community with an opportunity to discuss their needs with the GA development team. The main focus will be on extending the GA programming model to post-petascale architectures but other topics, including the addition of desirable features relevant for programming on existing platforms will also be entertained. The format will be informal discussion. The session leaders will provide a brief overview of issues associated with extending the GA programming model to the next generation of computers and current plans to deal with them and will then open up the discussion to session participants.

Getting Started with Institution-Wide Support for Supercomputing

12:15pm-1:15pm

Room: D133-134

Getting Started with Institution-Wide Support for Supercomputing

Primary Session Leader: David Stack (University of Wisconsin-Milwaukee)

Not every research team has the expertise, or desire, to manage their own supercomputing resources. As the need for high performance computing expands into less technical fields, increasing numbers of less savvy researchers are turning to their institutions for support for hardware, operating systems and applications. They are also looking for assistance in parallelizing and using applications. However, it is difficult to identify and dedicate institutional resources to build an institution-wide support structure for high performance computing in lean budget times. Attendees will be asked to share their challenges and successes in creating and funding organizations that have a mission to support supercomputing across their institution.

Green500 List

12:15pm-1:15pm

Room: A107-108



Green500 List

Primary Session Leader: Wu Feng (Virginia Tech)

Secondary Session Leader: Kirk Cameron (Virginia Tech)

The Green500 List, now entering its third year, seeks to encourage sustainable supercomputing by raising awareness to the energy efficiency of such systems. Since its official launch at SC07, the list has continued to evolve to serve the HPC community. This BoF will present: (1) an overview of the challenges of tracking energy efficiency in HPC; (2) new metrics and methodologies for measuring the energy efficiency of a HPC system; (3) highlights from the latest Green500 List; and (4) trends within the Green500. In addition, we will actively solicit constructive feedback from the HPC community to improve the impact of the Green500 on efficient high-performance system design. We will close with an awards presentation that will recognize the most energy-efficient supercomputers in the world.

HDF5: State of the Union

12:15pm -1:15pm

Room: D137-138

HDF5: State of the Union

Primary Session Leader: Quincey Kozioł (HDF Group)

Secondary Session Leader: Ruth Aydt (HDF Group)

HDF5 is an open-source technology suite for managing diverse, complex, high-volume data in heterogeneous computing and storage environments. HDF5 includes: (1) a versatile self-describing data model that can represent very complex data objects and relationships, and a wide variety of metadata; (2) a completely portable binary file format with no limit on the number or size of data objects; (3) a software library with time and space optimization features for reading and writing HDF5 data; and (4) tools for managing, manipulating, viewing, and analyzing data in HDF5 files. This session will provide a forum for HDF5 developers and users to interact. HDF5 developers will describe the current status of HDF5 and present future plans. Ample time will be allowed for questions and discussion, and users of HDF5 technologies will be encouraged to share their successes, challenges, and requests.

HPC Saving the Planet, One Ton of CO₂ at a Time

12:15pm-1:15pm

Room: E141-142



HPC Saving the Planet, One Ton of CO₂ at a Time

Primary Session Leader: Natalie Bates (self-employed)

Without dispute, the amount of CO₂ in our atmosphere today has been rapidly rising. Pre-industrial levels of carbon dioxide were about 280 ppmv and current levels are greater than 380 ppmv. This was caused by combustion of coal, oil and gas; the primary energy sources fueling industrialization. This BOF will explore the net energy impact of computing and information technologies. While these technologies consume energy, they also enable productivity enhancements and directly contribute to energy efficiency. Some even argue that computing and information technologies are vital, significant and critical in moving towards a low-carbon future. One report suggests that the ICT sector "could deliver approximately 7.8 Gt CO₂ emissions savings in 2020" or "15% of emissions in 2020." ("Smart 2020: Enabling the Low Carbon Economy in the Information Age," The Climate Group and Global eSustainability Initiative). Review, evaluate and discuss the impact of HPC on driving a more sustainable future.

Jülich Research on Petaflops Architectures Project

12:15pm-1:15pm

Room: D135-136

Jülich Research on Petaflops Architectures Project

Primary Session Leader: Gilad Shainer (Mellanox Technologies)

Secondary Session Leader: Thomas Lippert (Forschungszentrum Juelich)

The JuRoPa II, one of the leading European PetaScale supercomputer projects, is currently being constructed at the Forschungszentrum Jülich, in the German state of North Rhine-Westphalia - one of the largest interdisciplinary research centers in Europe. The new systems were being built through an innovative alliance between Mellanox, Bull, Intel, Sun Microsystems, ParTec and the Jülich Supercomputing Centre; the first such collaboration in the world. This new 'best-of-breed' system, one of Europe's most powerful, will support advanced research in many areas such as health, information, environment, and energy. It consists of two closely coupled clusters, JuRoPA, with more than 200 Teraflop/s performance, and HPC-FF, with more than

100 Teraflop/s. The latter will be dedicated for the European fusion research community. The session will introduce the project and the initiative, how it will affect future supercomputers systems and how it will contribute to Petascale scalable software development.

MPICH: A High-Performance Open-Source MPI Implementation

12:15pm-1:15pm

Room: E145-146

MPICH: A High-Performance Open-Source MPI Implementation

Primary Session Leader: Darius Buntinas (Argonne National Laboratory)

Secondary Session Leader: Rajeev Thakur (Argonne National Laboratory), Rusty Lusk (Argonne National Laboratory)

MPICH is a popular, open-source implementation of the MPI message passing standard. It has been ported to many platforms and used by several vendors and research groups as the basis for their own MPI implementations. Following last year's successful and well-attended BoF, we are organizing another session this year. This session will provide a forum for users of MPICH, as well as developers of MPI implementations derived from MPICH, to discuss experiences and issues in using and porting MPICH. New features and future plans for fault tolerance and hybrid programming with MPI and CUDA/OpenCL will be discussed. Representatives from MPICH-derived implementations will provide brief updates on the status of their efforts. MPICH developers will also be present for an open forum discussion. All those interested in MPICH usage, development and future directions are encouraged to attend.

Securing High Performance Government Networks with Open Source Deep Packet Inspection Applications

12:15pm-1:15pm

Room: B119

Securing High Performance Government Networks with Open Source Deep Packet Inspection Applications

Primary Session Leader: Joel Ebrahimi (Bivio Networks)

Secondary Session Leader: Joe McManus (Carnegie Mellon University / CERT), Jeff Jaime (Consultant)

Sophisticated government applications, like those used for cyber security, intelligence analysis and simulation, require robust data streaming to multiple agencies in real-time without sacrificing data security or network performance. Additionally, government must accomplish this task within budget. Fortunately, some of the most innovative network security, data analysis and traffic management solutions are available to government agencies as cost-effective open source software applications. The core enabler of many of these applications is deep packet inspection (DPI), which provides unprecedented visibility into and control over network traffic. When leveraged on a high-performance, DPI-enabled platform, these applications can dramatically boost security on government networks without sacrificing performance. The audience will learn how DPI can help agencies with high-compute and high-throughput networking needs ensure security and performance, including how open source DPI applications like YAF and SiLK help the Computer Emergency Response Team (CERT) at Carnegie Mellon identify viruses, unauthorized access, malware and other vulnerabilities.

What's New about INCITE in 2010?

12:15pm-1:15pm

Room: B117

What's New about INCITE in 2010?

Primary Session Leader: Julia White (INCITE)

Awarding an unprecedented 1.3 billion processor hours to the scientific community in 2010, the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) Program selects high-impact, computationally intensive projects to carry out simulations at the U.S. Department of Energy's Leadership Computing Facilities. Beginning in 2010, Leadership Computing Facilities at Argonne and Oak Ridge national laboratories will jointly manage INCITE, granting large allocations of processor hours to select projects in science, engineering, and computer science that address challenges of national significance. Awardees come from industry, academia, and government labs. SC09 attendees are invited to learn more about the program and provide input on how INCITE can best meet the leadership-computing needs of the scientific community. Leadership Computing Facilities executive team members and the INCITE manager will provide a short overview of what's new about INCITE. Several INCITE principal investigators will summarize their scientific achievements, and an open discussion



Computing for a Changing World

SC09 Tutorials

The Tutorials Program gives attendees the opportunity to explore and learn a wide variety of important topics related to high-performance computing, networking, and storage.

SC09 has put together an exciting program of 28 tutorials: 12 full day and 16 half day. The topics span the entire range of areas of interest to

conference attendees, including programming models (CUDA, OpenCL, MPI, OpenMP, hybrid programming, PGAS languages), emerging technologies, cloud computing, visualization, performance modeling and tools, parallel I/O, parallel debugging, high-performance networking, and many other topics.

Sunday, November 15

S01: Application Supercomputing and the Many-Core Paradigm Shift

8:30am-5pm

Presenters: Alice Koniges (Lawrence Berkeley National Laboratory), William Gropp (University of Illinois at Urbana-Champaign), Ewing (Rusty) Lusk (Argonne National Laboratory), Rolf Rabenseifner (High Performance Computing Center Stuttgart), David Eder (Lawrence Livermore National Laboratory)

This tutorial provides an overview of supercomputing application development with an emphasis on the many-core paradigm shift and programming languages. We assume a rudimentary familiarity with parallel programming concepts and focus on discussing architectures, terminology, parallel languages and development tools. The architecture overview examines TOP500-type systems and surveys designs that are likely precursors to many-core platforms. Parallel programming languages (MPI, OpenMP, HPF, UPC, CAF, Titanium) are introduced and compared. An example set of small program kernels for testing and understanding these languages is provided. A specific discussion of 'undiscovered MPI' as well as philosophy and performance of hybrid MPI/OpenMP is presented. Tips for optimizing and analyzing performance are covered. Examples of real code experiences on IBM, CRAY and large cluster machines are given. A short hands-on session using participants' laptops with an xterm/wireless may be included, pending available resources.

S02: Parallel Computing 101

8:30am-5pm

Presenters: Quentin F. Stout (University of Michigan), Christiane Jablonowski (University of Michigan)

This tutorial provides a comprehensive overview of parallel computing, emphasizing those aspects most relevant to the user. It is suitable for new users, managers, students and anyone seeking an overview of parallel computing. It discusses software and hardware, with an emphasis on standards, portability and systems that are widely available. The tutorial surveys basic parallel computing concepts, using examples selected from large-scale engineering and scientific problems. These examples illustrate using MPI on distributed memory systems, OpenMP on shared memory systems and MPI+OpenMP on hybrid systems. It discusses numerous parallelization approaches, and software engineering and performance improvement aspects, including the use of state-of-the-art tools. The tutorial helps attendees make intelligent decisions by covering the primary options that are available, explaining how they are used and what they are most suitable for.

Extensive pointers to the literature and web-based resources are provided to facilitate follow-up studies.

S03: A Hands-on Introduction to OpenMP

8:30am-5pm

Presenters: Larry Meadows (Intel Corporation), Mark Bull (Edinburgh Parallel Computing Centre), Tim Mattson (Intel Corporation)

OpenMP is the de facto standard for writing parallel applications for shared memory computers. With multi-core processors in everything from laptops to high-end servers, the need for multi-threaded applications is growing and OpenMP is one of the most straightforward ways to write such programs. We will cover the full standard emphasizing OpenMP 2.5 and the new features in OpenMP 3.0. This will be a hands-on tutorial. We will have a few laptops to share, but we expect students to use their own laptops (with Windows, Linux, or OS/X). We ask that students set up their laptops in advance by either purchasing an OpenMP compiler, acquiring a demo compiler or by using an open source environment (see www.openmp.org for details).

S04: High Performance Computing with CUDA

8:30am-5pm

Presenters: David P. Luebke (NVIDIA), Ian A. Buck (NVIDIA), Jon M. Cohen (NVIDIA), John D. Owens (University of California, Davis), Paulius Micikevicius (NVIDIA), John E. Stone (University of Illinois at Urbana-Champaign), Scott A. Morton (Hess Corporation), Michael A. Clark (Boston University)

NVIDIA's CUDA is a general-purpose architecture for writing highly parallel applications. CUDA provides several key abstractions---a hierarchy of thread blocks, shared memory and barrier synchronization---for scalable high-performance parallel computing. Scientists throughout industry and academia use CUDA to achieve dramatic speedups on production and research codes. The CUDA architecture supports many languages, programming environments and libraries, including C, Fortran, OpenCL, DirectX Compute, Python, Matlab, FFT, LAPACK implementations, etc. In this tutorial NVIDIA engineers will partner with academic and industrial researchers to present CUDA and discuss its advanced use for science and engineering domains. The morning session will introduce CUDA programming, motivate its use with many brief examples from different HPC domains and discuss tools and programming environments. The afternoon will discuss advanced issues, such as optimization and sophisticated algorithms/data structures, closing with real-world case studies from domain scientists using CUDA for computational biophysics, fluid dynamics, seismic imaging and theoretical physics.

S05: Parallel I/O in Practice**8:30am-5pm**

Presenters: Robert Ross (Argonne National Laboratory), Robert Latham (Argonne National Laboratory), Marc Unangst (Panasas), Brent Welch (Panasas)

I/O on HPC systems is a black art. This tutorial sheds light on the state-of-the-art in parallel I/O and provides the knowledge necessary for attendees to best leverage I/O resources available to them. We cover the entire I/O software stack from parallel file systems at the lowest layer, to intermediate layers (such as MPI-IO), and finally high-level I/O libraries (such as HDF-5). We emphasize ways to use these interfaces that result in high performance, and benchmarks on real systems are used throughout to show real-world results. This tutorial first discusses parallel file systems in detail (PFSs). We cover general concepts and examine four examples: GPFS, Lustre, PanFS, and PVFS. Next we examine the upper layers of the I/O stack, covering POSIX I/O, MPI-IO, Parallel netCDF, and HDF5. We discuss interface features, show code examples and describe how application calls translate into PFS operations. Finally, we discuss I/O best practice.

**S06: Open-Source Stack for Cloud Computing****8:30am-5pm**

Presenters: Thomas Kwan (Yahoo! Research), Milind Bhandarkar (Yahoo!), Mike Kozuch (Intel Research), Kevin Lai (HP Labs)

With the enormous interest in cloud computing as a viable platform for scientific and computing research, we will introduce you to an open-source software stack that you can use to develop and run cloud applications or to conduct cloud systems research. We will discuss in detail the four layers of this stack: (1) Pig, a parallel programming language for expressing large-scale data analysis programs; (2) Hadoop, a distributed file system and parallel execution environment that can run Pig/Map-Reduce programs; (3) Tashi, a cluster management system for managing virtual machines; and (4) PRS, a low-level service that manages VLAN-isolated computer, storage and networking resources. Pig, Hadoop, Tashi and PRS are Apache open-source projects. Yahoo! already uses Pig and Hadoop in production and HP, Intel and Yahoo! have deployed this stack on their Open Cirrus' cloud computing testbed, supporting cloud computing research conducted by multiple labs and universities.

S07: InfiniBand and 10-Gigabit Ethernet for Dummies**8:30am-Noon**

Presenters: Dhabaleswar K. (DK) Panda (Ohio State University), Pavan Balaji (Argonne National Laboratory), Matthew Koop (NASA Goddard Space Flight Center)

InfiniBand Architecture (IB) and 10-Gigabit Ethernet (10GE) technologies are generating a lot of excitement towards building next generation High-End Computing (HEC) systems. This tutorial will provide an overview of these emerging technologies, their offered features, their current market standing, and their suitability for prime-time HEC. It will start with a brief overview of IB, 10GE and their architectural features. An overview of the emerging OpenFabrics stack which encapsulates both IB and 10GE in a unified manner will be presented. IB and 10GE hardware/software solutions and the market trends will be highlighted. Finally, sample performance numbers highlighting the performance these technologies can achieve in different environments such as MPI, Sockets, Parallel File Systems and Multi-tier Datacenters, will be shown.

S08: Principles and Practice of Application Performance Measurement and Analysis on Parallel Systems**8:30am-Noon**

Presenters: Luiz DeRose (Cray Inc.), Bernd Mohr (Forschungszentrum Juelich)

In this tutorial we will present the principles of experimental performance instrumentation, measurement and analysis of HPC Applications, with an overview of the major issues, techniques and resources in performance tools development, as well as an overview of the performance measurement tools available from vendors and research groups. In addition, we will discuss cutting edge issues, such as performance analysis on large scale multi-core systems and automatic performance analysis. Our goals are three-fold: first, we will provide background information on methods and techniques for performance measurement and analysis, including practical tricks and tips, so you can exploit available tools effectively and efficiently. Second, you will learn about simple portable techniques for measuring the performance of your parallel application. Finally, we will discuss open problems in the area for students, researchers, and users interested in working in the field of performance analysis.



S09: VisIt - Visualization and Analysis for Very Large Data Sets

8:30am-Noon

Presenters: Hank Childs (Lawrence Berkeley National Laboratory), Sean Ahern (Oak Ridge National Laboratory)

Visualization and analysis plays a crucial role for scientific simulations as an enabling technology: exploring data, confirming hypotheses and communicating results. This tutorial will focus on VisIt, an open source visualization and analysis tool designed for processing large data. The tool is intended for more than just visualization and is built around five primary use cases: data exploration, quantitative analysis, comparative analysis, visual debugging and communication of results. VisIt has a client-server design for remote visualization, with a distributed memory parallel server. VisIt won an R&D 100 award in 2005, has been downloaded over 100,000 times and is being developed by a large community. VisIt is used to visualize and analyze the results of hero runs on eight of the top twelve machines on top500.org. The tutorial will introduce VisIt, demonstrate how to do basic things in VisIt and discuss how to do advanced analysis and visualizations.



S10: Power and Thermal Management in Data Centers

8:30am-Noon

Presenters: David H. Du (University of Minnesota), Krishna Kant (Intel Corporation)

Rapid growth in online services and high-performance computing has resulted in using virtualization to consolidate distributed IT resources into large data centers. Power management becomes a major challenge in massive scale systems infrastructures such as data centers. Various estimates show that the total of powering the nation's data centers in 2007 was about \$5B and is expected to be about \$7.5B by 2011. In this tutorial we provide an overview of recent work in power management. Specifically, we try to highlight the fundamental characteristics of various devices in data centers that impact power consumption and provide an overview of various technologies and research ideas that can reduce the energy cost. Further, we make an attempt to identify future research avenues in data center power management. The subjects covered include power management for servers, storage systems, networking equipment, HVAC and integrated solutions of the four major components.

S11: Emerging Technologies and Their Impact on System Design

1:30pm-5pm

Presenters: Norman P. Jouppi (Hewlett-Packard), Yuan Xie (Pennsylvania State University)

The growing complexity of advanced deep-submicron technology used to build supercomputing systems has resulted in many design challenges for future Exascale computing systems, such as system power limits, limited communication bandwidth and latency, and reduced system reliability due to increasing transient error rates. At the same time, various disruptive emerging technologies promise dramatically improved performance at the device level. How likely are these improvements and what impact will they have at the system level? In this tutorial, we present an overview of three emerging technologies: optical interconnect, 3D integration and new non-volatile memory technologies. We describe the fundamentals and current status of each technology, introduce the advantages of the technologies and discuss the design challenges inherent in adoption of the technologies. We conclude with their possible impact on large-scale system performance.



S12: Large Scale Visualization with ParaView

1:30pm-5pm

Presenters: Kenneth Moreland (Sandia National Laboratories), John Greenfield (Global Analytic Information Technology Services), W Alan Scott (Global Analytic Information Technology Services), Utkarsh Ayachit (Kitware, Inc.), Berk Geveci (Kitware, Inc.)

ParaView is a powerful open-source turnkey application for analyzing and visualizing large data sets in parallel. ParaView is regularly used by Sandia National Laboratories analysts to visualize simulations run on the Red Storm and ASC Purple supercomputers and by thousands of other users worldwide. Designed to be configurable, extendible and scalable, ParaView is built upon the Visualization Toolkit (VTK) to allow rapid deployment of visualization components. This tutorial presents the architecture of ParaView and the fundamentals of parallel visualization. Attendees will learn the basics of using ParaView for scientific visualization with hands-on lessons. The tutorial features detailed guidance in visualizing the massive simulations run on today's supercomputers and an introduction to scripting and extending ParaView. Attendees should bring laptops to install ParaView and follow along with the demonstrations.

S13: Large Scale Communication Analysis: An Essential Step in Understanding Highly Scalable Codes

1:30pm-5pm

Presenters: Andreas Knuepfer (Technical University Dresden), Dieter Kranzlmüller (Ludwig-Maximilians-University Munich), Martin Schulz (Lawrence Livermore National Laboratory), Christof Klausecker (Ludwig-Maximilians-University Munich)

The communication structure in large-scale codes can be complex and is often unpredictable. Further, modern applications are often composed from a large number of independent modules and libraries with interleaving patterns. Nevertheless, understanding an application communication structure, in particular at scale, is a crucial step towards debugging or optimizing any code. In this tutorial we will discuss practical approaches for users to understand the communication structure of their codes, and we will present three scalable tools targeting different levels of communication analysis for MPI programs: mpiP to study aggregated communication behavior, Vampir to study fine grain traces and an automated pattern analysis approach to detect and validate repeating patterns. Combined they will lead a user from a basic understanding of the underlying communication behavior to a detailed communication pattern analysis as required for verifying a code's correctness or for optimizing its performance.

S14: Designing High-End Computing Systems with InfiniBand and 10-Gigabit Ethernet

1:30pm-5pm

Presenters: Dhabaleswar K. (DK) Panda (Ohio State University), Pavan Balaji (Argonne National Laboratory), Matthew Koop (NASA Goddard Space Flight Center)

As InfiniBand Architecture (IB) and 10-Gigabit Ethernet (10GE) technologies mature in their support for next generation high-end Computing (HEC) systems, more and more scientists, engineers and researchers are becoming interested in learning about the details of these technologies. Large-scale deployments of these technologies are also bringing new challenges in terms of performance, scalability, portability and reliability. This tutorial will provide details about the advanced features of these emerging technologies. It will start with an overview of the current large-scale deployments of clusters and the associated challenges being faced. Advanced hardware and software features and their capabilities to alleviate the bottlenecks will be emphasized. Challenges in designing next generation systems with these advanced features will be focused. Finally, case studies and experiences in designing HPC clusters (with MPI-1 and MPI-2), Parallel File and Storage Systems, Multi-tier Datacenters and Virtualization schemes will be presented together with the associated performance numbers and comparisons.

Monday, November 16

M01: A Practical Approach to Performance Analysis and Modeling of Large-Scale Systems

8:30am-5pm

Presenters: Darren J. Kerbyson (Los Alamos National Laboratory), Adolfo Hoisie (Los Alamos National Laboratory), Scott Pakin (Los Alamos National Laboratory), Kevin J. Barker (Los Alamos National Laboratory)

This tutorial presents a practical approach to the performance modeling of large-scale scientific applications on high performance systems. The defining characteristic involves the description of a proven modeling approach, developed at Los Alamos, of full-blown scientific codes, validated on systems containing 10,000s of processors. We show how models are constructed and demonstrate how they are used to predict, explain, diagnose and engineer application performance in existing or future codes and/or systems. Notably, our approach does not require the use of specific tools but rather is applicable across commonly used environments. Moreover, since our performance models are parametric in terms of machine and application characteristics, they imbue the user with the ability to "experiment ahead" with different system configurations or algorithms/coding strategies. Both will be demonstrated in studies emphasizing the application of these modeling techniques including: verifying system performance, comparison of large-scale systems and examination of possible future systems.

M02: Advanced MPI

8:30am-5pm

Presenters: Ewing Lusk (Argonne National Laboratory), William Gropp (University of Illinois at Urbana-Champaign), Rob Ross (Argonne National Laboratory), Rajeev Thakur (Argonne National Laboratory)

MPI continues to be the dominant programming model on all large-scale parallel machines, such as IBM Blue Gene/P and Cray XT5, as well as on Linux and Windows clusters of all sizes. Another important trend is the widespread availability of multi-core chips, as a result of which the individual nodes of parallel machines are increasingly multicore. This tutorial will cover several advanced features of MPI that can help users program such machines and architectures effectively. Topics to be covered include parallel I/O, multithreaded communication, one-sided communication, dynamic processes and fault tolerance. In all cases, we will introduce concepts by using code examples based on scenarios found in real applications and present performance results on the latest machines. Attendees will leave the tutorial with an understanding of how to use these advanced features of

MPI and guidelines on how they might perform on different platforms and architectures.

M03: Developing Scientific Applications using Eclipse and the Parallel Tools Platform

8:30am-5pm

Presenters: Greg Watson (IBM Corporation), Beth Tibbitts (IBM Corporation), Jay Alameda (National Center for Supercomputing Applications), Jeff Overbey (University of Illinois at Urbana-Champaign)

The Eclipse Parallel Tools Platform (PTP) is an open-source Eclipse Foundation project (<http://eclipse.org/ptp>) for parallel scientific application development. The application development workbench for the National Center for Supercomputing Applications (NCSA) BlueWaters petascale system is based on Eclipse and PTP. Eclipse offers features normally found in a commercial quality integrated development environment: a syntax-highlighting editor, a source-level debugger, revision control, code refactoring and support for multiple languages, including C, C++, UPC, and Fortran. PTP extends Eclipse to provide additional tools and frameworks for the remote development of parallel scientific applications on a wide range of parallel systems. Key components of PTP include runtime system and job monitoring, a scalable parallel debugger and static analysis of MPI and OpenMP codes. The tutorial hands-on exercises emphasize C and Fortran MPI applications, but features available to support other languages and programming models, such as OpenMP, UPC, etc., will also be described.

M04: Programming using the Partitioned Global Address Space (PGAS) Model

8:30am-5pm

Presenters: Tarek El-Ghazawi (George Washington University), Vijay Sarswat (IBM Research), Bradford Chamberlain (Cray Inc.)

The Partitioned Global Address Space (PGAS) programming model provides ease-of-use through a global shared address space while emphasizing performance through locality awareness. Over the past several years, the PGAS model has been gaining rising attention. A number of PGAS languages are now ubiquitous, such as UPC, which runs on most high-performance computers. The DARPA HPCS program has also resulted in new promising PGAS languages, such as X10 and Chapel. In this tutorial we will discuss the fundamentals of parallel programming models and will focus on the concepts and issues associated with the PGAS model. We will follow with an in-depth introduction of three PGAS languages, UPC, X10 and Chapel. We will start with basic concepts, syntax and semantics and will include a

range of issues from data distribution and locality exploitation to advanced topics such as synchronization, memory consistency and performance optimizations. Application examples will also be shared.

M05: Productive Performance Engineering of Petascale Applications with POINT and VI-HPS

8:30am-5pm

Presenters: Rick Kufrin (National Center for Supercomputing Applications), Sameer Shende (University of Oregon), Brian Wylie (Juelich Supercomputing Centre), Andreas Knuepfer (Technical University Dresden), Allen Malony (University of Oregon), Shirley Moore (University of Tennessee, Knoxville), Nick Nystrom (Pittsburgh Supercomputing Center), Felix Wolf (Juelich Supercomputing Centre), Wolfgang Nagel (Technical University Dresden)

This tutorial presents state-of-the-art performance tools for leading-edge HPC systems, focusing on how the tools are used for performance engineering of petascale, scientific applications. Four parallel performance evaluation toolsets from the POINT (Performance Productivity through Open Integrated Tools) and VI-HPS (Virtual Institute-High Productivity Supercomputing) projects are discussed: PerfSuite, TAU, Scalasca and Vampir. We cover all aspects of performance engineering practice, including instrumentation, measurement (profiling and tracing, timing and hardware counters), performance data storage, analysis and visualization. Emphasis is placed on how performance tools are used in combination for identifying performance problems and investigating optimization alternatives. The tutorial will include hands-on exercises using a Live-DVD containing all of the tools, helping to prepare participants to apply modern methods for locating and diagnosing typical performance bottlenecks in real-world parallel programs at scale. In addition to the primary presenters, the principals in the POINT and VI-HPS projects will be present.

M06: Linux Cluster Construction

8:30am-5pm

Presenters: Matthew Woitaszek (National Center for Atmospheric Research), Michael Oberg (National Center for Atmospheric Research), Theron Voran (University of Colorado), Paul Marshall (University of Colorado)

The Linux Cluster Construction tutorial provides participants a hands-on opportunity to turn a small collection of Linux hosts into a Beowulf cluster ready to run MPI-based parallel applications. Intended for those new to cluster computing, the tutorial starts by introducing basic system administration skills and the software components necessary to run parallel jobs through a batch scheduler and then addresses system performance, com-

mon administrative tasks and cluster scalability. Participants should bring a laptop and expect to work in small groups. Each group is given superuser access to a private collection of hosts (running on virtual machines offsite) to provide hands-on cluster configuration experience. The tutorial should be helpful to students and researchers interested in developing knowledge about basic system operations in order to take advantage of their local resources or professionally-administered clusters. Attendees of the Linux Cluster Construction tutorial should emerge ready to enter next year's Cluster Challenge!

M07: Cloud Computing Architecture and Application Programming

8:30am-Noon

Presenters: Dennis Gannon (Microsoft Research), Daniel Reed (Microsoft Research), Rich Wolski (University of California, Santa Barbara), Roger Barga (Microsoft Research)

Cloud computing allows data centers to provide self-service, automatic and on-demand access to services such as data storage and hosted applications that provide scalable web services and large-scale scientific data analysis. While the architecture of a data center is similar to a conventional supercomputer, they are designed with a different goal. For example, cloud computing makes heavy use of virtualization technology for dynamic application scaling and data storage redundancy for fault tolerance. The massive scale and external network bandwidth of today's data centers make it possible for users and application service providers to pay only for the resources consumed. This tutorial will cover the basic cloud computing system architectures and the application programming models. Using examples from the scientific community, we will illustrate how very large numbers of users can use the cloud to access and analyze data produced by supercomputers or gathered by instruments.

M08: Modeling the Memory Hierarchy Performance of Current and Future Multicore Systems

8:30am-Noon

Presenters: Yan Solihin (North Carolina State University)

Cache designs are increasingly critical to the overall performance of computer systems, especially in the multicore and manycore design. In addition to the "memory wall" problem in which the memory access latency is too expensive to hide by the processor, there are other multicore-specific problems that are emerging. One problem is the "cache capacity contention" that occurs when multiple cores share the last level on-chip cache. Because current cache architecture uses core-oblivious management policies, unmanaged contention results in a huge performance

volatility of many applications. Another problem is the "bandwidth wall," a situation in which the lack of off-chip bandwidth limits the performance scalability of applications in future multi-core systems. The bandwidth wall occurs due to the growth of transistor density dwarfing the growth of off-chip pin bandwidth by roughly 50% each year. Understanding these problems and solving them are important for system designers, HPC application developers and performance tuners.

M09: Hybrid MPI and OpenMP Parallel Programming

8:30am-Noon

Presenters: Rolf Rabenseifner (High Performance Computing Center Stuttgart), Georg Hager (Erlangen Regional Computing Center), Gabriele Jost (Texas Advanced Computing Center / Naval Postgraduate School)

Most HPC systems are clusters of shared memory nodes. Such systems can be PC clusters with dual or quad boards and single or multi-core CPUs, but also "constellation" type systems with large SMP nodes. Parallel programming may combine the distributed memory parallelization on the node inter-connect with the shared memory parallelization inside of each node. This tutorial analyzes the strength and weakness of several parallel programming models on clusters of SMP nodes. Various hybrid MPI+OpenMP programming models are compared with pure MPI. Benchmark results of several platforms are presented. The thread-safety quality of several existing MPI libraries is also discussed. Case studies will be provided to demonstrate various aspects of hybrid MPI/OpenMP programming. Another option is the use of distributed virtual shared-memory technologies. Application categories that can take advantage of hybrid programming are identified. Multi-socket-multi-core systems in highly parallel environments are given special consideration. Details: <https://fs.hlr.de/projects/rabenseifner/publ/SC2009-hybrid.html>

M10: Expanding Your Debugging Options

8:30am-Noon

Presenters: Chris Gottbrath (TotalView Technologies), Edward Hinkel (TotalView Technologies), Nikolay Piskun (TotalView Technologies)

With today's ever increasing complexity in development environments, programmers are facing significant challenges in effectively debugging new or ported parallel codes in the HPC environment. Developers need increased flexibility in tackling these challenges. They need to be able to step back from the usual approach of dedicated interactive debug sessions at times. They need to be able to find better methods than the ubiquitous debug cycle: Go, Step, Step, Crash, Repeat. And they need to

take the venerable printf debugging approach to a whole new level. This half-day tutorial will present ways for developers to expand and simplify their troubleshooting efforts. Techniques introduced will include: interactive debugging of MPI programs; graphical interactive remote debugging; scripted unattended debugging; and reverse debugging for the hardest bugs. Participants should bring an x86 or x86-64 laptop with CD or DVD drive to participate in the hands-on portion of the tutorial.

M11: Configuring and Deploying GridFTP for Managing Data Movement in Grid/HPC Environments

1:30pm-5pm

Presenters: Rajkumar Kettimuthu (Argonne National Laboratory), John Bresnahan (Argonne National Laboratory), Mike Link (University of Chicago)

One of the foundational issues in HPC computing is the ability to move large (multi Gigabyte, and even Terabyte) data sets between sites. Simple file transfer mechanisms such as FTP and SCP are not sufficient either from the reliability or the performance perspective. Globus implementation of GridFTP is the most widely used Open Source production quality data mover available today. In the first half of this tutorial, we will quickly walk through the steps required for setting up GridFTP on Linux/Unix machines. Then we will explore the advanced capabilities of GridFTP such as striping, and a set of best practices for obtaining maximal file transfer performance with GridFTP. In the second half, we will do hands-on exercises.

M12: Python for High Performance and Scientific Computing

1:30pm-5pm

Presenters: William R. Scullin (Argonne National Laboratory), Massimo Di Pierro (DePaul University), Nichols A. Romero (Argonne National Laboratory), James B. Snyder (Northwestern University)

Python, a high-level portable multi-paradigm interpreted programming language is becoming increasingly popular with the scientific and HPC communities due to ease of use, large collection of modules, adaptability, and strong support from vendors and community alike. This tutorial provides an introduction to Python focused on HPC and scientific computing. Throughout, we provide concrete examples, hands-on examples and links to additional sources of information. The result will be a clear sense of possibilities and best practices using Python in HPC environments. We will cover several key concepts: language basics, NumPy and SciPy, parallel programming, performance issues, integrating C and Fortran, basic visualization, large production codes, and finding resources. While it is impossible to address all

libraries and application domains, at the end participants should be able to write a simple application making use of parallel programming techniques, visualize the output and know how to confidently proceed with future projects with Python.

M13: OpenCL: A Standard Platform for Programming Heterogeneous Parallel Computers

1:30pm-5pm

Presenters: Tim Mattson (Intel Corporation), Ian Buck (NVIDIA), Mike Houston (AMD)

OpenCL is a standard for programming heterogeneous computers built from CPUs, GPUs and other processors. It includes a framework to define the platform in terms of a host (e.g. a CPU) and one or more compute devices (e.g. a GPU) plus a C-based programming language for writing programs for the compute devices. Using OpenCL, a programmer can write task-based and data-parallel programs that use all the resources of the heterogeneous computer. In this tutorial, we will introduce OpenCL. We will walk through the specification explaining the key features and how to use them to write HPC software. We will then provide a series of case studies to show how OpenCL is used in practice. By the conclusion of the tutorial, people should be able to start writing complex scientific computing applications on their own using OpenCL.

M14: Memory Debugging Parallel Applications

1:30pm-5pm

Presenters: Chris Gottbrath (TotalView Technologies), Edward Hinkel (TotalView Technologies), Nikolay Piskun (TotalView Technologies)

Memory problems such as heap memory leaks and array bounds violations can be difficult to track down in serial environments, and the challenges are truly vexing when involving an MPI program running in a cluster environment. The limited memory per-node in large clusters makes understanding and controlling memory usage an important task for scientists and programmers with scalable codes. This half day tutorial on memory debugging will begin with a discussion of memory problems. Tutorial participants will learn how to use MemoryScape to debug memory problems in both serial and parallel applications. Participants will learn how to analyze memory leaks and array bounds violations and create various kinds of HTML, text and binary reports for sharing memory analysis results and automate memory testing. Participants should bring an x86 or x86-64 based laptop with CD or DVD to participate in the hands-on portion of the tutorial.



Computing for a Changing World

SC09 Workshops

Thirteen workshops have been scheduled at SC09. Workshops provide attendees with independently planned full-, half-, or multi-day sessions that complement the SC09 Technical

Program and extend its impact by providing greater depth of focus. Workshops are being held on November 15, 16, and 20.

Sunday, November 15

4th Petascale Data Storage Workshop

9am-5:30pm

Room: A106

Organizers: Garth A. Gibson (Carnegie Mellon University / Panasas)

Petascale computing infrastructures make petascale demands on information storage capacity, performance, concurrency, reliability, availability and manageability. This one-day workshop focuses on the data storage problems and emerging solutions found in petascale scientific computing environments. Special attention will be given to: issues in which community collaboration can be crucial, problem identification, workload capture, solution interoperability, standards with community buy-in, and shared tools. This workshop seeks contributions on relevant topics, including but not limited to: performance and benchmarking results and tools, failure tolerance problems and solutions, APIs for high performance features, parallel file systems, high bandwidth storage architectures, wide area file systems, metadata intensive workloads, autonomies for HPC storage, virtualization for storage systems, data-intensive and cloud storage, archival storage advances, resource management innovations, etc. Submitted extended abstracts (up to 5 pages, due Sept. 18, 2009) will be peer reviewed for presentation and publication on www.pdsi-scidac.org and in the ACM digital library.

5th International Workshop on High Performance Computing for Nano-science and Technology (HPCNano09)

9am-5:30pm

Room: A107

Organizers: Jun Ni (University of Iowa), Andrew Canning (Lawrence Berkeley National Laboratory), Lin-Wang Wang (Lawrence Berkeley National Laboratory), Thom Dunning (National Center for Supercomputing Applications)

This workshop's main theme is "Cyber Gateway for Nano Discoveries and Innovation." It is about how to conduct innovative research and development using cyberinfrastructure for large-scale computations in nano-science and nanotechnology. Several major subjects around the theme are: (1) sustainable computing for nano discovery and exploration; (2) defining a future roadmap in computational nanotechnology; (3) developing emergent cognition to accelerate innovative applications; and (4) catalyzing the change and reformation, recovery, and reinvestment for future energy, education, health care, and security, as well as neutralizing and fertilizing the business research develop-

ments for domestic and global economics blooming. Nanotechnology is an exciting field with many potential applications. Its impact is already being felt in materials, engineering, electronics, medicine and other disciplines. Current research in nanotechnology requires multi-disciplinary knowledge, not only in sciences and engineering but also in HPC. This workshop offers academic researchers, developers and practitioners an opportunity to discuss various aspects of HPC-related computational methods and problem solving techniques for nanotechnology research.

Component-Based High Performance Computing 2009 (Day 1)

9am-5:30pm

Room: A103

Organizers: Nanbor Wang (Tech-X Corporation), Rosa M. Badia (Technical University of Catalonia)

This workshop will bring together developers and users of component-based high-performance computing (CBHPC) frameworks and environments. The workshop aims to build an international research community around CBHPC issues, such as the role of component and framework technologies in high-performance and scientific computing, high-level patterns and features unique to HPC and tools for efficiently developing component-based high-performance applications, with target environments including individual massively parallel systems, the Grid and hybrid, hardware-accelerated high-performance computing environments.

Third International Workshop on High-Performance Reconfigurable Computing Technology and Applications (HPRCTA'09)

9am-5:30pm

Room: A108

Organizers: Volodymyr Kindratenko (National Center for Supercomputing Applications), Tarek El-Ghazawi (George Washington University), Eric Stahlberg (Wittenberg University)

High-Performance Reconfigurable Computing (HPRC), based on the combination of conventional microprocessors and field-programmable gate arrays (FPGA), is a rapidly evolving computing paradigm that offers a potential to accelerate computationally-intensive scientific applications beyond what is possible on today's mainstream HPC systems. The academic community has been actively investigating this technology for the past several years, and the technology has proven itself to be practical for a number of HPC applications. Many of the HPC vendors are now offering various HPRC solutions. The goal of this workshop is to provide a forum for computational scientists who use recon-

figurable computers and companies involved with the development of the technology to discuss the latest trends and developments in the field. Topics of interest include architectures, languages, compilation techniques, tools, libraries, run-time environments, performance modeling/prediction, benchmarks, algorithms, methodology, applications, trends and the latest developments in the field of HPRC.

Workshop on High Performance Computational Finance

9am-5:30pm

Room: A105

Workshop on High Performance Computational Finance

Organizers: David Daly (IBM Research), Jose Moreira (IBM Research), Maria Eleftheriou (IBM Research), Kyung Ryu (IBM Research)

The purpose of this workshop is to bring together practitioners, researchers, vendors, and scholars from the complementary fields of computational finance and high performance computing, in order to promote an exchange of ideas, discuss future collaborations and develop new research directions. Financial companies increasingly rely on high performance computers to analyze high volumes of financial data, automatically execute trades and manage risk. As financial market data continues to grow in volume and complexity and algorithmic trading becomes increasingly popular, there is increased demand for computational power. Recent events in the world economy have demonstrated a great need for better models and tools to perform risk analysis and risk management. Therefore, we are selecting risk management as a focus area for this workshop.

Monday, November 16



2009 Workshop on Ultra-Scale Visualization

9am-5:30pm

Room: A109

Organizers: Kwan-Liu Ma (University of California, Davis), Michael Papka (Argonne National Laboratory)

The output from leading-edge scientific simulations is so voluminous and complex that advanced visualization techniques are necessary to interpret the calculated results. With the progress we have made in visualization technology over the past twenty years, we are barely capable of exploiting terascale data to its full extent, but petascale datasets are on the horizon. This workshop aims at addressing this pressing issue by fostering communication among

visualization researchers and practitioners, high-performance computing professionals and application scientists. The technical program will include both invited and solicited talks. Attendees will be introduced to the latest and greatest research innovations in large data visualization and also learn how these innovations impact the scientific supercomputing and discovery process. For more information about this workshop, please visit <http://vis.cs.ucdavis.edu/Ultravis09>.

2nd Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS)

9am-5:30pm

Room: A105

Organizers: Ioan Raicu (University of Chicago), Ian Foster (University of Chicago / Argonne National Laboratory), Yong Zhao (Microsoft Corporation)

This workshop will provide the scientific community a forum for presenting new research, development and deployment efforts of loosely-coupled large scale applications on large scale clusters, Grids, Supercomputers, and Cloud Computing infrastructures. Many-task computing (MTC) encompasses loosely-coupled applications, which are generally composed of many tasks to achieve some larger application goal. This workshop will cover challenges that can hamper efficiency and utilization in running applications on large-scale systems, such as local resource manager scalability and granularity, efficient utilization of the raw hardware, parallel file system contention and scalability, reliability at scale and application scalability. We welcome paper submissions on all topics related to MTC. For more information, please see <http://dsl.cs.uchicago.edu/MTAGS09/>. Selected excellent work will be invited to submit extended versions to the IEEE Transactions on Parallel and Distributed Systems (TPDS) Journal, Special Issue on Many-Task Computing (http://dsl.cs.uchicago.edu/TPDS_MTC/).

4th Workshop on Workflows in Support of Large-Scale Science (WORKS09)

9am-5:30pm

Room: A108

Organizers: Ewa Deelman (Information Sciences Institute), Ian Taylor (Cardiff University)

Scientific workflows are now being used in a number of scientific disciplines such as astronomy, bioinformatics, earth sciences and many others. Workflows provide a systematic way of describing the analysis and rely on workflow management systems to execute the complex analyses on a variety of distributed resources. This workshop focuses on both application experiences and the

many facets of workflow systems that operate at a number of different levels, ranging from job execution to service management. The workshop therefore covers a broad range of issues in the scientific workflow lifecycle that include but are not limited to: designing workflow composition interfaces; workflow mapping techniques that may optimize the execution of the workflow; workflow enactment engines that need to deal with failures in the application and execution environment; and a number of computer science problems related to scientific workflows, such as semantic technologies, compiler methods, fault detection and tolerance.

Component-Based High Performance Computing 2009 (Day 2)

9am-5:30pm

Room: A103

Organizers: Nanbor Wang (Tech-X Corporation), Rosa M. Badia (Technical University of Catalonia)

This workshop will bring together developers and users of component-based high-performance computing (CBHPC) frameworks and environments. The workshop aims to build an international research community around CBHPC issues such as the role of component and framework technologies in high-performance and scientific computing, high-level patterns and features unique to HPC, and tools for efficiently developing component-based high-performance applications, with target environments including individual massively parallel systems, the Grid and hybrid, hardware-accelerated high-performance computing environments.

User Experience and Advances in Bridging Multicore's Programmability Gap

9am-5:30pm

Room: A106

Organizers: Scott Michel (Aerospace Corporation), Hans Zima (Jet Propulsion Laboratory), Nehal Desai (Aerospace Corporation)

Multicore's "programmability gap" refers to the mismatch between traditionally sequential software development and today's multicore and accelerated computing environments. New parallel languages break with the conventional HPC programming paradigm by offering high-level abstractions for control and data distribution, thus providing direct support for the specification of efficient parallel algorithms for multi-level system hierarchies based on multicore architectures. Together with the emergence of architecture-aware compilation technology, these developments signify important contributions to bridging this programmability gap. Last year's "Bridging Multicore's

Programmability Gap" workshop examined up-and-coming languages, such as Chapel, X-10 and Haskell, and their approaches to bridging the gap. The foci of this year's workshop are on user experience in using new languages for challenging applications in a multicore environment, the progress made in the area of the accelerated computing software development lifecycle, as well as new compilation and programming environment technology supporting emerging languages.



Using Clouds for Parallel Computations in Systems Biology

9am-5:30pm

Room: A104

Organizers: Susan Gregurick (DOE), Folker Meyer (Argonne National Laboratory), Peg Folta (Lawrence Livermore National Laboratory)

Modern genomics studies use many high-throughput instruments that generate prodigious amounts of data. For example, a single run on a current sequencing instrument generates up to 17 GB of data, or one-third of the genomics sequence space (our archives of complete genomic data currently comprise 51 GB). The situation is further complicated by the democratization of sequencing; many small centers can now independently create large sequence data sets. Consequently, the rate of sequence production is growing far faster than our ability to analyze the data. Cloud computing provides a tantalizing possibility for on-demand access to computing resources. Many computations fall under the "embarrassingly parallel" header and should be ideally suited for cloud computing. However, challenging issues remain, including data transfer and local data availability on the cloud nodes. This workshop will bring together computer scientists, bioinformaticists and computational biologists to discuss the feasibility of using cloud computing for sequence analysis.

Friday, November 20th

ATIP 1st Workshop on HPC in India: Research Challenges on Computing in India

8:30am-5pm

Room: E141-142

Organizers: David K. Kahaner (Asian Technology Information Program), Bharat Jayaraman (State University of New York at Buffalo), R. Govindarajan (Indian Institute of Science Bangalore), P. Venkat Rangan (University of California, San Diego)

ATIP's First Workshop on HPC in India, sponsored by the US National Science Foundation (NSF), will include a significant set of presentations, posters and panels from a delegation of Indian academic, research laboratory, industry experts, and graduate students, addressing topics including Government Plans, University Research, Infrastructure, and Industrial Applications. The workshop will feature both Indian and international vendor participation and a panel discussion to identify topics suitable for collaborative (US-India) research and the mechanisms for developing those collaborations. A key aspect is the unique opportunity for the US research community to interact with top Indian HPC scientists.

Early Adopters PhD Workshop: Building the Next Generation of Application Scientists

8:30am-5pm

Room: D137-138

Organizers: David Abramson (Monash University), Wojtek Goscinski (Monash University)

Successfully applying HPC can be a challenging undertaking to newcomers from fields outside of computing. This workshop provides graduate students who are adopting HPC an opportunity to present early stage research and gain valuable feedback. A panel of expert reviewers with significant experience will be invited to come and critique students' work and provide constructive feedback. The goal of this workshop is to help students identify shortcomings, introduce new approaches, discuss new technology, learn about relevant literature or define their future research goals.

Grid Computing Environments

8:30am-5pm

Room: D139-140

Organizers: Marlon Pierce (Indiana University), Gopi Kandaswamy (ISI)

Scientific portals and gateways are important components of many large-scale Grid and Cloud Computing projects. They are characterized by web-based user interfaces and services that securely access Grid and Cloud resources, data, applications and collaboration services for communities of scientists. As a result, the scientific gateway provides a user- and (with social networking) a community-centric view of cyberinfrastructure. Web technologies evolve rapidly, and trends such as Cloud Computing are changing the way many scientific users expect to interact with resources. Academic clouds are being created using open source cloud technologies. Important Web standards such as Open Social and OAuth are changing the way web portals are built, shared and secured. It is the goal of the GCE workshop series to provide a venue for researchers to present pioneering, peer-reviewed work on these and other topics to the international science gateway community.



Computing for a Changing World

SC09 SCinet

SCinet provides commodity Internet, research and experimental networks for use by SC conference exhibitors and attendees.

SCinet

Wireless Network Service Policy

SCinet provides wireless networks for use by all SC09 exhibitors and attendees at no charge. Wireless access points are installed on the exhibit floor, in ballrooms and meeting rooms, and in many common areas in the Convention Center. Network limitations, such as areas of reduced signal strength, limited client capacity or other coverage difficulties may be experienced in certain areas. As a courtesy to other users, please do not download large files via the wireless network.

Laptops and other wireless devices configured to request network configuration information via DHCP receive the needed network settings automatically upon entering the SCinet wireless coverage area. The wireless networks are governed by the policy posted on conference website. Please help us ensure a successful, pleasant experience for all SC09 attendees by observing the following policies: Exhibitors and attendees may not operate their own IEEE 802.11 (a, b, g, n or other standard) wireless Ethernet access points anywhere within the convention center, including within their own booths. Wireless clients may not operate in ad-hoc or peer-to-peer mode, due to the potential for interference with other wireless clients.

Exhibitors and attendees may not operate cordless phones, wireless video or security cameras, or any other equipment transmitting in the 2.4GHz or 5.2GHz spectrum.

SCinet reserves the right to disconnect any equipment that interferes with the SCinet wireless networks.

Network Security Policy

SCinet peers with the Internet and with agency and national wide area networks through a series of very high-speed connections. To maximize performance across these interfaces, there are no firewalls. In this regard, the SCinet network is a logical, albeit temporary, extension of the open Internet. Any Internet user in the world, including those associated with international organized crime, can communicate freely with any host connected to SCinet. Exhibitors and attendees are reminded that network security is a collective responsibility. The SCinet Network Security team focuses on the integrity and availability of the SCinet network infrastructure. Exhibitors and attendees must ensure their security tools (e.g. host-based firewalls, antivirus software) are properly configured, turned on and kept up to date. Please pay particular attention to systems that are normally operated on private or firewalled networks.

All wireless networks, including those provided by SCinet, are vulnerable by their very nature. Each attendee and exhibitor is responsible for protecting the confidentiality and integrity of their communications sessions. SCinet strongly discourages (but does not prevent) the use of insecure applications including TELNET, POP, IMAP, SNMPv2c and FTP. These applications are

subject to compromise because they send passwords to remote hosts in human readable, clear text format. Attendees are strongly encouraged to protect their sessions through mechanisms such as Secure Shell (SSH), Virtual Private Network (VPN) and Secure Sockets Layer (SSL).

How SCinet Works

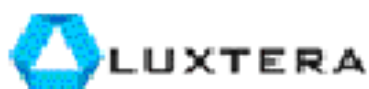
During SC09 week, Portland will host one of the most powerful, advanced networks in the world: SCinet. Built from the ground up each year, SCinet brings to life a highly sophisticated, high capacity networking infrastructure that can support the revolutionary applications and network experiments that have become the trademark of the SC conference. SCinet serves as the platform for exhibitors to demonstrate the advanced computing resources from their home institutions and elsewhere by supporting a wide variety of applications, including supercomputing and cloud computing.

Designed and built entirely by volunteers from universities, government and industry, SCinet connects multiple 10-gigabit per second (Gbps) circuits to the exhibit floor, linking the conference to research and commercial networks around the world, including Internet2, Department of Energy's ESnet, National LambdaRail, Level 3 Communications, Qwest and others.

Powerful enough to transfer more than 300 gigabits of data in just one second, SCinet features three major components. First, it provides a high performance, production-quality network with direct wide area connectivity that enables attendees and exhibitors to connect to networks around the world. Second, SCinet includes an OpenFabrics-based InfiniBand network that provides support for high-performance computing (HPC) applications and storage over a high performance, low-latency network infrastructure. This year's InfiniBand fabric will consist of Quad Data Rate (QDR) 40-gigabit per second (Gbps) circuits, providing SC09 exhibitors with up to 120Gbps connectivity. Third, SCinet builds an extremely high performance experimental network each year, called Xnet (eXtreme net). This serves as a window into the future of networking, allowing exhibitors showcase "bleeding-edge" (emerging, often pre-commercial or pre-competitive) networking technologies, protocols and experimental networking applications.

Volunteers from educational institutions, high performance computing centers, network equipment vendors, U.S. national laboratories, research institutions and research networks and telecommunication carriers work together to design and deliver the SCinet infrastructure. Industry vendors and carriers donate much of the equipment and services needed to build the local and wide area networks. Planning begins more than a year in advance of each SC conference and culminates with a high-intensity installation just seven days before the conference begins.





PACIFIC





Computing for a Changing World

SC09 Services

This section is a quick reference guide to services provided by SC09. These services are listed in alphabetical order for your convenience.

SC Services

SC Services

Access Policies

See the Registration section beginning on page ???

Accessibility

The Oregon Convention Center complies with ADA requirements and is wheelchair accessible, as are the Portland area buses and the MAX Light Rail trains.

Age Requirements Policy

- Technical Program attendees must be 16 years of age or older. Age verification is required.
- Exhibits-Only registration is available for children ages 12-16. Age verification is required.
- Children under 12 are not permitted in the Exhibit Hall other than on Family Day (Wednesday), 4pm-6pm. Children under 16 are not allowed in the Exhibit Hall during installation, dismantling or before or after posted exhibit hours. Anyone under 16 must be accompanied by an adult at all times while visiting the exhibits.

Airport Transportation

Portland's MAX Light Rail provides direct service from the convention center to the airport. Use the red line and purchase a ticket that is for "all zones" (\$2.30). Trains leave approximately every 15 minutes from 4am to 11pm on weekdays. The trip takes about 30 minutes.

Automated Teller Machines (ATMs)

Two ATMs are available in the convention center, one in the upper lobby near FedEx/Kinko's and the second near the meeting rooms downstairs (adjacent to Starbuck's).

Badge Policies

See the Registration section beginning on page ???

Baggage Check

For your convenience, a coat and bag check is available during conference hours in the main lobby area.

Banks/Currency Exchange

From the convention center, the closest banks are located in or across from Lloyd Center. This shopping mall is about 5 blocks east and 2 blocks north of the convention center, between Multnomah and Halsey and between 9th and 15th Avenues.

Business Center

The convention center has a self-service business area that includes coin- and credit-card-operated copiers and printers, as

well as a scanner and PC workstations. The facility is located on the Oregon Ballroom level. A FedEx/Kinko's is available just a few blocks away (at Weidler and 7th Ave; 503-284-2129), and they will pickup/deliver within 30 minutes for any large job.

Busing

Conference buses are available only to a handful of hotels. Most attendees will use the MAX Rapid Transit System; see "Public Transportation." If you are staying at one of the following hotels, buses will be available to/from the convention center during key hours on Saturday through Friday: Oxford Suites, Red Lion on the River, Hilton Vancouver, Marriott Waterfront, Hotel Modera, University Place, Residence Inn River Place. Signs with schedule information will be available at the hotels and near the bus drop-off point at the convention center (east side of the center).

Camera/Recording Policies

- No photography or audio/video recording is permitted at SC09. Abuse of this policy will result in revocation of the individual's registration credentials.
- SC09 employs a professional photographer and reserves the right to use all images that he/she takes during the conference for publication and promotion of future SC events.

Child Care

Child care will not be provided at SC09. Contact your hotel concierge for suggestions.

Coat Check

See "Baggage Check."

Conference Management Office

If you have questions regarding the management of SC09, stop by this office (Room F151/152) any time during conference program hours.

Exhibit Hours

The Exhibit Floor will be open to Technical Program attendees and individuals with Exhibits-Only badges, during the following hours:

- Tuesday-Wednesday, 10am-6pm
- Thursday, 10am-4pm

Early access will be given to Technical Program attendees only during the Gala Opening Reception on Monday evening (see "Social Events" in this section).

Exhibition Management Office

Exhibition Management representatives are available in Room A109 during conference program hours, to meet with exhibitors and help with plans for exhibiting at SC09 and SC10.

Exhibitor Registration

See the Registration section beginning on page ???

Food Services

The convention center operates several snack bars, food carts, and restaurants throughout the convention center. Consult the convention center map for locations.

First Aid

The Oregon Convention Center provides an onsite first aid room staffed with an emergency medical professional, on the lower level outside the door to Exhibit Hall A (adjacent to the International Attendees Center). In the event of a medical or other emergency, attendees can dial 911 from any pay phone or dial 0 from any house phone located in the facility for immediate assistance. In addition, all uniformed security personnel are available to assist you in any emergency.

Information

See "SC Your Way."

International Attendees Center

The SC Ambassadors and multi-lingual assistants are available in the lobby area outside Exhibit Hall A, to answer questions, offer suggestions, and help make connections among international attendees.

Internet Access

See the SCinet section beginning on page ???

Lost and Found

Found items may be turned in at either the SC Your Way booth (adjacent to Registration), or the Security Office (room A-102). To retrieve lost items, go directly to the Security Office.

Lost Badges

See the Registration section beginning on page ??

MAX Light Rail

See "Public Transportation."

Mailing Services

UPS agents will be available on-site on Thursday afternoon and Friday morning. At other times, there is a FedEx/Kinko's just a few blocks away (at Weidler and 7th Ave), or you can arrange shipping through your hotel concierge service.

Media Room

Media representatives and industry analysts should visit the Media Room (C125) for onsite registration. A valid media iden-

tification card, business card, or letter from a media outlet verifying a freelance assignment is required for media registration. The Media Room provides resources to media representatives for writing, researching and filing their stories and for interviewing conference participants and exhibitors. The facility is also available to SC09 exhibitors who wish to provide materials to, or arrange interviews with, media representatives and industry analysts covering the conference. A separate room will be available for conducting interviews and may be reserved through the Media Room coordinator during the conference.

Credentialed media representatives are allowed to photograph or record the SC exhibits and most Technical Program sessions, as long as they are not disruptive. Under no circumstances may anyone photograph or record the Gore keynote address. Whenever possible, the Communications Committee will provide an escort to assist with finding appropriate subjects.

Parking

Covered and uncovered parking is available on a first-come first-served basis at the convention center. The service is available 6am-midnight each day and costs \$8 per day.

Pay Phones

Public phones are available outside Exhibit Halls A and C, or in the Ginkoberry Concourse.

Press Passes

See "Media Room."

Printing/Photocopy Services

See "Business Center."

Proceedings

See the Registration section beginning on page ??

Public Transportation

Most attendees will use the MAX Light Rail to provide transportation to/from conference hotels and the convention center. Those locations are all within Portland's "fare-free zone," so no tickets are required. All three lines (Red, Blue, Green) that pass by the convention center go downtown, but Red and Blue should generally be used for getting to the conference hotels. The trains operate 18-22 hours per day, depending on day and destination. More information is available online at <http://scyourway.supercomputing.org>. Airport service involves a fee; see "Airport Transportation." The MAX station is located on the north side of the convention center.

Registration

See the Registration section beginning on page ???

SC Your Way

The SC Your Way booth, located adjacent to Registration, is the place to go with questions about the conference schedule, what activities will be held where/when, local area restaurants and activities, public transportation, and how to find things in the convention center. It will be open for questions at the following times:

- **Sunday, 8am-6pm**
- **Monday, 8am-9pm**
- **Tuesday-Thursday, 8am-6pm**
- **Friday, 8-11am**

Visit <http://scyourway.supercomputing.org> to plan your visit, create a customized schedule, or generate a personal map of the exhibits you want to visit.

Shipping Services

See "Mailing Services."

Special Assistance Desk

Located adjacent to Registration, this desk provides assistance with problems having to do with registration or purchases, including:

- Credit card problems (validations, errors)
- Lost badges (note the policies concerning replacement cost, under "Registration")
- Registration corrections and upgrades

Store

Review and purchase additional technical materials, SC09 clothing, and gift items for friends, family, and colleagues at the SC09 Store, located near Registration in the Ginkoberry Concourse.

The Store is open at the following times:

- Saturday, 1-6pm
- Sunday, 7am-6pm
- Monday, 7am-9pm
- Tuesday-Wednesday, 7:30am-6pm
- Thursday, 7:30am-5pm
- Friday, 8-11am

Student Volunteers

Undergraduate and graduate student volunteers assist with the administration of the conference, receiving in exchange conference registration, housing, and most meals. Student volunteers have the opportunity to experience and discuss the latest HPC technologies and to meet leading researchers from around the world while contributing to the success of the conference. SC09 attendees are encouraged to share information about the SC Student Volunteers program with their colleagues and to encourage students to apply for future conferences.

Technical Program Registration

See the Registration section beginning on page ???

Wireless Network Access

See the SCinet section beginning on page ???

SC09 Conference Hotels

Avalon Hotel & Spa

0455 SW Hamilton Court
Portland OR 97239
503-802-5800

The Benson Hotel

309 SW Broadway
Portland OR 97205
503-228-2000

Courtyard by Marriott (Portland City Center)

550 SW Oak St.
Portland OR 97204
503-505-5000

Courtyard by Marriott Portland (Lloyd Center)

435 NE Wasco St.
Portland OR 97232
503-224-3400
The Heathman Hotel
1001 SW Broadway
Portland Oregon 97205
503-241-4100

Hilton Portland and Executive Tower

921 SW Sixth Ave.
Portland OR 97204
503-226-1611

Hilton Vancouver Washington

301 West 6th St.
Vancouver WA 98660
360-993-4500

Hotel 50

50 SW Morrison St.
Portland OR 97204
503-221-0711

Hotel deLuxe

729 SW 15th Ave
Portland OR 97205
503-219-2094

Hotel Lucia

400 SW Broadway Ave.
Portland OR 97205
503-225-1717

Hotel Modera

515 SW Clay St.
Portland OR 97201
503-484-1084

Hotel Monaco

506 SW Washington St.
Portland OR 97204
503-222-0001

Inn at the Convention Center

420 NE Holladay St.
Portland OR 97232
503-233-6331

La Quinta Inn Portland Convention Center

431 NE Multnomah St.
503-233-7933 Oxford Suites
12226 N Jantzen Dr.
Portland OR 97217
503-283-3030

The Paramount Hotel

808 SW Taylor St.
Portland OR 97215
503-223-9900

Portland Marriott City Center

520 SW Broadway
Portland OR 97205
503-226-6300

Portland Marriott Downtown Waterfront

1401 SW Naito Parkway,
Portland OR 97201
503-226-7600

Red Lion on the River

909 N Hayden Island Dr.
Portland OR 97217
503-283-4466

Red Lion Hotel Portland

1021 NE Grand Ave.
Portland OR 97232
503-235-2100

Residence Inn by Marriott - Portland Downtown at RiverPlace

2115 SW River Pkwy
503-234-3200

Crowne Plaza Portland Downtown

1441 NE 2nd Ave
Portland OR 97232
503-233-2401

Doubletree Hotel Portland

1000 NE Multnomah St.
Portland OR 97232
503-281-6111

Embassy Suites Portland Downtown

319 SW Pine St.
Portland OR 97204
503-279-9000

The Governor Hotel

614 SW 11th Ave.
Portland OR 97205
503-224-3400

Hotel Lucia

400 SW Broadway Ave.
Portland OR 97205
503-225-1717

Hotel Modera

515 SW Clay St.
Portland OR 97201
503-484-1084

Hotel Monaco

506 SW Washington St.
Portland OR 97204
503-222-0001

Inn at the Convention Center

420 NE Holladay St.
Portland OR 97232
503-233-6331

La Quinta Inn Portland Convention Center

431 NE Multnomah St.
503-233-7933 Oxford Suites
12226 N Jantzen Dr.
Portland OR 97217
503-283-3030

Residence Inn by Marriott Portland Downtown/ Lloyd Center

1710 NE Multnomah St.
Portland OR. 97232
503-288-4849

The Nines

525 SW Morrison St.
Portland OR 97204
503-222-9996

University Place Hotel & Conference Center

310 SW Lincoln St.
Portland Oregon 97201
503-221-0140

Hotel Vintage Plaza

422 SW Broadway
Portland OR 97205
503-228-1212

The Westin Portland

750 SW Alder St.
Portland OR 97205
503-294-9000



Computing for a Changing World

SC09 Acknowledgements

It takes several years of hard work by hundreds of volunteers scattered around the world to plan and carry out an SC conference. SC09 gratefully acknowledges the many supporters of this year's conference, including representatives from academia, industry and the government.

The success of SC09 is a reflection of their dedication and commitment. Without the support of the following people, their organizations and their funding sources, SC09 would not be possible. Thank you!

SC09 Committee Members

Conference Committees

Wilfred Pinfold, Conference Chair
Intel Corporation

Barbara Hobbs, Executive Assistant
Intel Corporation

William Douglas Gropp, Technical Program Chair
University Of Illinois At Urbana-Champaign

Mike Bernhardt, Communications Chair
Libra Strategic Communications

Trish Damkroger, Infrastructure Chair
Lawrence Livermore National Laboratory

Jeffrey K. Hollingsworth, Finance Chair
University of Maryland

Becky Verastegui, Exhibits Chair
Oak Ridge National Laboratory

Barry V. Hess, Deputy General Chair
Sandia National Laboratories

Cherri M. Pancake,
SC Communities Chair
Oregon State University

Ralph A. McEldowney, Scinet Chair
Air Force Research Laboratory

David Harter, Production
Self-Employed

Society Liaisons - ACM

Donna Cappel
Ashley Cozzi
Jessica Fried

Society Liaisons- IEEE Computer Society

Thomas Baldwin
Lynne Harris
Anne Marie Kelly

Thrust Areas

Brent Gorda, Sustainability Chair
Lawrence Livermore National Laboratory
Peg Folta, Biocomputing Chair
Lawrence Livermore National Laboratory
John Hengeveld, 3D Internet Chair
Intel Corporation

Technical Program

(See Area Chairs Committee List On Page ??)

William Douglas Gropp, Technical Program Chair
University of Illinois at Urbana-Champaign

Ricky A. Kendall, Technical Program Deputy Chair
Oak Ridge National Laboratory

Andrew A. Chien, Technical Papers Chair
Intel Corporation

Satoshi Matsuo, Technical Papers Chair
Tokyo Institute of Technology

Fred Johnson, Tutorials Chair
Consultant

Rajeev Thakur, Tutorials Chair
Argonne National Laboratory

Daniel A. Reed, Plenary Speakers Chair
Microsoft Corporation

Dona Crawford, Keynote Chair
Lawrence Livermore National Laboratory

Harvey Wasserman, Masterworks Chair
Lawrence Berkeley National Laboratory

Toni Cortes, Birds-Of-A-Feather Chair
Barcelona Supercomputing Center

Padma Raghavan, Doctoral Showcase Chair
Pennsylvania State University

Robert F. Lucas, Disruptive Technologies Chair
Information Sciences Institute

Jack Dongarra, Posters Chair
University of Tennessee, Knoxville

Manuel Vigil, Challenges Chair
Los Alamos National Laboratory

Mary Hall, Awards Chair
University of Utah

Jack Dongarra, Awards Deputy Chair
University of Tennessee, Knoxville

Lois Curfman McInnes, Workshops Co-Chair

Argonne National Laboratory

Thomas Lawrence Sterling, Workshops Co-Chair

Louisiana State University

Bernd Mohr, Panels Chair

Juelich Supercomputing Centre

Tim Mattson, Panels Vice Chair
Intel Corporation

Exhibits

Becky Verastegui, Exhibits Chair
Oak Ridge National Laboratory

David Cooper, Exhibitor Liaison
Lawrence Livermore National Laboratory

Eric Sills, Exhibitor Liaison
North Carolina State University

Paul Graller, Exhibits Management
Hall-Erickson

Mike Weil, Exhibits Management
Hall-Erickson

Stephen E Hagstette Jr, Exhibits Logistics
Freeman Company

Darryl Monahan, Exhibits Logistics
Freeman Company

John Grosh, Exhibitor Forum Chair
Lawrence Livermore National Laboratory

Jeffery A. Kuehn, Exhibits - Sustainability Chair
Oak Ridge National Laboratory

Infrastructure

Trish Damkroger, Infrastructure Chair
Lawrence Livermore National Laboratory

Trey Breckenridge, Infrastructure Deputy Chair
Mississippi State University

Matt Link, Infrastructure Deputy Chair
Indiana University

Jim Costa, Special Projects
Sandia National Laboratories

Kevin Wohlever, Digital Archive Chair
Ohio Supercomputer Center

Barbara Fossum, Technical Program Liaison
Purdue University

James W. Ferguson, Space Chair
National Institute For Computational Sciences
Gary New, Electrical/Cabling Chair
National Center For Atmospheric Research
Rebecca Hartman-Baker, Signage Chair
Oak Ridge National Laboratory
Frank Behrendt, Sustainable Initiative
 Chair
Berlin Institute of Technology
Christine E. Cuicchi, Internet Access
 Chair
Navy DOD Supercomputing Resource Center
Eric Sills, AV/PC Chair
North Carolina State University
 AV/PC Contractor
Gabe Blakney, *AV Concepts*
Dan Sales, *AV Concepts*

Portland Experience

Janet Brown, Chair
Pittsburgh Supercomputing Center
Mary Peters, Local Transportation
Meet Green
Mary Peters, Event Management
Meet Green
Barbara Horner-Miller, Housing Chair
Arctic Region Supercomputing Center
Mark Farr, Housing Contractor
Housing Connection

Security

Tim Yeager, Security Chair
Air Force Research Laboratory

Security Staff

Christine E. Cuicchi, *Navy DOD Supercomputing Resource Center*
Jeff Graham, *Air Force Research Laboratory*
G.M. 'Zak' Kozak, *Texas Advanced Computing Center*
Eric McKinzie, *Lawrence Livermore National Laboratory*

Security Contractor

Peter Alexan, *RA Consulting*
Liz Clark, *RA Consulting*
Linda Duncan, Conference Office
 Manager
Oak Ridge National Laboratory
April Hanks, Conference Office Manager
 Assistant
Intel Corporation

Rebecca Mebane, Catering (Meeting
 Strategies)
Meet Green

Communications

Mike Bernhardt, Communications Chair
Libra Strategic Communications
Lauren Rotman, Deputy
 Communications Chair
Internet2
Jon Bashor, Special Advisor
Lawrence Berkeley National Laboratory
Vivian Benton, Conference Program
 Coordinator
Pittsburgh Supercomputing Center
Janet Brown, Proceedings
Pittsburgh Supercomputing Center
Angela Harris, Media Concierge
DOD HPC Modernization Program
Betsy Riley, Advance Media Registration
 & Media Room
DOE
Lizzie Bennett, Media Relations
Liaison PR
Jimme Peters, Media Relations
24/7 Consulting
Kathryn Kelley, Conference Newsletter
Ohio Supercomputer Center
Rich Brueckner, Social Media
 Coordinator
Sun Microsystems
John West, Blogs And Podcasts
US Army ERDC
Kathryn Kelley, VIP Tour Organizer
Ohio Supercomputer Center

Publicity And Features Team

Doug Black, *Libra Strategic Communications*
Karen Green, *Renaissance Computing Institute*
John Kirkley, *Kirkley Communications*
Jimme Peters, *24/7 Consulting*
John West, *US Army ERDC*

Writing / Editing Team

Doug Black, *Libra Strategic Communications*
Don Johnston, *Lawrence Livermore National Laboratory*
Kathryn Kelley, *Ohio Supercomputer Center*

John Kirkley, *Kirkley Communications*
Faith Singer-Villalobos, *Texas Advanced Computing Center*
John West, *US Army ERDC*

Broadcast Media

Mike Bernhardt, *Libra Strategic Communications*
Heidi Olson, *Olson & Associates*
Rich Brueckner, *Sun Microsystems*

Media Room Support

Angela Harris, *DOD HPC Modernization Program*
Don Johnston, *Lawrence Livermore National Laboratory*
Jimme Peters, *24/7 Consulting*

Concierge Support / Press Escorts

Rick Friedman, *Terascala*
Don Johnston, *Lawrence Livermore National Laboratory*
Jimme Peters, *24/7 Consulting*
John West, *US Army ERDC*

Design Contractors

Carlton Bruett, *Carlton Bruett Design*
Chris Hamb, *Carlton Bruett Design*

SC Communities

Cherri M. Pancake, Chair
Oregon State University
Jennifer Teig Von Hoffman, Deputy
 Chair
Boston University

Broader Engagement

Jennifer Teig Von Hoffman, Chair
Boston University

Broader Engagement Committee

Sadaf R. Alam, *Swiss National Supercomputing Centre*
Tony Baylis, *Lawrence Livermore National Laboratory*
Jason T. Black, *Florida A&M University*
Bonnie L. Bracey-Sutton, *George Lucas Educational Foundation*
Jamika D. Burge, *Pennsylvania State University*

Hongmei Chi, *Florida A&M University*
Charles Hardnett, *Spelman College*
Raquell Holmes, *University of Connecticut*
Mark A. Jack, *Florida A&M University*
Kathryn Kelley, *Ohio Supercomputer Center*
Brandeis H. Marshall, *Purdue University*
Yolanda Rankin, *IBM Almaden Research Center*
Cindy Sievers, *Los Alamos National Laboratory*
Tiki L. Suarez-Brown, *Florida A&M University*

SC Ambassadors

Ron Perrott, Chair
Queen's University, Belfast

SC Ambassadors Committee

David Abramson, *Monash University*
Bernd Mohr, *Juelich Supercomputing Centre*
Depei Qian, *Beihang University*

Sc Your Way

Beverly Clayton, Chair
Pittsburgh Supercomputing Center

Committee

L. Eric Greenwade, *Microsoft Corporation*
Heidi Lambek, *Intel Corporation*
Tad Reynales, *Calit2*
Adam Ryan, *Oregon State University*

Student Volunteers

Paul Fussell, Student Volunteers
 Co-Chair
Boeing

Barbara A. Kucera, Student Volunteers
 Co-Chair
University of Kentucky

Student Volunteers Committee

Tony Baylis, *Lawrence Livermore National Laboratory*
Barbara Fossum, *Purdue University*
Timothy Leite, *Visual Numerics, Inc.*
Bruce Loftis, *University of Tennessee, Knoxville*
Kate Mace, *Clemson University*
Diglio Simoni, *RTI International*
Kathy R. Traxler, *Louisiana State*

University

Caroline Carver Weilhamer, *Clemson University*
Robert Whitten Jr., *Oak Ridge National Laboratory*

Education Program

Scott Lathrop, Education Co-Chair
University of Chicago
Laura McGinnis, Education Co-Chair
Pittsburgh Supercomputing Center

Education Logistics

Dalma Hadziomerovic, *Krell Institute*
Michelle King, *Krell Institute*

Education Outreach

Tom Murphy, *Contra Costa College*
Robert M. Panoff, *Shodor*

Education Evaluation

Gypsy Abbott, *University of Alabama at Birmingham*

Education Communication

Mary Ann Leung, *Krell Institute*

Education Financials

Laura McGinnis, *Pittsburgh Supercomputing Center*

Education Special Projects

Charles Peck, *Earlham College*

Education Webmaster

Kristina (Kay) Wanous, *Earlham College*

Booth

Henry J. Neeman, *University of Oklahoma*
Kristina (Kay) Wanous, *Earlham College*

Student Programming Contest

David A. Joiner, *Kean University*
Tom Murphy, *Contra Costa College*
Charles Peck, *Earlham College*
Kristina (Kay) Wanous, *Earlham College*

Awards

Mary Ann Leung, *Krell Institute*

Robert M. Panoff, *Shodor*

Susan J. Ragan, *Maryland Virtual High School*

Technology Infrastructure

Andrew Fitz Gibbon, *Earlham College*
Kevin M. Hunter, *Earlham College*

Signage

Diane A. Baxter, *San Diego Supercomputer Center*

Student Volunteers

James Rome, *Oak Ridge National Laboratory*

Curriculum Resource Fair

Masakatsu Watanabe, *University of California, Merced*

Office

Dalma Hadziomerovic, *Krell Institute*
Michelle King, *Krell Institute*

Program Coordinators

Richard Gass, *University of Cincinnati*
Steven Gordon, *Ohio Supercomputer Center*

Summer Workshops

Workshop Coordinators

David A. Joiner, *Kean University*
Henry J. Neeman, *University of Oklahoma Chemistry*
Shawn Sendlinger, *North Carolina Central University*

Physics

Richard Gass, *University of Cincinnati*

Biology

Jeff Krause, *Shodor Education Foundation*

Computational Thinking

Robert M. Panoff, *Shodor*

Parallel Programming

Charles Peck, *Earlham College*

Math

Dan Warner, *Clemson University*

Engineering

Steven Gordon, *Ohio Supercomputer Center*

Conference Databases

Jeremiah Konkle, *Linklings LLC*

John Konkle, *Linklings LLC*

Mark Montague, *Linklings LLC*

Finance

Jeffrey K. Hollingsworth, Finance Chair
University of Maryland

Janet Mccord, Registration
University of Texas at Austin

Michele Bianchini-Gunn, Store
Lawrence Livermore National Laboratory

Ralph A. McEldowney, Scinet Chair
Air Force Research Laboratory

Beverly Clayton, Finance Committee
Pittsburgh Supercomputing Center

Finance Contractor

Donna Martin, *Talbot, Korvola & Warwick, LLP*

Anne Nottingham, *Talbot, Korvola & Warwick, LLP*

Brad Rafish, *Talbot, Korvola & Warwick, LLP*

Technical Program Committee

Applications

Esmond G Ng, Tech Papers Applications Area Chair
Lawrence Berkeley National Laboratory

Tech Papers Applications Area Committee

Takayuki Aoki, *Tokyo Institute of Technology*

William L. Barth, *Texas Advanced Computing Center*

George Biros, *Georgia Institute of Technology*

Edmond Chow, *D.E. Shaw Research*

Eduardo Francisco D'azevedo, *Oak Ridge National Laboratory*

Tony Drummond, *Lawrence Berkeley National Laboratory*

Anne C. Elster, *Norwegian University of Science & Technology*

Scott Emrich, *University of Notre Dame*
George I. Fann, *Oak Ridge National Laboratory*

Luc Giraud, *University of Toulouse*

Naga Govindaraju, *Microsoft Corporation*

Peter Graf, *National Renewable Energy Laboratory*

Laura Grigori, *INRIA*

Francois Gygi, *University of California, Davis*

Charles Hansen, *University of Utah*

Michael A Heroux, *Sandia National Laboratories*

Lie-Quan Lee, *SLAC National Accelerator Laboratory*

Hideo Matsuda, *Osaka University*

Kengo Nakajima, *University of Tokyo*

John Owens, *University of California, Davis*

Subhash Saini, *NASA Ames Research Center*

Ashok Srinivasan, *Florida State University*

Richard Vuduc, *Georgia Institute of Technology*

Lin-Wang Wang, *Lawrence Berkeley National Laboratory*

Gerhard Wellein, *Erlangen Regional Computing Center*

Theresa L. Windus, *Iowa State University*

Architecture/Networks

Keith Underwood, Tech Papers Architecture/Network Area Chair
Intel Corporation

Tech Papers Architecture/Network Area Committee

Dennis Abts, *Google*

Jung Ho Ahn, *Seoul National University*

Keren Bergman, *Columbia University*

Ron Brightwell, *Sandia National Laboratories*

Darius Buntinas, *Argonne National Laboratory*

John B. Carter, *IBM Corporation*

Ada Gavrilovska, *Georgia Institute of Technology*

Patrick Geoffray, *Myricom*

William Harrod, *DARPA*

Scott Hemmert, *Sandia National Laboratories*

Doug Joseph, *IBM Corporation*

Partha Kundu, *Intel Corporation*

Shubu Mukherjee, *Intel Corporation*

Hiroshi Nakashima, *Kyoto University*

Vivek S Pai, *Princeton University*

Scott Pakin, *Los Alamos National Laboratory*

Steve Parker, *NVIDIA*

Steve Scott, *Cray Inc.*

Kelly Shaw, *University Of Richmond*

Takeshi Shimizu, *Fujitsu Laboratories Ltd.*

Federico Silla, *Universidad Politecnica De Valencia*

Brian Towles, *D.E. Shaw Research*

Grids

Philip M. Papadopoulos, Tech Papers Grids Area Chair
San Diego Supercomputer Center

Tech Papers Grids Area Committee

David Abramson, *Monash University*
Henrique Andrade, *IBM T.J. Watson Research Center*

Henri Casanova, *University of Hawaii at Manoa*

Ann L. Chervenak, *University of Southern California*

Susumu Date, *Osaka University*
Narayan Desai, *Argonne National Laboratory*

Geoffrey Fox, *Indiana University*

Thomas Hacker, *Purdue University*

Weicheng Huang, *National Center for High-Performance Computing Taiwan*

Kate Keahey, *Argonne National Laboratory*

Thilo Kielmann, *Vrije Universiteit*

Laurent Lefevre, *INRIA*

Shava Smullen, *San Diego Supercomputer Center*

Yoshio Tanaka, *National Institute of Advanced Industrial Science & Technology*
Todd Tannenbaum, *University of Wisconsin-Madison*

Douglas Thain, *University of Notre Dame*

Rich Wolski, *University of California, Santa Barbara*

Performance

Xian-He Sun, Tech Papers Performance Area Chair
Illinois Institute of Technology

Tech Papers Performance Area Committee

Surendra Byna, *NEC Laboratories America*
Kirk Cameron, *Virginia Tech*
Thomas Fahringer, *University of Innsbruck*
John L. Gustafson, *Intel Corporation*
Adolfy Hoisie, *Los Alamos National Laboratory*
Donald Holmgren, *Fermi National Laboratory*
Chung-Ta King, *Tsinghua University*
David Lowenthal, *University of Arizona*
Allen Malony, *University of Oregon*
Bernd Mohr, *Juelich Supercomputing Centre*
Depei Qian, *Beihang University*
Padma Raghavan, *Pennsylvania State University*
Rajeev Thakur, *Argonne National Laboratory*
Jeffrey Vetter, *Oak Ridge National Laboratory*
David William Walker, *Cardiff University*
Cheng-Zhong Xu, *Wayne State University*
Zhiwei Xu, *Institute of Computing Technology*
Xiaodong Zhang, *Ohio State University*

System Software

Franck Cappello, Tech Papers System Software Area Chair
INRIA
Mitsuhiro Sato, Tech Papers System Software Area Chair
University of Tsukuba

Tech Papers System Software Area Committee

Rosa M. Badia, *Barcelona Supercomputing Center*
Peter Beckman, *Argonne National Laboratory*
George Bosilca, *University of Tennessee, Knoxville*
Greg Bronevetsky, *Lawrence Livermore National Laboratory*
Barbara Chapman, *University of Houston*
Toni Cortes, *Barcelona Supercomputing Center*
Bronis R. De Supinski, *Lawrence*

Livermore National Laboratory
Guang R. Gao, *University of Delaware*
Al Geist, *Oak Ridge National Laboratory*

Rich Graham, *Oak Ridge National Laboratory*
Costin Iancu, *Lawrence Berkeley National Laboratory*
Yutaka Ishikawa, *University of Tokyo*
Hai Jin, *Huazhong University of Science & Technology*
Julien Langou, *University of Colorado Denver*
Arthur Maccabe, *Oak Ridge National Laboratory*
Naoya Maruyama, *Tokyo Institute of Technology*
Fabrizio Petrini, *IBM T.J. Watson Research Center*
Keshav Pingali, *University of Texas at Austin*
Michael M. Resch, *High Performance Computing Center Stuttgart*
Olivier Richard, *Universite Joef Fourier*
Vivek Sarkar, *Rice University*
Thomas Lawrence Sterling, *Louisiana State University*
Shinji Sumimoto, *Fujitsu Laboratories Ltd.*
Michela Tafer, *University of Delaware*
Jesper Traeff, *NEC Laboratories Europe*

Storage

Joel Saltz, Tech Papers Storage Area Chair
Emory University

Tech Papers Storage Area Committee

Ghaleb Abdulla, *Lawrence Livermore National Laboratory*
Bill Allcock, *Argonne National Laboratory*
Alok Choudhary, *Northwestern University*
Garth Gibson, *Carnegie Mellon University*
Scott Klasky, *Oak Ridge National Laboratory*
Tahsin Kurc, *Emory University*
Ron Oldfield, *Sandia National Laboratories*
Karsten Schwan, *Georgia Institute of Technology*
Osamu Tatebe, *University of Tsukuba*
Mustafa Uysal, *HP Labs*

Tutorials Committee

Fred Johnson, *Tutorials Chair*

Consultant

Rajeev Thakur, *Tutorials Chair*
Argonne National Laboratory

Tutorials Committee

Pavan Balaji, *Argonne National Laboratory*
Alok Choudhary, *Northwestern University*
Almadena Chtchelkanova, *National Science Foundation*
John W. Cobb, *Oak Ridge National Laboratory*
Jack Dongarra, *University of Tennessee, Knoxville*
Ganesh Gopalakrishnan, *University of Utah*
Rama Govindaraju, *Google*
Yutaka Ishikawa, *University of Tokyo*
Alice E. Koniges, *Lawrence Berkeley National Laboratory*
Dieter Kranzlmüller, *Ludwig-Maximilians-University Munich*
Jeffery A. Kuehn, *Oak Ridge National Laboratory*
Zhiling Lan, *Illinois Institute of Technology*
Dr. Chockchai Leangsuksun, *Louisiana Tech University*
Kwan-Liu Ma, *University of California, Davis*
Xiaosong Ma, *North Carolina State University*
Manish Parashar, *Rutgers University*
Robert B. Ross, *Argonne National Laboratory*
Martin Schulz, *Lawrence Livermore National Laboratory*
Stephen F. Siegel, *University of Delaware*
Lauren L. Smith, *National Security Agency*
Thomas Lawrence Sterling, *Louisiana State University*
Craig B. Stunkel, *IBM T.J. Watson Research Center*
Naoki Sueyasu, *Fujitsu*
Jesper Traeff, *NEC Laboratories Europe*
Putchong Uthayopas, *Kasetsart University*
Harvey Wasserman, *Lawrence Berkeley National Laboratory*

Birds-Of-A-Feather
Toni Cortes, *Birds-Of-A-Feather Chair*
Barcelona Supercomputing Center

Birds-Of-A-Feather Committee

George Almasi, *IBM Corporation*
Daniel S. Katz, *University of Chicago*
Karl W. Schulz, *University of Texas at Austin*
Eno Thereska, *Microsoft Corporation*
Doctoral Showcase
Padma Raghavan, Doctoral Showcase
 Chair
Pennsylvania State University

Doctoral Showcase Committee
George Biros, *Georgia Institute of Technology*
Umit Catalyurek, *Ohio State University*
Jing-Ru Cheng, *U.S. Army ERDC*
Paul Hovland, *Argonne National Laboratory*
X. Sherry Li, *Lawrence Berkeley National Laboratory*
Suzanne Shontz, *Pennsylvania State University*

Posters

Jack Dongarra, Posters Chair
University of Tennessee, Knoxville

Posters Algorithms Area Chair
Rob Schreiber, *HP Labs*

Posters Algorithms Committee
David Abramson, *Monash University*
David A. Bader, *Georgia Institute of Technology*
Mark A. Baker, *University of Reading*
John Gilbert, *University of California, Santa Barbara*
Martyn Guest, *Cardiff University*
Erik Hagersten, *Uppsala University*
Piotr Luszczek, *University of Tennessee, Knoxville*
Raymond Namyst, *INRIA*
Rob Schreiber, *HP Labs*
Stan Scott, *Queen's University, Belfast*
Alan Sussman, *University of Maryland*
Patrick Haven Worley, *Oak Ridge National Laboratory*

Posters Applications Area Chair
David William Walker, *Cardiff University*

Posters Applications Committee
David Abramson, *Monash University*

Mark A. Baker, *University of Reading*
Martyn Guest, *Cardiff University*
Stan Scott, *Queen's University, Belfast*
Patrick Haven Worley, *Oak Ridge National Laboratory*
Posters Architecture Area Chair
Taisuke Boku, *University of Tsukuba*

Posters Architecture Committee
Wu Feng, *Virginia Tech*
Tomohiro Kudoh, *National Institute of Advanced Industrial Science & Amp; Technology*
Andres Marquez, *Pacific Northwest National Laboratory*
Hiroshi Nakamura, *University of Tokyo*
Lloyd W Slonaker, Jr., *AFRL DOD Supercomputing Resource Center*

Posters Software Area Chair
Frederic Desprez, *INRIA*

Posters Software Committee
Eddy Caron, *ENS Lyon*
Benjamin Depardon, *ENS Lyon*
Frederic Desprez, *INRIA*
Vladimir Getov, *University of Westminster*
Hidemoto Nakada, *National Institute of Advanced Industrial Science & Amp; Technology*
Martin Quinson, *LORIA*
Keith Seymour, *University of Tennessee, Knoxville*

Challenges

Manuel Vigil, Challenges Chair
Los Alamos National Laboratory

Bandwidth Challenge
Stephen Q. Lau, Bandwidth Challenge
 Co-Chairs
University of California, San Francisco

Kevin Walsh, Bandwidth Challenge
 Co-Chairs
University of California, San Diego

Bandwidth Challenge Committee
Greg Goddard, *Strangeloop Networks*
Debbie Montano, *Juniper Networks*

Storage Challenge

Raymond L. Paden, Storage Challenge
 Chair
IBM Corporation

Alan Sussman, Storage Challenge Co-Chair
University of Maryland

Storage Challenge Committee
Randy Kreiser, *Data Direct Networks*
Charles Lively, *Texas A&M University*
Thomas Ruwart, *I/O Performance Inc.*
John R. Sopka, *EMC*
Virginia To, *VTK Solutions LLC*

Student Cluster Competition
Jeanine Cook, Co-Chair
New Mexico State University
George (Chip) Smith, Co-Chair
Lawrence Berkeley National Laboratory

Student Cluster Competition Committee
Bob Beck, *University of Alberta*
Sam Coleman, *Retired*
Thomas Deboni, *Retired*
Douglas Fuller, *Arizona State University*
Earl Malmrose, *Zareason.Com*
Hai Ah Nam, *Oak Ridge National Laboratory*
Tom Spelce, *Lawrence Livermore National Laboratory*

Scinet

Ralph A. McEldowney, Scinet Chair
Air Force Research Laboratory
Jamie Van Randwyk, Scinet Deputy
 Chair
Sandia National Laboratories
Tracey D. Wilson, Scinet Vice Chair
Computer Sciences Corporation

Eli Dart, Scinet Executive Director
Energy Sciences Network

Architecture
Charles D. Fisher, Architecture Chair
Oak Ridge National Laboratory

Commodity Network
Rex Duncan, Commodity Network Chair

*Oak Ridge National Laboratory***Commodity Network Committee****Mike Beckman**, *US Army Space & Missile Defense Command***Jeffrey Richard Schwab**, *Purdue University***Bandwidth Challenge****Stephen Q. Lau**, Bandwidth Challenge Co-Chair*University of California, San Francisco***Kevin Walsh**, Bandwidth Challenge Co-Chair*University of California, San Diego***Bandwidth Challenge Committee****Greg Goddard**, *Strangeloop Networks*Debbie Montano, *Juniper Networks***WAN Transport****David Crowe**, Wan Transport Chair
*Network for Education and Research in Oregon***WAN Transport Committee****Ben Blundell**, *Louisiana State University*Chris Costa, *CENIC***Ken Goodwin**, *Pittsburgh Supercomputing Center***Chris Griffin**, *Florida LambdaRail***Bonnie Hurst**, *Renaissance Computing Institute***Bill Jensen**, *University of Wisconsin-Madison***Akbar Kara**, *Lonestar Education & Research Network***Lonnie Leger**, *Louisiana State University***Ron Milford**, *Indiana University***Dave Pokorney**, *Florida LambdaRail***Ryan Vaughn**, *Florida LambdaRail***Wayne Wedemeyer**, *University of Texas System***Kenny Welshons**, *Louisiana State University***Matthew J Zekauskas**, *Internet2***IT Services****Martin Swany**, IT Services Chair
*University of Delaware***IT Services Committee****Jon Dugan**, *ESNET***Guilherme Fernandes**, *University of Delaware***John Hoffman**, *Air Force Research Laboratory***Ezra Kissel**, *University of Delaware***Reed Martz**, *University of Delaware***Davey Wheeler**, *National Center For Supercomputing Applications***Logistics****Ken Brice**, Logistics Co-Chairs*Army Research Laboratory***John R. Carter**, Logistics Co-Chairs*Air Force Research Laboratory***Logistics Committee****Bob Williams****Equipment****Jeffery Alan Mauth**, Equipment Chair
*Pacific Northwest National Laboratory***Help Desk****Caroline Carver Weilhamer**, Help Desk Chair
*Clemson University***Help Desk Committee****Kate Mace**, *Clemson University***David Stewart**, *Lawrence Berkeley National Laboratory***Fiber****Warren Birch**, Fiber Co-Chairs*Army Research Laboratory***Annette Kitajima**, Fiber Co-Chairs*Sandia National Laboratories***Fiber Committee****Virginia Bedford**, *Arctic Region Supercomputing Center***John Christman**, *ESNET***Da Fye**, *Oak Ridge National Laboratory***Zachary Giles**, *Oak Ridge National Laboratory***Kevin R. Hayden**, *Argonne National Laboratory***Lance V. Hutchinson**, *Sandia National Laboratories***Jay Krous**, *Lawrence Berkeley National Laboratory***Mark Schultz**, *Air Force Research Laboratory*

Oliver Seuffert,

David Stewart, *Lawrence Berkeley National Laboratory***Measurement****Jeff W. Boote**, Measurement Chair
*Internet2***Measurement Committee****Aaron Brown**, *Internet2***Prasad Calyam**, *Ohio Supercomputer Center***Richard Carlson**, *Internet2***Scot Colburn**, *NOAA***Jon Dugan**, *ESNET***John Hicks**, *Indiana University***Russ Hobby**, *Internet2***Stuart Johnston**, *InMon Corporation***Loki Jorgenson**, *Apparent Networks***Neil H. McKee**, *InMon Corporation***Sonia Panchen**, *InMon Corporation***Jason Zurawski**, *University of Delaware***Network Security****Scott Campbell**, Network Security Co-Chairs*Lawrence Berkeley National Laboratory***Carrie Gates**, Network Security Co-Chairs*CA Labs***Aashish Sharma**, *National Center For Supercomputing Applications***Network Security Committee****James Hutchins**, *Sandia National Laboratories***John Johnson**, *Pacific Northwest National Laboratory***Jason Lee**, *Lawrence Berkeley National Laboratory***Tamara L. Martin**, *Argonne National Laboratory***Joe Mcmanus**, *CERT***Jim Mellander**, *Lawrence Berkeley National Laboratory***Scott Pinkerton**, *Argonne National Laboratory***Elliot Proebstel**, *Sandia National Laboratories***Christopher Roblee**, *Lawrence Livermore*

*National Laboratory***Piotr Zbiegiel**, *Argonne National Laboratory***Physical Security****Jeff Graham**, Physical Security Chair
*Air Force Research Laboratory***Physical Security Committee****Stephen Q. Lau**, *University of California, San Francisco***X-Net****E. Paul Love**, X-Net Co-Chairs
Internet Consulting of Vermont
Rod Wilson, X-Net Co-Chairs
*Nortel***X-Net Committee****Akbar Kara**, *Lonestar Education & Research Network***Routing****Linda Winkler**, Routing Chair
*Argonne National Laboratory***Routing Committee****Eli Dart**, *Energy Sciences Network*
Pieter De Boer, *Sara Computing & Networking Services*
Thomas Hutton, *San Diego Supercomputer Center*
Andrew Lee, *Indiana University*
Craig Leres, *Lawrence Berkeley National Laboratory*
Corby Schmitz, *Argonne National Laboratory*
Brent Sweeny, *Indiana University*
JP Velders, *University of Amsterdam*
Alan Verlo, *University of Illinois at Chicago***Wireless****Mark Mitchell**, Wireless Chair
*Sandia National Laboratories***Power****William R. Wing**, Power Chair
*Darkstrand***Open Fabrics****Eric Dube**, Open Fabrics Co-Chairs
Bay Microsystems
Cary Whitney, Open Fabrics Co-Chairs*Lawrence Berkeley National Laboratory***Open Fabrics Committee****Tom Ammon**, *University of Utah*
Troy R. Benjegerdes, *Ames Laboratory*
Lutfor Bhuiyan, *Intel Corporation*
Kelly Cash, *QLogic*
Rupert Dance, *Lamprey Networks*
Matt Davis, *Zarlink Semiconductor*
Marc Dizoglio, *RAID, Inc.*
Parks Fields, *Los Alamos National Laboratory*
Jason Gunthorpe, *Obsidian Strategies*
Sean Hafeez, *Arastra*
Ercan Kamber, *RAID, Inc.*
Bill Lee, *Mellanox Technologies*
Scott Martin, *Fortinet, Inc.*
Kevin Mcgrattan, *Cisco Systems*
Makia Minich, *Sun Microsystems*
Adithyaram Narasimha, *Luxtera*
John Petrilla, *Avago Technologies*
Jad Ramey, *RAID, Inc.*
Hal Rosenstock, *Obsidian Strategies*
Scot Schultz, *AMD*
Gilad Shainer, *Mellanox Technologies*
Tom Stachura, *Intel Corporation*
Kevin Webb, *RAID, Inc.*
Beth Wickenhiser, *Voltaire*
Mitch Williams, *Sandia National Laboratories*
Tracey D. Wilson, *Computer Sciences Corporation*
Communications
Lauren Rotman, Communications Chair
*Internet2***Communications Committee****William Kramer**, *National Center For Supercomputing Applications***Vendor Support****Vendor Liaisons****James Archuleta**, *Ciena*
David Boertjes, *Nortel*
Christopher Carroll, *Force10 Networks*
Shawn Carroll, *Qwest Communications*
Niclas Comstedt, *Force10 Networks*
George Delisle, *Force10 Networks*
Fred Finlay, *Infinera*
Rick Hafey, *Infinera*
Rob Jaeger, *Juniper Networks***John Lankford**, *Ciena*
Robert Marcoux, *Juniper Networks*
Kevin Mcgrattan, *Cisco Systems*
Casey Miles, *Brocade*
Debbie Montano, *Juniper Networks*
Lakshmi Pakala, *Juniper Networks*
Scott Pearson, *Brocade*
Manjuel Robinson, *Fujitsu*
Bill Ryan, *Brocade*
Jason Strayer, *Cisco Systems*
John Walker, *Infinera Corporation***Interconnect****Patrick Dorn**, Interconnect Chair
*National Center For Supercomputing Applications***Sustainability****Troy R. Benjegerdes**, Sustainability Chair
*Ames Laboratory***Student Volunteers****Kate Mace**, Student Volunteers Chair
*Clemson University***Student Volunteers Committee Member****Caroline Carver Weilhamer**, *Clemson University***Steering Committee****Patricia J. Teller**, Chair, *University of Texas at El Paso*
David H. Bailey, *Lawrence Berkeley National Laboratory*
Donna Cappo, *ACM*
Dona Crawford, *Lawrence Livermore National Laboratory*
Candace Culhane, *DOD*
Jack Dongarra, *University of Tennessee, Knoxville*
Lynne Harris, *IEEE Computer Society*
Barry V. Hess, *Sandia National Laboratories*
Jeffrey K. Hollingsworth, *University of Maryland*
Anne Marie Kelly, *IEEE Computer Society*
Scott Lathrop, *University of Chicago*
Satoshi Matsuoka, *Tokyo Institute of Technology*

Wilfred Pinfold, *Intel Corporation*
Daniel A. Reed, *Microsoft Corporation*
James H. Rogers, *Oak Ridge National Laboratory*
Rob Schreiber, *HP Labs*
Jennifer Teig Von Hoffman, *Boston University*
Becky Verastegui, *Oak Ridge National Laboratory*
Laurence T. Yang, *St. Francis Xavier University, Canada*

Steering Committee Admins

Linda Duncan, *Oak Ridge National Laboratory*
Michele Klammer, *Microsoft Corporation*
Katie Thomas, *Lawrence Livermore National Laboratory*
Industry Advisory Committee
Wilfred Pinfold, Chair, *Intel Corporation*
Frank Baetke, *Hewlett-Packard*
Javad Boroumand, *Cisco Systems*
Rich Brueckner, *Sun Microsystems*
Donna Cappel, *ACM*
Denney Cole, *Portland Group*
Phil Fraher, *Visual Numerics, Inc.*
George Funk, *Dell*
Barry V. Hess, *Sandia National Laboratories*
Michael Humphrey, *Altair*
Wes Kaplow, *Qwest Communications*
Anne Marie Kelly, *IEEE Computer Society*
Dennis Lam, *NEC Corporation*
Mike Lapan, *Verari Systems*
Timothy Leite, *Visual Numerics, Inc.*
Karen Makedon, *Platform Computing*
Jay Martin, *Data Direct Networks*
Kevin McGrattan, *Cisco Systems*
Michaela Mezo, *Juniper Networks*
David Morton, *Linux Network*
Takeshi Murakami, *Hitachi, Ltd.*
Raymond L. Paden, *IBM Corporation*
Dorothy Remoquillo, *Fujitsu*
Ellen Roder, *Cray Inc.*
Joan Roy, *SGI*
Daniel Schornak, *CSC*
Raju Shah, *Force10 Networks*
Patricia J. Teller, *University of Texas at*

El Paso

Ed Turkel, *Hewlett-Packard*
Becky Verastegui, *Oak Ridge National Laboratory*

SC09 Sponsoring Societies

ACM

ACM, the Association for Computing Machinery, is the world's largest educational and scientific society uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking. <http://www.acm.org>.

IEEE Computer Society

With nearly 100,000 members, the IEEE Computer Society is the world's leading organization of computer professionals. Founded in 1946, the Computer Society is the largest of the 37 societies of the Institute of Electrical and Electronics Engineers (IEEE).

The Computer Society's vision is to be the leading provider of technical information and services to the world's computing professionals. The Society is dedicated to advancing the theory, practice and application of computer and information processing technology.

[illegible]

SC10 Ad

SC10 Ad

Computing for a Changing World



Portland,
OR | 2009

Sponsors:
ACM SIGARCH/
IEEE Computer Society



Association for
Computing Machinery



IEEE
computer
society

Oregon Convention Center • Portland, Oregon • November 14-20, 2009

<http://sc09.supercomputing.org>