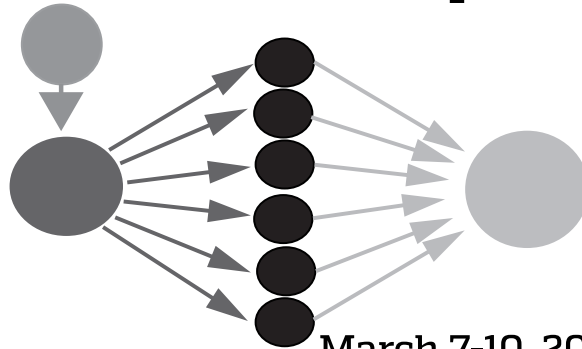


# Final Program and Abstracts

## SIAM Conference on **Parallel Processing for Scientific Computing**



March 7-10, 2018  
Waseda University  
Tokyo, Japan

This conference is sponsored by the  
SIAM Activity Group on Supercomputing  
and the Japan Society for Industrial  
& Applied Mathematics (JSIAM).



The PP18 Event Mobile App

Scan the QR code with any QR reader and download the  
TripBuilder EventMobile™ app to your iPhone, iPad,  
iTouch or Android mobile device.

You can also visit <http://www.tripbuildermedia.com/apps/siampp18>



Society for Industrial and Applied Mathematics

3600 Market Street, 6th Floor

Philadelphia, PA 19104-2688 USA

Telephone: +1-215-382-9800 Fax: +1-215-386-7999

Conference E-mail: [meetings@siam.org](mailto:meetings@siam.org)

Conference Web: [www.siam.org/meetings/](http://www.siam.org/meetings/)

Membership and Customer Service:

(800) 447-7426 (USA & Canada) or +1-215-382-9800 (worldwide)

[www.siam.org/meetings/pp18](http://www.siam.org/meetings/pp18)

## Table of Contents

Program-At-A-Glance...	See separate handout
General Information.....	2
Get-togethers.....	4
Invited Plenary Speakers.....	6
Program Schedule.....	9
Poster Session.....	48
Abstracts.....	91
Speaker and Organizer Index.....	229
Meeting Room Map.....	Back Cover

## Organizing Committee Co-Chairs

### Satoshi Matsuoka

Tokyo Institute of Technology, Japan

### Kengo Nakajima

Information Technology Center, The  
University of Tokyo, Japan

### Olaf Schenk

Universita della Svizzera italiana,  
Switzerland

## Organizing Committee

### Jed Brown

University of Colorado Boulder, United  
States

### Christian Engwer

University of Münster, Germany

### Luc Giraud

Inria, France

### Michael Heroux

Sandia National Laboratories, United  
States

### George Karypis

University of Minnesota, United States

### Harald Köstler

Friedrich-Alexander-Universität  
Erlangen, Germany

### Yutong Lu

National University of Defense  
Technology, China

### Simon McIntosh-Smith

University of Bristol, United Kingdom

### Miriam Mehl

Universität Stuttgart, Germany

### Tetsuya Sakurai

University of Tsukuba, Japan

### Sameer Shende

University of Oregon, United States

### Jeffrey Vetter

Oak Ridge National Laboratory, United  
States

### Weichung Wang

National Taiwan University, Taiwan

### Felix Wolf

Technische Universität Darmstadt,  
Germany

## SIAM Registration Desk

The conference registration desk is  
open 8:00 AM – 5:00 PM, Wednesday,  
March 7 through Saturday, March 10,  
2018.

## Venue Address

All sessions and conference registration  
will take place at Waseda University,  
Tokyo, Japan

Building 63,  
Nishi Waseda Campus,  
Waseda University  
3-4-1 Ookubo, Shinjuku-ku, Tokyo 169-  
8555, Japan

Website: [https://www.waseda.jp/top/en/  
access/nishiwaseda-campus](https://www.waseda.jp/top/en/access/nishiwaseda-campus)

## Child Care

Please contact [siampp18@ml.jsiam.org](mailto:siampp18@ml.jsiam.org)  
for information about the availability of  
child care services.

## Corporate Members and Affiliates

SIAM corporate members provide  
their employees with knowledge about,  
access to, and contacts in the applied  
mathematics and computational sciences  
community through their membership  
benefits. Corporate membership is more  
than just a bundle of tangible products  
and services; it is an expression of  
support for SIAM and its programs.  
SIAM is pleased to acknowledge its  
corporate members and sponsors.  
In recognition of their support, non-  
member attendees who are employed by  
the following organizations are entitled  
to the SIAM member registration rate.

### Corporate/Institutional Members

The Aerospace Corporation

Air Force Office of Scientific Research

Amazon

Aramco Services Company

Bechtel Marine Propulsion Laboratory

The Boeing Company

CEA/DAM

Department of National Defence (DND/  
CSEC)

DSTO- Defence Science and  
Technology Organisation

Hewlett-Packard

Huawei FRC French R&D Center

IBM Corporation

IDA Center for Communications  
Research, La Jolla

IDA Center for Communications  
Research, Princeton

Institute for Defense Analyses, Center  
for Computing Sciences

Lawrence Berkeley National Laboratory

Lawrence Livermore National Labs

Lockheed Martin

Los Alamos National Laboratory

Max-Planck-Institute for Dynamics of  
Complex Technical Systems

Mentor Graphics

National Institute of Standards and  
Technology (NIST)

National Security Agency (DIRNSA)  
Naval PostGrad

Oak Ridge National Laboratory,  
managed by UT-Battelle for the  
Department of Energy

Sandia National Laboratories

Schlumberger-Doll Research

United States Department of Energy

U.S. Army Corps of Engineers,  
Engineer Research and Development  
Center

US Naval Research Labs

*List current December 2017.*

## Funding Agency

SIAM and the Conference Organizing Committee wish to extend their thanks and appreciation to the U.S. National Science Foundation for its support of this conference.



**Leading the applied mathematics community**  
***Join SIAM and save!***

Join your peers in supporting the premier professional society for applied mathematicians and computational scientists. SIAM members receive subscriptions to SIAM Review, SIAM News and SIAM Unwrapped, and enjoy substantial discounts on SIAM books, journal subscriptions, and conference registrations.

Free Student Memberships are available to students who attend an institution that is an Academic Member of SIAM, are members of Student Chapters of SIAM, or are nominated by a Regular Member of SIAM.

Join online today at [www.siam.org/joinsiam](http://www.siam.org/joinsiam) or contact SIAM Customer Service:

Telephone: +1-215-382-9800  
(worldwide); or 800-447-7426 (U.S. and Canada only)

Fax: +1-215-386-7999

E-mail: [membership@siam.org](mailto:membership@siam.org)

Postal mail: Society for Industrial and Applied Mathematics, 3600 Market Street, 6th floor, Philadelphia, PA 19104-2688 USA

## Standard Audio/Visual Set-Up in Meeting Rooms

SIAM does not provide computers for any speaker. When giving an electronic presentation, speakers must provide their own computers. SIAM is not responsible for the safety and security of speakers' computers. The following equipment will be available:

Building 52	HDMI and VGA
Building 57	VGA (D-sub 15pin)
Building 63	VGA (D-sub 15pin)

## Internet Access

University wifi helps you get connected on your laptop, mobile or tablet. Eduroam wifi can also be used if your home institution provides it.

## Important Notice to Poster Presenters

The poster session is scheduled for March 8 at 7:15 PM. Poster presenters are expected to set up their poster material on the provided 180cm x 120cm poster boards in Room 63-1F. Posters must be removed by 6 pm on March 10.

## SIAM Books and Journals

SIAM books are available at a discounted price during the conference. Complimentary copies of journals

are also available on site. If a SIAM books representative is not available, orders may be placed according to the instructions on the Titles on Display form.

## Industry Exhibitors

The conference organizers thank the following industry exhibitors for their generous support.

### Platinum Status

DDN Storage  
HP Enterprise  
FUJITSU  
UBM  
Intel  
Mellanox Technologies

### Silver Status

HPC Tech  
NVIDIA  
arm

## Registration Fee Includes

- Admission to all technical sessions
- Business Meeting
- Coffee breaks daily
- Room set-ups and audio/visual equipment
- Welcome Reception and Poster Session

## Comments?

Comments about SIAM meetings are encouraged! Please send to:

Cynthia Phillips, SIAM Vice President for Programs (vpp@siam.org).

## Get-togethers

- **Welcome Reception**  
Wednesday, March 7, 6:15 – 8:00 PM
- **Business Meeting**  
Thursday, March 8, 6:30 – 7:15 pm
- **Poster Session**  
Thursday, March 8, 7:15 – 9:15 PM
- **Conference Banquet** (*A separate fee of 3,000 JPY is required to attend.*)  
Friday, March 9, 7:00 – 9:00 PM  
Event will be held at the Hotel Chinzanso Tokyo  
If you are interested in purchasing a ticket to attend, please see the registration desk located in room 63-1F.

## Statement on Inclusiveness

As a professional society, SIAM is committed to providing an inclusive climate that encourages the open expression and exchange of ideas, that is free from all forms of discrimination, harassment, and retaliation, and that is welcoming and comfortable to all members and to those who participate in its activities. In pursuit of that commitment, SIAM is dedicated to the philosophy of equality of opportunity

and treatment for all participants regardless of gender, gender identity or expression, sexual orientation, race, color, national or ethnic origin, religion or religious belief, age, marital status, disabilities, veteran status, field of expertise, or any other reason not related to scientific merit. This philosophy extends from SIAM conferences, to its publications, and to its governing structures and bodies. We expect all members of SIAM and participants in SIAM activities to work towards this commitment.

## Please Note

SIAM is not responsible for the safety and security of attendees' computers. Do not leave your personal electronic devices unattended. Please remember to turn off your cell phones and other devices during sessions.

## Recording of Presentations

Audio and video recording of presentations at SIAM meetings is prohibited without the written permission of the presenter and SIAM.

## Social Media

SIAM is promoting the use of social media, such as Facebook and Twitter, in order to enhance scientific discussion at its meetings and enable attendees to connect with each other prior to, during and after conferences. If you are tweeting about a conference, please use the designated hashtag to enable other attendees to keep up with the Twitter conversation and to allow better archiving of our conference discussions. The hashtag for this meeting is #SIAMPP18.

SIAM's Twitter handle is @TheSIAMNews.

## Changes to the Printed Program

The printed program and abstracts were current at the time of printing, however, please review the online program schedule (<http://meetings.siam.org/program.cfm?CONFCODE=pp18>) or use the mobile app for the most up-to-date information.

## The PP18 Event Mobile App Powered by TripBuilder®

To enhance your conference experience, we're providing a state-of-the-art mobile app to give you important conference information right at your fingertips. With this TripBuilder EventMobile™ app, you can:

- Create your own custom schedule
- View Sessions, Speakers, Exhibitors and more
- Take notes and export them to your email
- View Award-Winning TripBuilder Recommendations for the meeting location
- Get instant Alerts about important conference info

## The PP18 Event Mobile App

Scan the QR code with any QR reader and download the TripBuilder EventMobile™ app to your iPhone, iPad, iTouch or Android mobile device.

You can also visit <http://www.tripbuildermedia.com/apps/siampp18>



# SIAM Activity Group on ASupercomputing (SIAG/SC)

[www.siam.org/activity/supercomputing](http://www.siam.org/activity/supercomputing)

## A GREAT WAY TO GET INVOLVED!

The SIAM Activity Group on Supercomputing provides a forum for computational mathematicians, computer scientists, computer architects, and computational scientists to exchange ideas on mathematical algorithms and computer architecture needed for high-performance computer systems.

### ACTIVITIES INCLUDE:

- Special Sessions at SIAM meetings
- Biennial conference
- SIAG/Supercomputing Career Prize
- SIAG/Supercomputing Early Career Prize
- SIAG/Supercomputing Best Paper Prize

### BENEFITS OF SIAG/AG MEMBERSHIP:

- Listing in the SIAG's online membership directory
- Additional \$15 discount on registration at the SIAM Conference on Parallel Processing
- Electronic communications....
- Eligibility for candidacy for SIAG/SC office
- Participation in the selection of SIAG/SC officers

### PROCESSING

- Electronic communications about recent developments in your specialty.
- Eligibility for candidacy for SIAG/SC office
- Participation in the selection of SIAG/SC officers

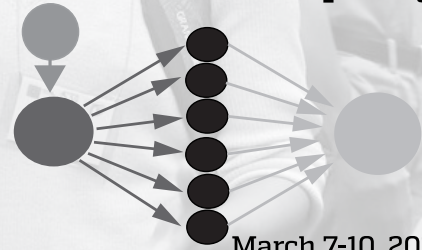
### ELIGIBILITY:

- Be a current SIAM member.

### COST:

- \$15 per year
- Student members can join 2 activity groups for free!

SIAM Conference on  
**Parallel Processing  
for Scientific Computing**



March 7-10, 2018  
Waseda University  
Tokyo, Japan

### 2018-19 SIAG/SC OFFICERS

Chair:	Rich Vuduc, Georgia Institute of Technology
Vice Chair:	Olaf Schenk, Università della Svizzera italiana
Program Director:	Ulrike Meier Yang, Lawrence Livermore National Lab
Secretary:	Matthias Bolten, Bergische Universität Wuppertal

### TO JOIN:

SIAG/AG: [my.siam.org/forms/join\\_siag.htm](http://my.siam.org/forms/join_siag.htm)  
SIAM: [www.siam.org/joinsiam](http://www.siam.org/joinsiam)



## Invited Plenary Speakers

*All invited speaker presentations are scheduled to take place in Room 63-201/202.*

**Wednesday, March 7**

**5:15 PM - 6:00 PM**

**IP1** Verified Numerical Computation and Large Scale Computing

**Shin'ichi Oishi**, *Waseda University, Japan*

**Thursday, March 8**

**8:30 AM - 9:15 AM**

**IP2** Computing Beyond Moore's Law

**John Shalf**, *Lawrence Berkeley National Laboratory, USA*

**1:45 PM - 2:30 PM**

**IP3** Toward Community Software Ecosystems  
for High-Performance Computational Science

**Lois Curfman McInnes**, *Argonne National Laboratory, USA*

## Invited Plenary Speakers

**Friday, March 9**

**8:30 AM - 9:15 AM**

**IP4** Enabling Multi-Peta-Scale Fully Implicit Simulations  
of Atmospheric Dynamics on Sunway TaihuLight

**Haohuan Fu**, *National Supercomputing Center, Wuxi,  
China and Tsinghua University, China*

**Chao Yang**, *Chinese Academy of Sciences, China*

**4:40 PM - 5:25 PM**

**IP5** Fast Scalable Implicit Solver for Low-Ordered Unstructured  
Finite Element Analysis and Its Application

**Tsuyoshi Ichimura**, *University of Tokyo, Japan*

**Saturday, March 10**

**8:30 AM - 9:15 AM**

**IP6** ChainerMN: Scalable Distributed Deep Learning with Chainer

**Takuya Akiba**, *Preferred Networks, Inc., Japan*

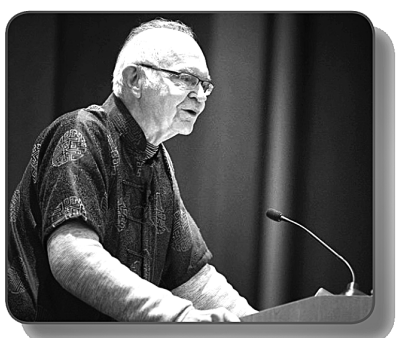
**4:20 PM - 5:05 PM**

**IP7** Communication Hiding Through Pipelining in Krylov Method

**Wim I. Vanroose**, *University of Antwerp, Belgium*



**SIAM PRESENTS IS AN AUDIO-VISUAL ARCHIVE  
COMPRISED OF MORE THAN 2,000 PRESENTATIONS  
POSTED IN OVER 40 SEARCHABLE TOPICS, INCLUDING:**



- algebraic geometry
- atmospheric and oceanographic science
- computational science
- data mining
- geophysical science
- optimization
- uncertainty quantification and more...

The collection, *Featured Lectures from our Archives*, includes audio and slides from more than 30 conferences since 2008, including talks by invited and prize speakers, select minisymposia, and minitutorials. Presentations from SIAM meetings are being added throughout the year.



In addition you can view short video clips of speaker interviews from sessions at Annual Meetings starting in 2010.

Plans for adding more content are on the horizon. Keep an eye out!

The audio, slide, and video presentations are part of SIAM's outreach activities to increase the public's awareness of mathematics and computational science in the real world, and to bring attention to exciting and valuable work being done in the field. Funding from SIAM, the National Science Foundation, and the Department of Energy was used to partially support this project.



New presentations are posted every few months as the program expands with sessions from additional SIAM meetings. Users can search for presentations by category, speaker name, and/or key words.

**[www.siam.org/meetings/presents.php](http://www.siam.org/meetings/presents.php)**



SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS • 3600 MARKET STREET, 6TH FLOOR • PHILADELPHIA, PA 19104-2688 USA  
® PHONE: +1-215-382-9800 • FAX +1-215-386-7999 • [SERVICE@SIAM.ORG](mailto:SERVICE@SIAM.ORG) • [WWW.SIAM.ORG](http://WWW.SIAM.ORG)



# SIAM Books

Visit the SIAM booth to see these and other SIAM books

**CONFERENCE  
ATTENDEES:**  
Get a free t-shirt  
when you buy two  
or more books!

## Barriers and Transport in Unsteady Flows: A Melnikov Approach

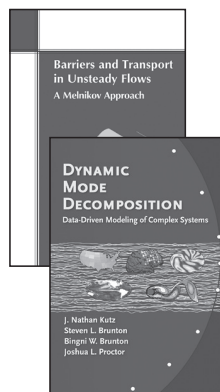
Sanjeeva Balasuriya

*Mathematical Modeling and Computation 21*

How do coherent structures exchange fluid with their surroundings? What is the impact on global mixing? What is the "boundary" of the structure, and how does it move? Can these questions be answered from time-varying observational data? This book addresses these issues from the perspective of the differential equations that must be obeyed by fluid particles. The concepts are illustrated with an array of theoretical and applied examples that arise from oceanography and microfluidics.

2016 • xiv + 264 pages • Softcover • 978-1-611974-57-7

List \$84.00 • Attendee \$67.20 • SIAM Member \$58.80 • MM21



## Differential Dynamical Systems, Revised Edition

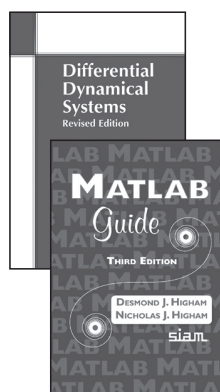
James D. Meiss

*Mathematical Modeling and Computation 22*

This new edition contains several important updates and revisions. It begins with coverage of linear systems, including matrix algebra; the focus then shifts to foundational material on nonlinear differential equations, making heavy use of the contraction-mapping theorem. Subsequent chapters deal specifically with dynamical systems concepts—flow, stability, invariant manifolds, the phase plane, bifurcation, chaos, and Hamiltonian dynamics.

2017 • xviii + 392 pages • Softcover • ISBN 978-1-611974-63-8

List \$87.00 • Attendee \$69.60 • SIAM Member \$60.90 • MM22



## Model Emergent Dynamics in Complex Systems

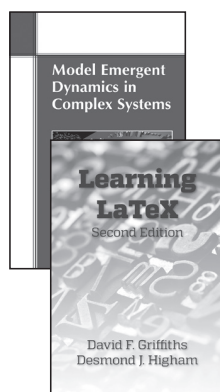
A. J. Roberts

*Mathematical Modeling and Computation 20*

Arising out of the growing interest in and applications of modern dynamical systems theory, this book explores how to derive relatively simple dynamical equations that model complex physical interactions. The authors use sound theory to explore algebraic techniques, develop interesting applications, and discover general modeling principles. The book unifies into one powerful and coherent approach the many varied extant methods for mathematical model reduction and approximation.

2014 • xii + 748 pages • Softcover • 978-1-611973-55-6

List \$114.00 • Attendee \$91.20 • SIAM Member \$79.80 • MM20



## Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems

J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor

The recently developed dynamic mode decomposition (DMD) is an innovative tool for integrating data with dynamical systems theory. The DMD has deep connections with traditional dynamical systems theory and many recent innovations in compressed sensing and machine learning. This is the first book to address the DMD algorithm and it presents a pedagogical and comprehensive approach to all aspects of DMD currently developed or under development.

2016 • xvi + 234 pages • Softcover • 978-1-611974-49-2

List \$69.00 • Attendee \$55.20 • SIAM Member \$48.30 • OT149

## MATLAB Guide, Third Edition

Desmond J. Higham and Nicholas J. Higham

This third edition of *MATLAB Guide* completely revises and updates the best-selling second edition and is more than 25 percent longer. The book remains a lively, concise introduction to the most popular and important features of MATLAB® and the Symbolic Math Toolbox. Key features are a tutorial in Chapter 1 that gives a hands-on overview of MATLAB, a thorough treatment of MATLAB mathematics, including the linear algebra and numerical analysis functions and the differential equation solvers, and a web page that provides example program files, updates, and links to MATLAB resources.

2017 • xxvi + 476 pages • Hardcover • 978-1-611974-65-2

List \$62.00 • Attendee \$49.60 • SIAM Member \$43.40 • OT150

## Learning LaTeX, Second Edition

David F. Griffiths and Desmond F. Higham

"I never would have thought it possible, but the new edition is a substantial improvement with the additional coverage of BiBTeX, Beamer, and posters. Learning LaTeX should be handed to new graduate students in mathematical sciences along with their office key and ID card."

— Don Estep, Colorado State University

2016 • x + 103 pages • Softcover • 978-1-611974-41-6

List \$29.00 • Attendee \$23.20 • SIAM Member \$20.30 • OT148

Prices expire June 25, 2017. All prices are in US dollars.

**Attendees  
get 20% off list.  
SIAM members  
get 30% off.**

## ORDER at [bookstore.siam.org](http://bookstore.siam.org)

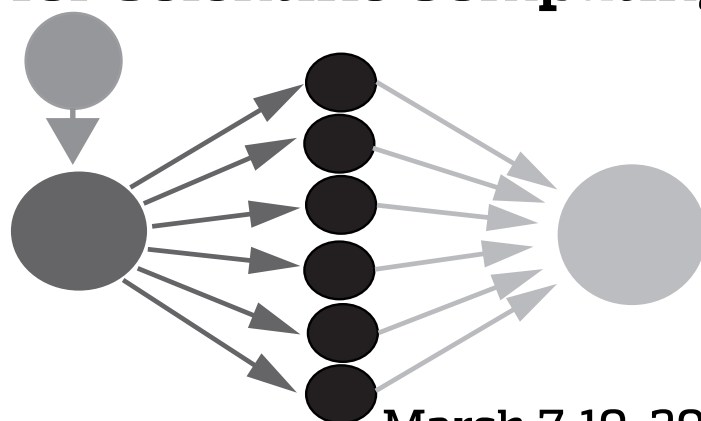
Use your credit card (AMEX, Discover, MasterCard, or VISA) when ordering online, by phone at +1-215-382-9800 worldwide or toll free at 800-447-SIAM in USA and Canada, or by fax at +1-215-386-7999 • Send check or money order to: SIAM, Dept. BKDS17, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688. Please mention "BKDS17" when you order.

**siam** SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS

## Notes

## Program Schedule

SIAM Conference on  
**Parallel Processing  
for Scientific Computing**



March 7-10, 2018  
Waseda University  
Tokyo, Japan

## Wednesday, March 7

### Registration

8:00 AM-5:00 PM

Room:63-1F

## MS1

### Productive Programming using Parallel Models, Tools and Scientific Workflows

1:00 PM-2:40 PM

Room:52-103

This minisymposium will focus on increasing programmers' and users' productivity. This MS will highlight parallelization of Next-Generation Sequencing (NGS) algorithm by applying parallel programming models and establishing a productive bioinformatics workflow thus facilitating biologists to focus more on science and less on programming. To enhance the approach, we will discuss how OpenStack, Dockers, Containers could be used to migrate such parallel tools across platforms and create an easily usable tool with the least intervention necessary. The symposium will also highlight interactive computing and data analysis resources like Jetstream that has a critical impact on the HPC community; not always well-served by traditional systems. Additionally, package manager, SPACK will be presented, which allows application developers to rely on packages that automatically manage/build software requiring several dependency libraries and allow different users to use different versions of software. With hardware evolving so rapidly, we need to embrace techniques that can expedite software production, maintainability, and adaptability.

Organizer: Sunita Chandrasekaran

University of Delaware, USA

#### 1:00-1:20 Achieving Performance While Preserving Portability for NGS Application

Sunita Chandrasekaran, University of Delaware, USA

#### 1:25-1:45 HPC As a Tool for Better Science – The Trend Towards Applications As a Service

Guido Juckeland and Axel Huebl, Helmholtz-Center Dresden-Rossendorf, Germany

#### 1:50-2:10 Decluttering HPC Software Chaos with SPACK

Todd Gamblin, Lawrence Livermore National Laboratory, USA

#### 2:15-2:35 Increasing User and Application Diversity Through Programmable Cyberinfrastructure with Jetstream

David Hancock and Robert Henschel, Indiana University, USA

Wednesday, March 7

## MS2

### State-of-the-Art FFT --- Algorithms, Implementations, and Applications - Part I of II

1:00 PM-2:40 PM

Room:52-104

#### For Part 2 see MS13

The fast Fourier transform (FFT) is widely used in many areas of science and engineering. This minisymposium is an opportunity to discuss high performance and parallel implementations of FFT, mathematical encapsulations of FFT algorithms that are amenable to automatic implementations tuned to hardware platforms. It is also a venue to discuss applications and performance results of FFT on current and emerging platforms such as many-core processors, GPUs, and distributed-memory systems.

Organizer: Daisuke Takahashi  
University of Tsukuba, Japan

Organizer: Franz Franchetti  
Carnegie Mellon University, USA

Organizer: Samar A. Aseeri  
King Abdullah University of Science & Technology (KAUST), Saudi Arabia

Organizer: Benson K. Muite  
University of Michigan, USA

#### 1:00-1:20 Implementation of Parallel FFTs on Cluster of Intel Xeon Phi Processors

Daisuke Takahashi, University of Tsukuba, Japan

#### 1:25-1:45 SPIRAL FFT

Franz Franchetti, Carnegie Mellon University, USA

#### 1:50-2:10 Pipelining Fast Fourier Transform on the OpenPOWER Cluster

Jun Doi, IBM Research - Tokyo, Japan

#### 2:15-2:35 Automatic FFT Kernel Generation for CUDA GPUs

Akira Nukada, Tokyo Institute of Technology, Japan

Wednesday, March 7

## MS3

### On Batched BLAS Standardization - Part I of II

1:00 PM-2:40 PM

Room:52-201

#### For Part 2 see MS14

It is becoming increasingly common in modern high-performance computing (HPC) to solve a large linear algebra problem by subdividing it into thousands of smaller problems which can be solved independently. To address the need for efficient libraries which allow users to perform batches of small BLAS operations in parallel, and to make efficient use of their hardware, many APIs have been investigated. While each of these new APIs has been designed to speed up specific applications on a target computing platform, the linear algebra community has recently begun standardization efforts to settle on a common batched BLAS interface. The main goal of this Minisymposium is to disseminate, discuss, and gathering agreed-upon ideas for the next official standard, while highlighting potential disagreements and points of contention to be resolved during the standardization process.

Organizer: Mawussi Zounon  
*The University of Manchester, UK*

Organizer: Azzam Haidar  
*University of Tennessee, Knoxville, USA*

Organizer: Siva Rajamanickam  
*Sandia National Laboratories, USA*

#### 1:00-1:20 Standardization and Optimization Opportunities for Batching Linear Algebra on Intel® Processors

*Sarah Knepper, Murat E. Guney, Kazushige Goto, Louise Huot, Arthur Mitrano, and Shane Story, Intel Corporation, USA; Mawussi Zounon, The University of Manchester, UK*

#### 1:25-1:45 NVIDIA's Batched BLAS, Challenges, and C++ Abstractions

*Cris Cecka, NVIDIA, USA*

#### 1:50-2:10 Vector-friendly Batched BLAS and LAPACK Kernels : Design and Applications

*Siva Rajamanickam, Kyungjoo Kim, Andrew Bradley, Mehmet Deveci, Simon D. Hammond, and Christian Trott, Sandia National Laboratories, USA*

#### 2:15-2:35 Batched BLAS in High-Order Finite Element Methods

*Johann Dahm, IBM Research, USA; Steve Rennich, NVIDIA, USA; David Medina, Rice University, USA; Veselin Dobrev and Tzanio Kolev, Lawrence Livermore National Laboratory, USA*

Wednesday, March 7

## MS4

### Solving Complex Partial Differential Equations on Massively Parallel Machines - Part I of II

1:00 PM-2:40 PM

Room:52-202

#### For Part 2 see MS15

Today's simulation models for technical and biological applications are mostly highly complex, multi-scale and multi-physics. They require very high resolution to achieve the desired accuracy. This calls for their efficient implementation on supercomputers. From the German priority program SPP 1648 -- Software for Exascale Computing this minisymposium presents three groups. ExaMAG targets the structure formation of the universe, ExaFSA fluid-structure-acoustic interactions, and ExaStencils pursues a domain-specific design approach for stencil codes, a fundamental building block of most solvers. We present methods and software advancing the state-of-the-art towards modular, portable, efficient and scalable simulation frameworks.

Organizer: Miriam Mehl  
*University of Stuttgart, Germany*

Organizer: Harald Koestler  
*University of Erlangen-Nuernberg, Germany*

Organizer: Christian F. Klingenberg  
*Wurzburg University, Germany*

#### 1:00-1:20 Large-Scale Applications for Real-World Problems Solving Partial Differential Equations on Heterogeneous GPU Machines

*Takayuki Aoki, Tokyo Institute of Technology, Japan*

#### 1:25-1:45 Coupling Parallel Simulation Codes: Problems and Solutions

*Benjamin Uekermann and Hans-Joachim Bungartz, Technical University of Munich, Germany; Miriam Mehl and Florian Lindner, University of Stuttgart, Germany; Benjamin R  th, Technical University of Munich, Germany; Amin Totounferoush, University of Stuttgart, Germany*

*continued in next column*

*continued on next page*



Wednesday, March 7

## MS4

### Solving Complex Partial Differential Equations on Massively Parallel Machines - Part I of II

1:00 PM-2:40 PM

continued

#### 1:50-2:10 The Universe in a Computer: How Numerical Methods are Essential

*Christian F. Klingenberg*, Wurzburg University, Germany

#### 2:15-2:35 Performance Analysis and Tuning of ExaFSA Codes

*Kazuhiko Komatsu*, Tohoku University, Japan

Wednesday, March 7

## MS5

### Approximate Computing Towards Exascale - Part I of II

1:00 PM-2:40 PM

Room:52-204

#### For Part 2 see MS16

Guaranteed numerical precision of each elementary step in a complex computation has been the mainstay of traditional computing systems for many years. This era is at its twilight: to overcome the “power wall” in Exascale systems, a shift from traditional computing paradigms is now mandatory. This minisymposium will investigate the theoretical and practical understanding of the energy efficiency boost obtainable when accuracy requirements on data being processed, stored and communicated can be lifted for intermediate calculations. The target applications range from Big Data Analytic and Deep Learning, up to classical scientific computing simulations in HPC environments.

Organizer: *Cristiano Malossi*  
*IBM Research-Zurich, Switzerland*

Organizer: *Costas Bekas*  
*IBM Research, USA*

Organizer: *Dimitrios S. Nikolopoulos*  
*Queen's University, Belfast, United Kingdom*

Organizer: *Enrique S. Quintana-Ortí*  
*Universidad Jaume I, Spain*

**1:00-1:20 The Transprecision Computing Paradigm and its Impact**  
*Cristiano Malossi, IBM Research-Zurich, Switzerland*

**1:25-1:45 Dynamic Accuracy Schemes for Deep Neural Networks**  
*Marc Casas, Barcelona Supercomputing Center, Spain*

#### 1:50-2:10 Exploiting Auxiliary Floating Point Types to Reduce Total Energy Consumption of Applications with Precision Constraints

*Giuseppe Tagliavini, Andrea Marongiu, Davide Rossi, Michele Lombardi, and Michela Milano*, University of Bologna, Italy

#### 2:15-2:35 To Reduce Numerical Precision to Achieve Higher Accuracy in Weather and Climate Modelling

*Peter D. Dueben*, European Weather Centre, United Kingdom

*continued in next column*

Wednesday, March 7

## MS6

### Extreme Scale Solvers for Coupled Systems - Part I of II

1:00 PM-2:40 PM

Room:52-301

#### For Part 2 see MS17

Exascale computers will exhibit billion-way parallelism. Computing on such extreme scale needs methods scaling perfectly with optimal complexity. This minisymposium combines talks on crucial aspects of extreme scale solving. The solver must be of optimal complexity, which is more and more severe with increasing problem size, and scale efficiently on extreme scales of parallelism. To that end, the minisymposium brings together talks on parallel adaptive multigrid methods in space and time, as well as optimization and uncertainty quantification techniques. Also reducing power consumption will be a topic of the minisymposium.

Organizer: Gabriel Wittum

King Abdullah University of Science & Technology (KAUST), Saudi Arabia

#### 1:00-1:20 Uncertainty Quantification Using Tensor Methods and Large Scale HPC

Lars Grasedyck, RWTH-Aachen, Germany

#### 1:25-1:45 Efficient Parallelization by Space-Time Discretization and Solution

Rolf Krause, Seif Ben Bader, and Marco Favino, Università della Svizzera italiana, Switzerland; Christian Hesch, University of Siegen, Germany; Alessio Quaglino, Università della Svizzera italiana, Switzerland

#### 1:50-2:10 Scalable Shape Optimization Methods for Structured Inverse Modeling Using Large Scale HPC

Volker H. Schulz, University of Trier, Germany

#### 2:15-2:35 Parallel Adaptive and Robust Multigrid

Gabriel Wittum, King Abdullah University of Science & Technology (KAUST), Saudi Arabia

Wednesday, March 7

## MS7

### Resilience for Extreme Scale Computing - Part I of IV

1:00 PM-2:40 PM

Room:52-302

#### For Part 2 see MS18

The reliability of large scale HPC systems has been a major concern due to the increasing complexity of hardware and system software in addition to the tight power budget for system operations. Despite the reliability improvement in the recent hardware, the users are still responsible for mitigating the failures in their applications. In fact, the US DOE Exascale Computing Project aims at achieving 1 week of meantime between “application” failures, assuming checkpoint and restart is integrated by default. In this minisymposium, we will discuss the recent progress in the resilience for high performance computing systems from a variety of perspectives, including hardware, systems, runtime, algorithm and applications. We will also discuss challenges and opportunities that facilitate resilience-centric HPC system co-design for reducing the operational budget of these systems while maintaining their capability.

Organizer: Keita Teranishi

Sandia National Laboratories, USA

Organizer: Emmanuel Agullo

Inria, France

Organizer: George Bosilca

University of Tennessee, Knoxville, USA

Organizer: Christian

Engelmann

Oak Ridge National Laboratory, USA

Organizer: Luc Giraud

Inria, France

#### 1:00-1:20 Resilience Design Patterns: A Structured Approach to Resilience at Extreme Scale

Saurabh Hukerikar and Christian Engelmann, Oak Ridge National Laboratory, USA

#### 1:25-1:45 Data Analytics and Machine Learning for Supercomputer Datacenter Design

Nathan A. DeBardeleben and Elisabeth Baseman, Los Alamos National Laboratory, USA

#### 1:50-2:10 Accurate and Portable Compiler-based Fault Injection in HPC Programs

Ignacio Laguna, Lawrence Livermore National Laboratory, USA; Georgios Georgakoudis and Dimitrios S. Nikolopoulos, Queen's University, Belfast, United Kingdom; Martin Schulz, Lawrence Livermore National Laboratory, USA

#### 2:15-2:35 Title Not Available At Time Of Publication

Frank Cappello, Argonne National Laboratory, USA

continued in next column

Wednesday, March 7

## MS8

### Parallel Multigrid Methods and Iterative Solvers - Part I of II

1:00 PM-2:40 PM

Room:52-303

#### For Part 2 see MS19

The solution of large systems of equations and eigenvalue problems is needed in many applications in computational science and engineering. Often these applications require the use of massively parallel computers, due to excessive need of memory or compute power. As a consequence scalable numerical methods are needed. Iterative methods are usually used to solve these problems, often multigrid methods are the best choice as solver or preconditioner. In order to exploit modern architectures and the vast amount of parallelism different techniques have to be employed, e.g., to reduce communication. These techniques are often similar in different types of solvers, e.g., blocking is used to increase spatial locality. Further, the use of multigrid is also possible for non-linear problems, again asking for similar parallelization techniques. In this minisymposium recent developments in highly scalable iterative methods are presented.

Organizer: Matthias Bolten

*University of Wuppertal, Germany*

Organizer: Thomas K. Huckle

*Technische Universität München, Germany*

#### 1:00-1:20 Structured Multigrid at Extreme Scales

*Luke Olson and Andrew Reisner, University of Illinois at Urbana-Champaign, USA; David Moulton, Los Alamos National Laboratory, USA*

#### 1:25-1:45 Improving Parallel Structured Multigrid Methods with Block Smoothers

*Matthias Bolten, University of Wuppertal, Germany*

#### 1:50-2:10 Multigrid Smoothers for the Stokes Problem

*Lisa Claus, Bergische Universität, Germany*

#### 2:15-2:35 The Analysis for Robustness of Sa-Amg Method by Extraction of Near-Kernel Vectors

*Naoya Nomura, University of Tokyo, Japan*

Wednesday, March 7

## MS9

### Tensor Decomposition for High Performance Data Analytics - Part I of III

1:00 PM-2:40 PM

Room:52-304

#### For Part 2 see MS20

With the explosion of Big Data, finding fast and scalable solutions for mining and analyzing large amounts of data is becoming increasingly important. In this regard, tensor decomposition has recently received much attention due to its ability to identify latent properties in high-order data. Unfortunately, the sparseness and the high-order nature of real-world data makes decomposing tensors slow and inefficient. This minisymposium will present research into new and improved methods for tensor decomposition for both shared-memory and distributed systems, as well as their application for analyzing real data.

Organizer: Jee Choi

*IBM T.J. Watson Research Center, USA*

Organizer: Keita Teranishi

*Sandia National Laboratories, USA*

Organizer: Richard Vuduc

*Georgia Institute of Technology, USA*

#### 1:00-1:20 High-performance Dense Tucker Decomposition on GPU Clusters

*Jee Choi, IBM T.J. Watson Research Center, USA; Xing Liu, IBM Research, USA*

#### 1:25-1:45 Computing Sparse Tensor Decompositions using Dimension Trees

*Oguz Kaya, Inria Bordeaux Sud-Ouest, France; Bora Ucar, LIP-ENS Lyon, France*

#### 1:50-2:10 A Sparse Tensor Algebra Compiler

*Fredrik Kjolstad, Massachusetts Institute of Technology, USA; Shoaib Kamil, Adobe Systems, USA; Stephen Chou, Massachusetts Institute of Technology, USA; David Lugato, CEA, France; Saman Amarasinghe, Massachusetts Institute of Technology, USA*

#### 2:15-2:35 HiCOO: Hierarchical Storage of Sparse Tensors

*Jiajia Li, Jimeng Sun, and Richard Vuduc, Georgia Institute of Technology, USA*

*continued in next column*

Wednesday, March 7

## MS10

### Resource Management, Scheduling, Workflows: Critical Middleware for HPC and Clouds- Part I of II

1:00 PM-2:40 PM

Room:57-201

#### For Part 2 see MS21

As applications and systems become more complex, the middleware needed to execute on such platforms needs to be able to keep up: systems need to be able to deal with a variety of resources and resources types, incl. power, network and I/O requirements; need to enable scheduling options to map applications to heterogeneous architectures, incl. complex network topologies; and need to provide capabilities to support emerging workflows, incl. V&V and UQ workloads. These needs require a new set of resource management, scheduling and workflow solutions, both for HPC and cloud environments. In this three part minisymposium we assemble a series of expert speakers covering these critical middleware components who will both report on the fundamental problems we face in this area as well as discuss possible solutions. We will cover how these three areas are fundamentally connected and need to be solved in an integrated software stack. We will also discuss how they interact with the rest of the system, both at the system and application level and what requirements they impose on the rest of the system. Overall, we aim at an interactive session that allows us to discuss these critical software components with the community and to foster new collaborations enabling a tighter integration not only within the software stack, but also with the rest of the system.

*continued in next column*

Organizer: Tapasya Patki  
*Lawrence Livermore National Laboratory, USA*

Organizer: Martin Schulz  
*Lawrence Livermore National Laboratory, USA*

Organizer: Masaaki Kondo  
*University of Tokyo, Japan*

#### 1:00-1:20 Enabling Hierarchical Scheduling in Next-Generation HPC Systems with Flux

*Tapasya Patki, Lawrence Livermore National Laboratory, USA*

#### 1:25-1:45 Rethinking Performance in HPC Systems: Monitoring, Analytics, and Resource Management

*Ayse Coskun, Boston University, USA*

#### 1:50-2:10 Energy and Time-to-Solution Optimization with GEOPM

*Jonathan Eastep, Intel Corporation, USA*

#### 2:15-2:35 The Interplay of Workflow Execution and Resource Provisioning

*Rafael Ferreira da Silva, University of Southern California, USA*

Wednesday, March 7

## MS11

### High Performance Computational Electromagnetics - Part I of II

1:00 PM-2:40 PM

Room:57-202

#### For Part 2 see MS22

Electromagnetic field analysis is one of key technologies for designing electric machines and devices. Therefore, there is a strong demand of acceleration of an electromagnetic field analysis for an efficient design process. This minisymposium (MS) focuses on various techniques for fast electromagnetic field analyses. One of main topics of the MS is introduction of high performance computing technologies including parallel processing into the analysis. Recent processors are equipped with many cores and they support vector instruction sets with a wide SIMD length. However, it is not easy to effectively use them in practical simulations. In the MS, we will discuss how to efficiently use these state-of-the-art processors in the computational electromagnetics. Namely, parallel processing of finite element analysis, improvement of linear iterative solvers, a use of parallel in time method and related techniques will be reported in the MS. Moreover, the MS also pays special attention to improvement of numerical methods for the computational electromagnetics. It involves development in mesh-less methods, improvement of subspace correction methods, and increased simulation accuracy.

Organizer: Takeshi Iwashita  
*Hokkaido University, Japan*

Organizer: Takeshi Mifune  
*Kyoto University, Japan*

Organizer: Yasuhito Takahashi  
*Doshisha University, Japan*

*continued on next page*

Wednesday, March 7

## MS11

### High Performance Computational Electromagnetics - Part I of II

1:00 PM-2:40 PM

continued

**1:00-1:20 Parallel-in-Time Magnetic Field Analysis of Electric Machines Based on Time-Periodic Finite-Element Method and Time-Periodic Explicit Error Correction Method**  
*Yasuhito Takahashi*, Doshisha University, Japan

**1:25-1:45 Analysis of Rotating Machines by Hierarchical Domain Decomposition Method**  
*Shin-ichiro Sugimoto*, Tokyo University of Science, Suwa, Japan

**1:50-2:10 Fast Finite Element Analyses for Electric Machines by Using Massively Parallel Processing**  
*Kazuki Semba*, Hirokatsu Katagiri, Tatsuya Asanuma, Masahiko Miwa, Hiroyuki Sano, and Takashi Yamada, JSOL Corporation, Japan

**2:15-2:35 Parallelization of Preconditioned Krylov Subspace Method Using Block-Multicolor Ordering in Linear Systems Derived from Finite Element Method**  
*Tomonori Tsuburaya*, Fukuoka University, Japan; Yoshifumi Okamoto, Hosei University, Japan

Wednesday, March 7

## CP1

### Scalable Linear Solvers - Part I of II

1:00 PM-2:40 PM

Room:52-101

Chair: *Hartwig Anzt*, University of Tennessee, USA

**1:00-1:20 Fast Scalable Solvers for Charge Models in Molecular Dynamics Applications**  
*H. Metin Aktulga*, Michigan State University, USA

**1:25-1:45 A Scalable Block-Iterative (Non)-Linear Solver for Adaptive 3D Organic Semiconductor Simulation**  
*Carlo De Falco*, Politecnico di Milano, Italy

**1:50-2:10 Leveraging Kokkos/Tpetra for Performance Portability in the Thyra Abstraction Layer**  
*Alexander Toth*, Roger Pawlowski, and Ross Bartlett, Sandia National Laboratories, USA

**2:15-2:35 BDDC and FETI-DP Methods in PETSC**  
*Stefano Zampini*, King Abdullah University of Science & Technology (KAUST), Saudi Arabia

Wednesday, March 7

## CP2

### HPC Applications in Physics

1:00 PM-2:40 PM

Room:52-102

Chair: *Wolfgang Bangert*, Colorado State University, USA

**1:00-1:20 Performance Assessment of Hybrid Parallelism for Large-Scale Reservoir Simulation on Petascale Architecture**  
*Amani Alonazi*, King Abdullah

University of Science & Technology (KAUST), Saudi Arabia; Marcin Rogowski and Ahmed Al-Zawawi, Saudi Aramco Oil Company, Saudi Arabia; David E. Keyes, King Abdul Aziz University, Saudi Arabia

**1:25-1:45 Radial-Basisfunction Interpolation for Black-Box Multi-Physics Simulation**  
*Florian Lindner* and Miriam Mehl, University of Stuttgart, Germany; Benjamin Uekermann, Technical University of Munich, Germany

**1:50-2:10 Persistent Homology of Geospatial Signals**  
*David Nicholaeff*, Descartes Labs, USA

**2:15-2:35 A Parallel Library of P-Adaptive High Order Finite Elements for Geophysical Applications**  
*Giovanni Tumolo*, International Centre for Theoretical Physics, Trieste, Italy; Luca Bonaventura, Politecnico di Milano, Italy

### Coffee Break

2:40 PM-3:10 PM



Room:63-1F



Wednesday, March 7

## MS12

### Advanced Approaches for Environmental Engineering Problems

3:10 PM-4:50 PM

Room:52-103

Latest political decisions concerning climate change not only underline the need for new solutions, they also illustrate the importance of environmental engineering for better understanding and prediction of ecological challenges. By nature, these problems tend to be inherent complex, thus entailing sophisticated approaches based on modern multi-physics, multi-scale, massive parallel simulations. Applications vary from harvesting regenerative sources such as geothermic energy, air pollution problems using multi-phase fluid flow, environmental noise reduction, seismological analyses, flooding and tsunami scenarios as well as land reclamation. Furthermore, highly efficient visualisation methods need to be developed for visualising huge datasets generated by such parallel computations.

Organizer: Ralf-Peter Mundani  
*Technische Universität München, Germany*

Organizer: Jérôme Frisch  
*RWTH Aachen University, Germany*

### 3:10-3:30 Large Eddy Simulation of Thermal Atmospheric Environment in Urban Boundary Layer

*Atsushi Inagaki*, Tokyo Institute of Technology, Japan; Naoyuki Onodera, Japan Atomic Energy Agency, Japan; Manabu Kanda and Takayuki Aoki, Tokyo Institute of Technology, Japan

### 3:35-3:55 Advanced Load-Balancing Techniques for Massive Parallel Flood Simulations

*Christoph M. Ertl*, Technische Universität München, Germany; Nevena Perovic, Technische Universität München, Germany; Bobby Minola Ginting, Technical University of Munich, Germany; Ralf-Peter Mundani and Ernst Rank, Technische Universität München, Germany

### 4:00-4:20 Petascale Earthquake Simulation Using High Order Methods

*Michael Bader*, Technische Universität München, Germany

### 4:25-4:45 Large Scale Numerical Simulations of Geothermal Potentials Using a Distributed Parallel Data Structure

*Jérôme Frisch*, RWTH Aachen University, Germany; Ralf-Peter Mundani, Technische Universität München, Germany

Wednesday, March 7

## MS13

### State-of-the-Art FFT --- Algorithms, Implementations, and Applications - Part II of II

3:10 PM-4:50 PM

Room:52-104

#### For Part 1 see MS2

The fast Fourier transform (FFT) is widely used in many areas of science and engineering. This minisymposium is an opportunity to discuss high performance and parallel implementations of FFT, mathematical encapsulations of FFT algorithms that are amenable to automatic implementations tuned to hardware platforms. It is also a venue to discuss applications and performance results of FFT on current and emerging platforms such as many-core processors, GPUs, and distributed-memory systems.

Organizer: Daisuke Takahashi  
*University of Tsukuba, Japan*

Organizer: Franz Franchetti  
*Carnegie Mellon University, USA*

Organizer: Samar A. Aseeri  
*King Abdullah University of Science & Technology (KAUST), Saudi Arabia*

Organizer: Benson K. Muite  
*University of Michigan, USA*

### 3:10-3:30 Fast Fourier Transforms (fft)

*Samar A. Aseeri*, King Abdullah University of Science & Technology (KAUST), Saudi Arabia

### 3:35-3:55 Use of the Fast Fourier Transform in Solving Partial Differential Equations

*Benson K. Muite*, University of Michigan, USA

### 4:00-4:20 Parallel Fast Gauss Transform

*Shravan Veerapaneni*, University of Michigan, USA

### 4:25-4:45 Implementation of OpenFFT and Its Application to Industrial Problems

*Truong Vinh Truong Duy*, Nissan ARC, Japan

Wednesday, March 7

## MS14

### On Batched BLAS Standardization - Part II of II

3:10 PM-4:50 PM

Room: 52-201

#### For Part 1 see MS3

It is becoming increasingly common in modern high-performance computing (HPC) to solve a large linear algebra problem by subdividing it into thousands of smaller problems which can be solved independently. To address the need for efficient libraries which allow users to perform batches of small BLAS operations in parallel, and to make efficient use of their hardware, many APIs have been investigated. While each of these new APIs has been designed to speed up specific applications on a target computing platform, the linear algebra community has recently begun standardization efforts to settle on a common batched BLAS interface. The main goal of this minisymposium is to disseminate, discuss, and gathering agreed-upon ideas for the next official standard, while highlighting potential disagreements and points of contention to be resolved during the standardization process.

Organizer: Mawussi Zounon

*The University of Manchester, UK*

Organizer: Negin Bagherpour

*University of Manchester, United Kingdom*

Organizer: Azzam Haidar

*University of Tennessee, Knoxville, USA*

Organizer: Siva Rajamanickam

*Sandia National Laboratories, USA*

#### 3:10-3:30 Magma Batched Computations: Current Development and Trend

Azzam Haidar, University of Tennessee, Knoxville, USA; Tomov Stan and Ahmad Ahmad, University of Tennessee, USA; Jack J. Dongarra, University of Tennessee, ORNL, and University of Manchester, USA

#### 3:35-3:55 CuHinesBatch, a Cuda Implementation to Solve Batches of Hines Systems. Human Brain Project

Pedro Valero Lara, Barcelona Supercomputing Center, Spain

#### 4:00-4:20 Batched Triangular Dense Linear Algebra for Very Small Matrices on GPUs

Ali M. Charara, David E. Keyes, and Hatem Ltaief, King Abdul Aziz University, Saudi Arabia

#### 4:25-4:45 Batched DGEMM Operations in Density Matrix Renormalization Group

Eduardo F. D'Azevedo, Oak Ridge National Laboratory, USA; Arghya Chatterjee, Georgia Institute of Technology, USA; Wael R. Elwasif, Oscar Hernandez, and Gonzalo Alvarez, Oak Ridge National Laboratory, USA

Wednesday, March 7

## MS15

### Solving Complex Partial Differential Equations on Massively Parallel Machines - Part II of II

3:10 PM-4:50 PM

Room: 52-202

#### For Part 1 see MS4

Today's simulation models for technical and biological applications are mostly highly complex, multi-scale and multi-physics. They require very high resolution to achieve the desired accuracy. This calls for their efficient implementation on supercomputers. From the German priority program SPP 1648 -- Software for Exascale Computing this minisymposium presents three groups. ExaMAG targets the structure formation of the universe, ExaFSA fluid-structure-acoustic interactions, and ExaStencils pursues a domain-specific design approach for stencil codes, a fundamental building block of most solvers. We present methods and software advancing the state-of-the-art towards modular, portable, efficient and scalable simulation frameworks.

Organizer: Miriam Mehl

*University of Stuttgart, Germany*

Organizer: Harald Koestler

*University of Erlangen-Nuernberg, Germany*

Organizer: Christian F.

Klingenberg

*Wuerzburg University, Germany*

#### 3:10-3:30 Simulating Cosmic Structure Formation

Naoki Yoshida, University of Tokyo, Japan

#### 3:35-3:55 Parallel Implicit DG Algorithms Without Linear System Inversion. Applications to Plasma Physics

Philippe Helluy, University of Strasbourg, France

#### 4:00-4:20 Generation of Highly Parallel Multigrid Solvers for CFD Applications

Harald Koestler and Sebastian Kuckuk, University of Erlangen-Nuernberg, Germany

#### 4:25-4:45 Implementation Techniques of Domain-Specific Languages for Parallel Solvers

Shigeru Chiba, University of Tokyo, Japan

*continued in next column*

Wednesday, March 7

## MS16

### Approximate Computing Towards Exascale - Part II of II

3:10 PM-4:50 PM

Room:52-204

#### For Part 1 see MS5

Guaranteed numerical precision of each elementary step in a complex computation has been the mainstay of traditional computing systems for many years. This era is at its twilight: to overcome the “power wall” in Exascale systems, a shift from traditional computing paradigms is now mandatory. This minisymposium will investigate the theoretical and practical understanding of the energy efficiency boost obtainable when accuracy requirements on data being processed, stored and communicated can be lifted for intermediate calculations. The target applications range from Big Data Analytic and Deep Learning, up to classical scientific computing simulations in HPC environments.

Organizer: Cristiano Malossi  
*IBM Research-Zurich, Switzerland*

Organizer: Costas Bekas  
*IBM Research, USA*

Organizer: Dimitrios S. Nikolopoulos  
*Queen's University, Belfast, United Kingdom*

Organizer: Enrique S. Quintana-Ortí  
*Universidad Jaume I, Spain*

#### 3:10-3:30 New Approaches to Memory Reliability Management for Big Data Workloads

*Dimitrios S. Nikolopoulos, Queen's University, Belfast, United Kingdom*

#### 3:35-3:55 Adjoint Methods for Stochastic Approximate Computing

*Uwe Naumann, Jens Deussen, and Jonathan Hüser, RWTH-Aachen, Germany*

#### 4:00-4:20 Adaptive Precision Preconditioning for Sparse Linear Systems

*Enrique S. Quintana-Ortí, Universidad Jaume I, Spain; Hartwig Anzt, University of Tennessee, USA; Jack J. Dongarra, University of Tennessee, ORNL, and University of Manchester, USA; Goran Flegar, Universitat Jaume I, Spain*

#### 4:25-4:45 Exploiting Extended Hardware Margins to Improve Energy Efficiency

*Christos D. Antonopoulos, University of Thessaly, Greece*

Wednesday, March 7

## MS17

### Extreme Scale Solvers for Coupled Systems - Part II of II

3:10 PM-4:50 PM

Room:52-301

#### For Part 1 see MS6

Exascale computers will exhibit billion-way parallelism. Computing on such extreme scale needs methods scaling perfectly with optimal complexity. This minisymposium combines talks on crucial aspects of extreme scale solving. The solver must be of optimal complexity, which is more and more severe with increasing problem size, and scale efficiently on extreme scales of parallelism. To that end, the minisymposium brings together talks on parallel adaptive multigrid methods in space and time, as well as optimization and uncertainty quantification techniques. Also reducing power consumption will be a topic of the minisymposium.

Organizer: Gabriel Wittum  
*King Abdullah University of Science & Technology (KAUST), Saudi Arabia*

#### 3:10-3:30 Scaling Studies for Large Scale Multigrid Computations

*Andreas Vogel, Ruhr-University Bochum, Germany; Sebastian Reiter, Goethe University Frankfurt, Germany; Arne Nägel, Goethe University, Germany; Gabriel Wittum, King Abdullah University of Science & Technology (KAUST), Saudi Arabia*

#### 3:35-3:55 Meshing and Adaption of Unstructured Grid Hierarchies for Large Scale Parallel Multigrid Computations

*Sebastian Reiter, Goethe University Frankfurt, Germany*

#### 4:00-4:20 Robust and Adaptive Non-Linear Solvers

*Arne Nägel, Goethe University, Germany*

#### 4:25-4:45 Modeling Programming Languages and Potential Applications for HPC

*Michael Hoffer, University of Frankfurt, Germany*

Wednesday, March 7

## MS18

### Resilience for Extreme Scale Computing - Part II of IV

3:10 PM-4:50 PM

Room:52-302

For Part 1 see MS7

For Part 3 see MS30

The reliability of large scale HPC systems has been a major concern due to the increasing complexity of hardware and system software in addition to the tight power budget for system operations. Despite the reliability improvement in the recent hardware, the users are still responsible for mitigating the failures in their applications. In fact, the US DOE Exascale Computing Project aims at achieving 1 week of meantime between “application” failures, assuming checkpoint and restart is integrated by default. In this minisymposium, we will discuss the recent progress in the resilience for high performance computing systems from a variety of perspectives, including hardware, systems, runtime, algorithm and applications. We will also discuss challenges and opportunities that facilitate resilience-centric HPC system co-design for reducing the operational budget of these systems while maintaining their capability.

Organizer: Keita Teranishi

Sandia National Laboratories, USA

Organizer: Emmanuel Agullo

Inria, France

Organizer: George Bosilca

University of Tennessee, Knoxville, USA

Organizer: Christian

Engelmann

Oak Ridge National Laboratory, USA

Organizer: Luc Giraud

Inria, France

### 3:10-3:30 Algorithm Based Fault Recovery of Adaptively Refined Parallel Multilevel Grids

Linda Stals, The Australian National University, Australia

### 3:35-3:55 Recent Advances on the Algorithm-based Fault Tolerance of the Sakurai-Sugiura Eigensolver

Yasunori Futamura, Akira Imakura, and Tetsuya Sakurai, University of Tsukuba, Japan

### 4:00-4:20 Pattern-based Modeling of Fail-stop and Soft-error Resilience for Iterative Linear Solvers

Rizwan Ashraf, Christian Engelmann, and Saurabh Hukerikar, Oak Ridge National Laboratory, USA

### 4:25-4:45 An Algorithm-Based Fault-Tolerant Framework for High-dimensional PDEs

Michael Obersteiner and Alfredo Parra Hinojosa, Technical University of Munich, Germany; Mario Henne, University of Stuttgart, Germany; Hans-Joachim Bungartz, Technical University of Munich, Germany; Dirk Pfluger, University of Stuttgart, Germany

Wednesday, March 7

## MS19

### Parallel Multigrid Methods and Iterative Solvers - Part II of II

3:10 PM-4:50 PM

Room:52-303

For Part 1 see MS8

The solution of large systems of equations and eigenvalue problems is needed in many applications in computational science and engineering. Often these applications require the use of massively parallel computers, due to excessive need of memory or compute power. As a consequence scalable numerical methods are needed. Iterative methods are usually used to solve these problems, often multigrid methods are the best choice as solver or preconditioner. In order to exploit modern architectures and the vast amount of parallelism different techniques have to be employed, e.g., to reduce communication. These techniques are often similar in different types of solvers, e.g., blocking is used to increase spatial locality. Further, the use of multigrid is also possible for non-linear problems, again asking for similar parallelization techniques. In this minisymposium recent developments in highly scalable iterative methods are presented.

Organizer: Matthias Bolten

University of Wuppertal, Germany

Organizer: Thomas K. Huckle

Technische Universität München, Germany

### 3:10-3:30 A Recursive Multilevel Trust Region Method for Phase-field Models of Fracture

Alena Kopanicakova, Università della Svizzera italiana, Switzerland; Carola Bilgen and Kerstin Weinberg, Universität Siegen, Germany; Rolf Krause, Università della Svizzera italiana, Switzerland

continued in next column

continued on next page



**3:35-3:55 Horizontal Vectorization of Block Krylov Solvers**

*Christian Engwer*, Fahlke Jorrit, and  
Andreas Nüßing, University of  
Münster, Germany

**4:00-4:20 Iterative Parallel Methods for Deriving ILU-Type Preconditioners**

*Thomas K. Huckle*, Technische  
Universität München, Germany;  
Hartwig Anzt, University of  
Tennessee, USA; Matthias Bolten,  
University of Wuppertal, Germany

**4:25-4:45 Parallel Iterative Methods in the ELPA Eigensolver**

*Michael Rippl*, Technische Universität  
München, Germany

Wednesday, March 7

**MS20****Tensor Decomposition for High Performance Data Analytics - Part II of III**

3:10 PM-4:50 PM

*Room: 52-304*

**For Part 1 see MS9**

**For Part 3 see MS56**

With the explosion of Big Data, finding fast and scalable solutions for mining and analyzing large amounts of data is becoming increasingly important. In this regard, tensor decomposition has recently received much attention due to its ability to identify latent properties in high-order data. Unfortunately, the sparseness and the high-order nature of real-world data makes decomposing tensors slow and inefficient. This minisymposium will present research into new and improved methods for tensor decomposition for both shared-memory and distributed systems, as well as their application for analyzing real data.

Organizer: Jee Choi

*IBM T.J. Watson Research Center, USA*

Organizer: Keita Teranishi

*Sandia National Laboratories, USA*

Organizer: Richard Vuduc

*Georgia Institute of Technology, USA*

**3:10-3:30 Role of Tensors in Machine Learning**

*Anima Anandkumar*, Amazon.com, USA

**3:35-3:55 Accelerating the Tucker Decomposition with Compressed Sparse Tensors**

*Shaden Smith* and George Karypis,  
University of Minnesota, USA

**4:00-4:20 The SPLATT Toolkit on the K Computer**

*Yiyu Tan*, RIKEN Advanced Institute  
for Computational Science, Japan;  
Toshiyuki Imamura, RIKEN, Japan

**4:25-4:45 Shared-Memory Parallelization of MITKRP for Dense Tensors**

*Koby B. Hayashi*, Grey Ballard, Tobia  
Michael, and Yuijie Jiang, Wake  
Forest University, USA

Wednesday, March 7

**MS21****Resource Management, Scheduling, Workflows: Critical Middleware for HPC and Clouds- Part II of II**

3:10 PM-4:50 PM

*Room: 57-201*

**For Part 1 see MS10**

As applications and systems become more complex, the middleware needed to execute on such platforms needs to be able to keep up: systems need to be able to deal with a variety of resources and resources types, incl. power, network and I/O requirements; need to enable scheduling options to map applications to heterogeneous architectures, incl. complex network topologies; and need to provide capabilities to support emerging workflows, incl. V&V and UQ workloads. These needs require a new set of resource management, scheduling and workflow solutions, both for HPC and cloud environments. In this three part minisymposium we assemble a series of expert speakers covering these critical middleware components who will both report on the fundamental problems we face in this area as well as discuss possible solutions. We will cover how these three areas are fundamentally connected and need to be solved in an integrated software stack. We will also discuss how they interact with the rest of the system, both at the system and application level and what requirements they impose on the rest of the system. Overall, we aim at an interactive session that allows us to discuss these critical software components with the community and to foster new collaborations enabling a tighter integration not only within the software stack, but also with the rest of the system.

*continued on next page*



Wednesday, March 7

## MS21

**Resource Management, Scheduling, Workflows: Critical Middleware for HPC and Clouds- Part II of II**

3:10 PM-4:50 PM

continued

Organizer: Tapasya Patki

*Lawrence Livermore National Laboratory, USA*

Organizer: Martin Schulz

*Lawrence Livermore National Laboratory, USA*

Organizer: Masaaki Kondo

*University of Tokyo, Japan*

**3:10-3:30 Approaches to the Power Consumption Problem on the K Computer**

*Atsuya Uno, RIKEN Advanced Institute for Computational Science, Japan*

**3:35-3:55 Cache Partitioning as an Extension to Co-Scheduling: Can Applications Benefit from It?**

*Carsten Trinitis, TU Munich, Germany*

**4:00-4:20 COMPSs-Mobile: Parallel Programming for Mobile Cloud Computing**

*Francisco Javier Conejero, Barcelona Supercomputing Center, Spain*

**4:25-4:45 Robust Middleware Services for In-Situ Workflows**

*Pradeep Subedi, Rutgers University, USA*

Wednesday, March 7

## MS22

**High Performance Computational Electromagnetics - Part II of II**

3:10 PM-4:50 PM

*Room:57-202*

**For Part 1 see MS11**

Electromagnetic field analysis is one of key technologies for designing electric machines and devices. Therefore, there is a strong demand of acceleration of an electromagnetic field analysis for an efficient design process. This minisymposium (MS) focuses on various techniques for fast electromagnetic field analyses. One of main topics of the MS is introduction of high performance computing technologies including parallel processing into the analysis. Recent processors are equipped with many cores and they support vector instruction sets with a wide SIMD length. However, it is not easy to effectively use them in practical simulations. In the MS, we will discuss how to efficiently use these state-of-the-art processors in the computational electromagnetics. Namely, parallel processing of finite element analysis, improvement of linear iterative solvers, a use of parallel in time method and related techniques will be reported in the MS. Moreover, the MS also pays special attention to improvement of numerical methods for the computational electromagnetics. It involves development in mesh-less methods, improvement of subspace correction methods, and increased simulation accuracy.

Organizer: Takeshi Iwashita

*Hokkaido University, Japan*

Organizer: Takeshi Mifune

*Kyoto University, Japan*

Organizer: Yasuhito Takahashi

*Doshisha University, Japan*

**3:10-3:30 Parallelization of Variable Preconditioned Krylov Subspace Method for Linear System Obtained from Meshless Approaches**

*Soichiro Ikuno, Tokyo University of Technology, Japan; Taku Itoh, Nihon University, Japan*

**3:35-3:55 An Ultrafast Multi-GPU Eigensolver for Band Structures of Three-Dimensional Photonic Crystals**

*Weichung Wang, National Taiwan University, Taiwan*

**4:00-4:20 Performance Evaluation of Tiled 3D FDTD Solver on Recent Multicore Processors**

*Takeshi Iwashita and Takeshi Fukaya, Hokkaido University, Japan*

**4:25-4:45 Geometric Block Diagonal Preconditioning Technique for Electromagnetic Finite Element Analysis Using Voxel Mesh**

*Takeshi Mifune, Kyoto University, Japan*

*continued in next column*

Wednesday, March 7

## CP3

### Performance Tuning & Tools

3:10 PM-4:50 PM

Room:52-101

Chair: Georg Hager, Erlangen Regional Computing Center, Germany

#### 3:10-3:30 A Programming Framework for Performance Tuning in Julia

Ayumu Gomi and Daisuke Takahashi,  
University of Tsukuba, Japan

#### 3:35-3:55 PerfMemPlus: A Memory Performance Analysis Tool Based on Linux Perf

Christian R. Helm and Kenjiro Taura,  
University of Tokyo, Japan

#### 4:00-4:20 Acceleration of MuPAT on MATLAB/scilab Using Parallel Processing

Hotaka Yagi and Emiko Ishiwata, Tokyo University of Science, Japan; Hidehiko Hasegawa, University of Tsukuba, Japan

#### 4:25-4:45 Performance Optimization of the SCB Kernel on the X86 Multi-Core Platform

Peng Zhang, Institute of Applied Physics and Computational Mathematics, Beijing, P.R. China; Zhang Yang, Institute of Applied Physics and Computational Mathematics, China

Wednesday, March 7

## CP4

### Matrix Factorization

3:10 PM-4:50 PM

Room:52-102

Chair: Sherry Li, Lawrence Livermore National Laboratory, USA

#### 3:10-3:30 Recent Development in Distributed-Memory Hierarchical Interpolative Factorization

Yingzhou Li, Duke University, USA;  
Lexing Ying, University of Texas, USA

#### 3:35-3:55 A Method to Calculate SVD of a Sparse T-S Matrix with Iterative Correction of Orthogonality

Hiroshi Murakami, Tokyo Metropolitan University, Japan

#### 4:00-4:20 OpenMP Tasks for QR Factorization in PLASMA

Jakub Sístek, University of Manchester, United Kingdom;  
Jakub Kurzak, University of Tennessee, Knoxville, USA; Julien Langou, University of Colorado, Denver, USA

#### 4:25-4:45 A Hybrid Parallel Block Low-Rank Solver with Flexible Factorization Algorithms and Communication Schemes

Clement Weisbecker, Cleve Ashcraft, and Francois-Henry Rouet, Livermore Software Technology Corporation, USA

### Welcome Remarks

5:00 PM-5:15 PM

Room:63-201/202

Wednesday, March 7

## IP1

### Verified Numerical Computation and Large Scale Computing

5:15 PM-6:00 PM

Room:63-201/202

Chair: Tetsuya Sakurai, University of Tsukuba, Japan

In this talk, we overview the states of art of the verified numerical computation. One of the stresses is on large scale computation. We first treat computer assisted rigorous inclusion of eigenvalues of the self-adjoint operator on bounded domain. Especially, we report our recent work on how to calculate rigorous lower bounds of eigenvalues. Especially, we summarize how to get a rigorous lower bound of minimum eigenvalue. Then, we briefly review its application to computer assisted proof of nonlinear elliptic boundary value problems through the Newton-Kantorovich type arguments. Namely, inclusion of minimum eigenvalue is useful for bounding the operator norm of inverses of linear operators. In the process of inclusion of eigenvalues, usually large-scale computation is needed. We will show an example in which, using the Kei-computer, several eigenvalues are rigorously included for a square matrix with more than 1 million-dimension.

Shin'ichi Oishi

Waseda University, Japan

### Welcome Reception

6:15 PM-8:00 PM

Room:63-1F



## Thursday, March 8

### Registration

8:00 AM-5:00 PM

Room:63-1F

## IP2

### Computing Beyond Moore's Law

8:30 AM-9:15 AM

Room:63-201/202

Chair: Michael Heroux, Sandia National Laboratories, USA

Moore's Law is a techno-economic model that has enabled the Information Technology (IT) industry to nearly double the performance and functionality of digital electronics roughly every two years within a fixed cost, power and area. Within a decade, the technological underpinnings for the process Gordon Moore described will come to an end as lithography gets down to atomic scale. At that point, it will be feasible to create lithographically produced devices with characteristic dimensions in the 3nm–5nm range. This range corresponds to a dozen or fewer Si atoms across critical device features and will therefore be a practical limit for controlling charge in a classical sense. The classical technological driver that has underpinned Moore's law for the past 50 years is already failing and is anticipated to flatten by 2025. This talk provides an updated view of what a 2021-2023 system might look like and the challenges ahead, based on our most recent understanding of technology roadmaps. It also will discuss the tapering of historical improvements in lithography, and how it affects options available to continue scaling of successors to the first exascale machine.

John Shalf

Lawrence Berkeley National Laboratory, USA

Thursday, March 8

### Intermission

9:15 AM-9:25 AM

### Panel

9:25 AM-10:25 AM

Room:63-201/202

### Coffee Break

10:25 AM-10:50 AM

Room:63-1F



Thursday, March 8

## MS23

### Language-Level Abstractions for Asynchronous Runtimes - Part I of II

10:50 AM-12:30 PM

Room:52-102

#### For Part 2 see MS35

Next-generation architectures are motivating more asynchronous, multi-threaded applications. Programming models and languages are now incorporating asynchronous tasking and performance portability as first-class design considerations. Expressive programming models should enable runtime flexibility and therefore high performance across different application and platforms. With expressive models, applications can focus on algorithms with runtimes developed by system experts managing execution resources. These sessions explore progress in developing intuitive programming models and abstractions that map to underlying asynchronous, distributed, or multi-threaded runtimes.

Organizer: David S. Hollman  
Sandia National Laboratories, USA

Organizer: Jonathan J. Lifflander  
Sandia National Laboratories, USA

#### 10:50-11:10 Way of the DARMA: Sequential Semantics for Deterministic-by-default Task Parallelism

David S. Hollman, Sandia National Laboratories, USA

#### 11:15-11:35 The Flexible Computational Science Infrastructure (FLESCSI) Programming System: Overview & Productivity

Ben Bergen, Los Alamos National Laboratory, USA

#### 11:40-12:00 Charm++: Interoperation and Support for Higher Level Abstractions

Phil Miller, University of Illinois at Urbana-Champaign, USA

#### 12:05-12:25 Scaling Implicitly Parallel Programs to a Thousand Nodes with Regent

Elliott Slaughter, SLAC National Accelerator Laboratory, USA

Thursday, March 8

**MS24****Structure Preserving Numerical Methods in Large Scale Scientific Computations**

10:50 AM-12:30 PM

Room:52-103

In modern scientific computations, the use of high performance computing (HPC) techniques is indispensable. The size of the target problems, however, has still kept growing more and more huge, and the development of scientific computing relying only on HPC techniques might not be effective as before. On the other hand, in the area of mathematical numerical analysis, structure-preserving methods for differential equations have drawn much attention and proved its strength over recent decades. For example, symplectic methods for Hamiltonian problems, operator splitting methods, and variational integrators. They cleverly utilize some mathematical (often physical) structure of the target problems, and realize reliable computations compared to generic methods. A striking drawback of such methods is, however, that the numerical schemes tend to be implicit, and sometimes they are not as practical as expected. In this minisymposium, we spot this issue, and seek the potential possibility of the combination of structure-preserving methods and HPC techniques, which seems a promising new research direction toward the next generation of large scale scientific computations. The topics will include parallelization in structure-preserving methods, structure-preserving model reductions, structure-preserving numerical linear algebra, and also various implementation issues arising in those topics.

Organizer: Takayasu Matsuo  
University of Tokyo, Japan

Organizer: Yuto Miyatake  
Nagoya University, Japan

**10:50-11:10 Toward Modern Scientific Computations Based on Structure-Preserving Methods**

Takayasu Matsuo, University of Tokyo, Japan; Yuto Miyatake, Nagoya University, Japan

**11:15-11:35 Structure-Preserving Model Reduction of Forced Hamiltonian Systems**

Liqian Peng, Sandia National Laboratories, USA

**11:40-12:00 A Mixed Precision Semi-Lagrangian Method for Modern Hardware Architectures**

Lukas Einkemmer, University of Innsbruck, Austria

**12:05-12:25 Energy-Preserving Parareal Algorithm for the Hamilton Equation**

Ai Ishikawa, Takaharu Yaguchi, and Mitsuo Yokokawa, Kobe University, Japan

Thursday, March 8

**MS25****Large-Scale Electronic Structure Calculations**

10:50 AM-12:30 PM

Room:52-104

**For Part 2 see MS37**

Electronic structure calculations and their applications are among the most challenging and computationally demanding science and engineering problems. This minisymposium aims at presenting and discussing new electronic structure calculation algorithms and parallel implementations that are suitable for modern computing architectures. It also highlights recent software development efforts that enable new algorithms to be easily implemented on massively parallel many-core systems and GPUs. The presentations included in this minisymposium cover a variety of topics ranging from linear scaling density functional theory based approaches to fast Hartree-Fock and other accurate wavefunction methods to fast matrix computations required in electronic structure calculations and software infrastructures required to support a user/developer-friendly interface between application software and numerical libraries. They will highlight both the latest success stories and the remaining challenges.

Organizer: Chao Yang

Lawrence Berkeley National Laboratory, USA

**10:50-11:10 The ELSI Infrastructure for Large-Scale, Massively Parallel Electronic Structure Theory**

Volker Blum, Duke University, USA

continued in next column

continued on next page

Thursday, March 8

## MS25

### Large-Scale Electronic Structure Calculations

10:50 AM-12:30 PM

continued

#### 11:15-11:35 A Conjugate Gradient Method for Electronic Structure Calculations

Xiaoying Dai, Chinese Academy of Sciences, China

#### 11:40-12:00 Parallel Computation of the Single Particle Density Matrix for Electronics Structure Calculations

Daniel Osei-Kuffuor, Lawrence Livermore National Laboratory, USA

#### 12:05-12:25 Eigenstate Calculations using Sakurai-Sugiura Method with the Large-scale Electronic Structure Calculation Code CONQUEST

Ayako Nakata and Tsuyashi Miyazaki, National Institute for Materials Science, Japan; Yasunori Futamura and Tetsuya Sakurai, University of Tsukuba, Japan

Thursday, March 8

## MS26

### Deep Learning from HPC Perspective: Opportunities and Challenges - Part I of II

10:50 AM-12:30 PM

Room:52-201

#### For Part 2 see MS38

Deep learning is becoming an indispensable tool in a wide variety of scientific and engineering domains thanks to its groundbreaking advancement in data classification, recognition, and prediction. For instance, machine learning models such as convolutional networks and recurrent neural networks have been reported to demonstrate competitive or even outperforming accuracy as traditional numerical simulations in various scientific domains. However, training such models is an extremely compute intensive task, taking hours and even days with latest generations of processors. Consequently, parallel and distributed model training has been one of the hot topics in deep learning. In this minisymposium, we will discuss opportunities and challenges in high-performance deep learning from applications and systems perspectives. We will motivate the importance of high performance deep learning by presenting use cases of deep learning in science and engineering problems, and then present recent efforts in scaling deep learning in large-scale HPC systems, including both systems level as well as algorithmic innovations. Our minisymposium will allow attendees to learn about the state of the art and future directions in large-scale deep learning.

Organizer: Naoya Maruyama  
Lawrence Livermore National Laboratory, USA

Organizer: Brian Van Essen  
Lawrence Livermore National Laboratory, USA

Organizer: Aleksandr Drozd  
Tokyo Institute of Technology, Japan

Organizer: Mohamed Wahib  
RIKEN, Japan

#### 10:50-11:10 Scalable Deep Learning for Scientific Data Sets

Brian Van Essen, Lawrence Livermore National Laboratory, USA

#### 11:15-11:35 Accelerated Deep Learning Advances in Fusion Energy Science Predictions

William Tang, Princeton University, USA; Julian Kates-Harbeck, Harvard University, USA; Alexey Syatkovskiy, Princeton University, USA

#### 11:40-12:00 Exascale Deep Learning for Text Comprehension

Arvind Ramanathan, Shang Gao, Michael Young, John Qiu, James Blair Christian, Hong-Jun Yoon, and Georgia Tourassi, Oak Ridge National Laboratory, USA; Paul Fearn, National Cancer Institute, USA

#### 12:05-12:25 Challenges in Training Large Neural Networks for Natural Language Processing

Raphael Shu and Hideki Nakayama, University of Tokyo, Japan

continued in next column



Thursday, March 8

## MS27

### Parallel Lattice Boltzmann Methods - Part I of II

10:50 AM-12:30 PM

Room:52-202

#### For Part 2 see MS39

In the field of computational fluid dynamics lattice Boltzmann methods (LBM) pose an alternative to classical Navier-Stokes solvers for the simulation of time-dependent, incompressible flows. The locality of the LBM makes it an ideal method for distributed memory system and allows for implementations with excellent scalability on GPUs and CPUs. LBM have been applied to a broad range of problems, from complex geometries like porous media to studies of turbulent flows. This Minisymposium offers an overview of the recent advancements in leading LBM implementations. We cover high performance implementations of advanced topics like refinement or particle-coupling methods on CPUs and GPUs.

Organizer: Christian  
Godenschwager

*University of Erlangen-Nuernberg,  
Germany*

Organizer: Takayuki Aoki  
*Tokyo Institute of Technology, Japan*

Organizer: Ulrich Rüde  
*University of Erlangen-Nuernberg,  
Germany*

#### 10:50-11:10 A Framework for Large-Scale, High-Performance Lattice Boltzmann Simulations in Complex Geometries

*Christian Godenschwager, Martin Bauer,  
Florian Schornbaum, and Ulrich  
Rüde, University of Erlangen-  
Nuernberg, Germany*

#### 11:15-11:35 An Octree-Based AMR Lattice Boltzmann Method for Multi-GPU Aerodynamics Simulations

*Yuta Hasegawa and Takayuki Aoki,  
Tokyo Institute of Technology, Japan*

*continued in next column*

#### 11:40-12:00 Acceleration of Turbulent Wind Simulation Using Locally Mesh-Refined Lattice Boltzmann Method

*Naoyuki Onodera and Yasuhiro Idomura,  
Japan Atomic Energy Agency, Japan*

#### 12:05-12:25 Extending Existing Simulation Software Packages with Spatial Adaptivity in a Minimal-Invasive Way

*Michael Lahnert, University of Stuttgart,  
Germany; Takayuki Aoki, Tokyo  
Institute of Technology, Japan;  
Malte Brunn, University of Stuttgart,  
Germany; Carsten Burstedde,  
Universität Bonn, Germany; Steffen  
Hirschmann, Dirk Pflüger, and  
Miriam Mehl, University of Stuttgart,  
Germany*

Thursday, March 8

## MS28

### Scalable Communication-Avoiding and -Hiding Krylov Subspace Methods - Part I of II

10:50 AM-12:30 PM

Room:52-204

#### For Part 2 see MS40

With HPC hardware advancing towards the exascale regime, the gap between computation (i.e. flops) and communication (i.e. internode communication, as well as data movement within the memory hierarchy) keeps steadily increasing. To address the growing time and energy costs of communication, classic numerical algorithms - which were primarily optimized for flop performance - have to be thoroughly revised or even reformulated entirely to be able to benefit from the increasing levels of parallelism. Parallel scalability can be improved by either reducing the number of global synchronization bottlenecks (communication avoiding), and/or overlapping communication latency with computations where possible (communication hiding). Although many of these re-engineered algorithms are equivalent to their classic counterparts in exact arithmetic, their numerical stability in a practical finite precision framework is often reduced. This minisymposium aims to give an overview of recent developments in the design, implementation and numerical analysis of sparse linear algebra solvers, with a strong emphasis on communication-avoiding and -hiding Krylov subspace methods for solving linear systems.

Organizer: Siegfried Cools  
*University of Antwerp, Belgium*

Organizer: Erin C. Carson  
*New York University, USA*

*continued on next page*

Thursday, March 8

## MS28

### Scalable Communication-Avoiding and -Hiding Krylov Subspace Methods - Part I of II

10:50 AM-12:30 PM

continued

#### 10:50-11:10 High Performance Variants of Krylov Subspace Methods

Emmanuel Agullo, Inria, France; *Erin C. Carson*, New York University, USA; Siegfried Cools, University of Antwerp, Belgium; James Demmel, University of California, Berkeley, USA; Pieter Ghysels, Lawrence Berkeley National Laboratory, USA; Luc Giraud, Inria, France; Miro Rozloznik, Czech Academy of Sciences, Czech Republic; Zdenek Strakos, Charles University, Czech Republic; Petr Tichy and Miroslav Tuma, Czech Academy of Sciences, Czech Republic; Wim Vanroose, Antwerp University, Belgium; Emrullah Fatih Yetkin, Inria, France

#### 11:15-11:35 About Parallel Variants of GMRES Algorithm

Jocelyne Erhel, Inria-Rennes, France

#### 11:40-12:00 Enlarged GMRES for Reducing Communication

Hussam Al Daas, UPMC-Inria-TOTAL, France; Laura Grigori, Inria, France; Pascal Henon, Total E&P, USA; Philippe Ricoux, TOTAL SA, France

#### 12:05-12:25 Iteration-Fusing Conjugate Gradient

Sicong Zhuang and Marc Casas, Barcelona Supercomputing Center, Spain

Thursday, March 8

## MS29

### Performance Engineering from the Node Level to the Extreme Scale- Part I of II

10:50 AM-12:30 PM

Room:52-301

#### For Part 2 see MS41

Achieving hardware and energy efficiency is important for current large-scale numerical simulations and will be a key component in the exascale era. In a world of heterogeneous, highly parallel computer architectures with deep memory hierarchies, complex application scenarios, and a broad spectrum of algorithms, software developers can no longer rely on mystic “black-box” performance engineering computer programs. Instead, a thorough analysis and understanding of the complex interaction of software, data structures, algorithms, and hardware features, a.k.a. performance engineering, is required for implementing codes that allow for portable performance on the computer generations to come. The minisymposium addresses a broad range of topics in performance engineering for modern HPC architectures, ranging from recent advances in performance models and tools supporting a “white-box” performance engineering approach to application performance tuning cases studies. The presentations should point out the potentials and limitations of “white-box” performance engineering activities and demonstrate the wide spectrum of performance models used in the PE activities including simple performance expectations, automatic model parameter selections, and analytic models.

Organizer: Gerhard Wellein  
*Erlangen Regional Computing Center, Germany*

Organizer: Georg Hager  
*Erlangen Regional Computing Center, Germany*

#### 10:50-11:10 If It Doesn't Work, We Learn Something: Instructive Case Studies in Performance Engineering

*Georg Hager*, Erlangen Regional Computing Center, Germany

#### 11:15-11:35 Fast Matrix-Free High-Order Discontinuous Galerkin Kernels: Performance Optimization and Modeling

*Martin Kronbichler*, Technische Universität München, Germany; Niklas Fehn, Technical University of Munich, Germany; Katharina Kormann, Max Planck Institute for Plasma Physics, Germany; Wolfgang A Wall, Technical University of Munich, Germany

#### 11:40-12:00 Efficient and Robust Parallel ILU Preconditioners for Quantum Eigenvalue Problems in ppOpen-HPC/ESSEX-II

*Masatoshi Kawai*, Akihiro Ida, and Kengo Nakajima, University of Tokyo, Japan

#### 12:05-12:25 Roofline: a Throughput Oriented Performance Model

*Leonid Oliker* and Tuomas Koskela, Lawrence Berkeley National Laboratory, USA; Jack Deslippe, National Energy Research Scientific Computing Center, USA; Zakhar Matveev, Intel Corporation, USA; Samuel Williams, Lawrence Berkeley National Laboratory, USA

*continued in next column*

Thursday, March 8

**MS30****Resilience for Extreme Scale Computing - Part III of IV**

10:50 AM-12:30 PM

Room:52-302

**For Part 2 see MS18****For Part 4 see MS42**

The reliability of large scale HPC systems has been a major concern due to the increasing complexity of hardware and system software in addition to the tight power budget for system operations. Despite the reliability improvement in the recent hardware, the users are still responsible for mitigating the failures in their applications. In fact, the US DOE Exascale Computing Project aims at achieving 1 week of meantime between “application” failures, assuming checkpoint and restart is integrated by default. In this minisymposium, we will discuss the recent progress in the resilience for high performance computing systems from a variety of perspectives, including hardware, systems, runtime, algorithm and applications. We will also discuss challenges and opportunities that facilitate resilience-centric HPC system co-design for reducing the operational budget of these systems while maintaining their capability.

Organizer: Keita Teranishi

*Sandia National Laboratories, USA*

Organizer: Emmanuel Agullo

*Inria, France*

Organizer: George Bosilca

*University of Tennessee, Knoxville, USA*

Organizer: Christian Engelmann

*Oak Ridge National Laboratory, USA*

Organizer: Luc Giraud

*Inria, France**continued in next column***10:50-11:10 Making HPC Systems Resilient with Parallel Objects**

*Esteban Meneses, Costa Rica Institute of Technology, Costa Rica; Laxmikant Kale, University of Illinois at Urbana-Champaign, USA*

**11:15-11:35 Designing and Modelling Selective Replication for Fault-tolerant HPC Applications**

*Omer Subasi, Pacific Northwest National Laboratory, USA*

**11:40-12:00 Resilience With Asynchronous Many Task (AMT) Programming Models**

*Sanjay Chatterjee, General Atomics, USA; Keita Teranishi, Sandia National Laboratories, USA; Vivek Sarkar, Georgia Institute of Technology, USA*

**12:05-12:25 Fault Tolerance Features of a Multi SPMD Scientific Workflow Programming Model**

*Miwako Tsuji, RIKEN Advanced Institute for Computational Science, Japan*

Thursday, March 8

**MS31****Future HPC Operating Systems and Runtimes**

10:50 AM-12:30 PM

Room:52-303

Custom Operating Systems and Runtimes are fundamental to enabling scalable parallel scientific computing applications. The general approach in HPC systems software is to maximize the system resources given to the application while still providing a rich set of application primitives to simplify and enhance application development. The proposed minisymposium gathers 4 talks, covering systems software support for future heterogeneous architectures, specifically new Operating Systems, proving runtimes correct, the challenges of heterogeneity and future systems architectures and their systems software. The objective is to summarize the latest developments in HPC operating systems and runtime research and how the capabilities they provide relate to construction of HPC applications. The goal is to provide applications researchers a view of future systems software and how these new capabilities relate to the construction of scientific applications on future HPC systems.

Organizer: Noah Evans

*Sandia National Laboratories, USA***10:50-11:10 The Challenges of Formally Proving HPC Systems Software**

*Noah Evans, Sandia National Laboratories, USA*

**11:15-11:35 Operating and Runtime Systems Challenges for HPC Systems**

*Arthur McCabe, Oak Ridge National Laboratory, USA*

**11:40-12:00 Vanguard: Maturing the Arm Software Ecosystem for U.S. DOE Supercomputing**

*Kevin Pedretti, Sandia National Laboratories, USA*

**12:05-12:25 Recent Advances in the IHK/McKernel Lightweight Multi-Kernel Operating System**

*Balazs Gerofti, RIKEN, Japan*

Thursday, March 8

## MS32

### Particle Method Based Applications and Their Parallel Optimizations Towards Exascale-Part I of II

10:50 AM-12:30 PM

Room:52-304

#### For Part 2 see MS44

A large number of industrial problems can be modelled using particle-based methods. The particle method based application software is one of the main development areas in recent funded DOE Exascale Computing Projects(ECP). The reason is that particles based methods provide extremely fine-grained parallelism and allow the exploitation of asynchrony. Efficient parallel particle method applications require the efficient implementation when computing with billions and trillions of particles. The main objective of this proposal is to getting together experts to specifically address issues on the following common topics for large scale application simulations, scalable distributed computing (domain decomposition, dynamic load balancing), optimised data mapping (structured and unstructured communication), efficient parallel I/O, nearest neighbour lists searching using tree algorithms(particular for particle refinement) and cell linked lists, particle-to-mesh, and mesh-to-particle interpolation, sparse linear solver for incompressible problems on various modern computing architectures and particle based applications involving complex geometries.

Organizer: Xiaohu Guo

Science and Technology Facilities Council,  
United Kingdom

Organizer: Xiangyu Hu

Technical University of Munich, Germany

Organizer: Tao Cui

Chinese Academy of Sciences, China

Organizer: Petros

Koumoutsakos

ETH Zürich, Switzerland

Organizer: Mark Adams

Lawrence Berkeley National Laboratory,  
USA

Organizer: Ivo F. Sbalzarini

TU Dresden, Germany

#### 10:50-11:10 A Lagrangian Inertial Centroidal Voronoi Particle Method for Dynamic Load Balancing in Particle-Based Simulations

Zhe Ji, Lin Fu, Xiangyu Hu, and

Nikolaus Adams, Technical

University of Munich, Germany

#### 11:15-11:35 Swift: Using Task-Based Parallelism, Asynchronous MPI and Dynamic Workload-Based Domain Decomposition to Achieve Near-Perfect Load-Balancing for Particle-Based Hydrodynamics and Gravity

Matthieu Schaller, Durham University,  
United Kingdom

#### 11:40-12:00 Developing An Extensible, Portable, Scalable Toolkit for Massively Parallel Incompressible SPH

Xiaohu Guo, Science and Technology

Facilities Council, United Kingdom

#### 12:05-12:25 OpenFPM: An Open-source Framework for Highly Scalable Hybrid Particle-mesh Simulations

Ivo F. Sbalzarini and Pietro Incardona,

TU Dresden, Germany

Thursday, March 8

## MS33

### Hierarchical Low Rank Approximation Methods - Part I of VI

10:50 AM-12:30 PM

Room:57-201

#### For Part 2 see MS45

Hierarchical Low-Rank Approximation Methods such as H-matrices, Hierarchical Semi-Seperable matrices, and Hierarchical Interpolative Factorization are increasing in popularity. These methods have near linear complexity and can be used as a fast approximate inverse of both dense matrices arising from integral equations and sparse matrices arising from partial differential equations. Accuracy is controllable by increasing the rank of the approximated blocks or by changing the subdivision criteria of the off-diagonal blocks. Therefore, these methods can be used as robust direct solvers with near linear complexity if high accuracy is chosen or they can be used as preconditioners if lower accuracy is used. With the advent of these methods, large dense linear algebra operations are scarcely needed anymore. In terms of parallelism, the hierarchical nature of these methods pose a challenge. Task-based approaches seem to be effective for shared memory parallelism, while load-balance becomes a problem on distributed memory. The recursive block structure also results in the computation of many small dense blocks. The recent development in batched dense linear algebra libraries facilitates the efficient computation of such small dense blocks. This minisymposium consists of cutting edge research on hierarchical low-rank approximation methods, their applications, parallelization, and extension.

continued in next column

continued on next page



Organizer: Rio Yokota  
Tokyo Institute of Technology, Japan

Organizer: Akihiro Ida  
University of Tokyo, Japan

Organizer: Chao Chen  
Stanford University, USA

Organizer: Eric F. Darve  
Stanford University, USA

**10:50-11:10 An Overview of Fast Sparse Linear Solvers Based on Approximate Factorizations**

*Eric Darve*, Stanford University, USA; *Erik G. Boman*, Sandia National Laboratories, USA; *Chao Chen*, Stanford University, USA; *Mathieu Faverge*, Université de Bordeaux, Inria, LaBRI, France; *Siva Rajamanickam*, Sandia National Laboratories, USA; *Pierre Ramet*, Université de Bordeaux, Inria, LaBRI, France; *Jean Roman*, Inria, France; *Toru Takahashi*, Nagoya University, Japan

**11:15-11:35 Randomized Methods for Accelerating Structured Matrix Computations**

*Gunnar Martinsson*, University of Oxford, United Kingdom

**11:40-12:00 GOFMM: A Parallel Geometry-oblivious N-body Algorithm for Sketching SPD Matrices**

*Chenhan Yu* and *James Levitt*, University of Texas at Austin, USA; *Severin Reiz*, Technical University of Munich, Germany; *George Biros*, University of Texas at Austin, USA

**12:05-12:25 New Variations of a Fast Hierarchical Low-Rank Solver for Sparse Systems**

*Erik G. Boman*, Sandia National Laboratories, USA; *Chao Chen* and *Eric F. Darve*, Stanford University, USA; *Siva Rajamanickam* and *Ray S. Tuminaro*, Sandia National Laboratories, USA

Thursday, March 8

## MS34

### Architecture-Aware Graph Analytics - Part I of II

10:50 AM-12:30 PM

Room: 57-202

**For Part 2 see MS46**

Graphs are used to model, manipulate, distribute, and analyze the complex relations within datasets emerging from various fields. These graphs can have compact but irregular structures, resulting in irregular data access patterns for graph analysis algorithm. Such irregularities make it difficult to achieve efficient parallelism and data access locality, and these difficulties correspondingly increase with the increasing complexity of modern computer architectures. Current architectures provide massive amount of parallelisms through GPUs and KNLs, along with complex memory hierarchies. Achieving high performance in graph analysis requires algorithms to consider both the massive amount of available parallelisms as well as effectively exploit the memory hierarchy. This minisymposium focuses on the effects of such architectural shifts on the design of parallel graph-based algorithms.

Organizer: Mehmet Deveci  
Sandia National Laboratories, USA

Organizer: George M. Slota  
Rensselaer Polytechnic Institute, USA

**10:50-11:10 Exploiting High Bandwidth Memory for Graph Algorithms**

*George M. Slota*, Rensselaer Polytechnic Institute, USA; *Siva Rajamanickam*, Cynthia Phillips, and *Jonathan Berry*, Sandia National Laboratories, USA

**11:15-11:35 Processing Massive Scale-Free Graphs on HPC with NVRAM**

*Roger Pearce*, Lawrence Livermore National Laboratory, USA

**11:40-12:00 TriData: High Performance Linear Algebra-Based Data Analytics**

*Michael Wolf*, Daniel M. Dunlavy, Rich Lehoucq, and Jonathan W. Berry, Sandia National Laboratories, USA; *Daniel Bourgeois*, Rice University, USA

**12:05-12:25 Scalable Graph Alignment on Modern Architectures**

*Umit V. Catalyurek* and *Abdurrahman Yasar*, Georgia Institute of Technology, USA; *Bora Ucar*, LIP-ENS Lyon, France

*continued in next column*



Thursday, March 8

## CP5

### Scalable Solvers - Part II of II

10:50 AM-12:30 PM

Room:52-101

Chair: Paolo Bientinesi, RWTH-Aachen, Germany

#### 10:50-11:10 Application of Transactional Memory to Scalable Linear Solvers

Barna Bihari and Ulrike Yang, Lawrence Livermore National Laboratory, USA

#### 11:15-11:35 Performance Evaluation of a Modified Communication-Avoiding Generalized Minimal Residual Method on Many Core Platforms

Yasuhiro Idomura, Takuya Ina, Akie Mayumi, and Susumu Yamada, Japan Atomic Energy Agency, Japan; Kazuya Matsumoto, University of Aizu, Japan; Yuuichi Asahi, CEA/DSM/IRFM Cadarache, France; Toshiyuki Imamura, RIKEN, Japan

#### 11:40-12:00 A Novel Adaptive Algebraic Multigrid Preconditioner for Large-scale Problems

Victor A. Paludetto Magri, Massimiliano Ferronato, Andrea Franceschini, and Carlo Janna, University of Padova, Italy

#### 12:05-12:25 Autotuning Qr Factorizations

Wissam M. Sid-Lakhdar, Lawrence Berkeley National Laboratory, USA

## Lunch Break

12:30 PM-1:45 PM

Attendees On Their Own

Thursday, March 8

## IP3

### Toward Community Software Ecosystems for High-Performance Computational Science

1:45 PM-2:30 PM

Room:63-201/202

Chair: Miriam Mehl, University of Stuttgart, Germany

Software---crosscutting technology that connects advances in mathematics, computer science, and domain-specific science and engineering---is a cornerstone of long-term collaboration and scientific progress. As we leverage unprecedented high-performance computing resources to work toward predictive science, software complexity is increasing due to multiphysics and multiscale modeling, the coupling of simulations and data analytics, and the demand for greater reproducibility in the midst of disruptive architectural changes. Applications increasingly require the combined use of independent software packages, which have diverse sponsors, priorities, and processes for development and release. These challenges create the unique opportunity to improve how scientific software is designed, developed, and sustained---with explicit work toward scientific software ecosystems. This presentation will introduce the xSDK, or Extreme-scale Scientific Software Development Kit, where community-defined policies are increasing the quality and interoperability across numerical libraries as needed by the DOE Exascale Computing Project. We will also discuss complementary efforts to increase scientific software productivity and sustainability.

Lois Curfman McInnes

Argonne National Laboratory, USA

## Intermission

2:30 PM-2:40 PM

Thursday, March 8

## MS35

### Language-Level Abstractions for Asynchronous Runtimes - Part II of II

2:40 PM-4:20 PM

Room:52-102

#### For Part 1 see MS23

Next-generation architectures are motivating more asynchronous, multi-threaded applications. Programming models and languages are now incorporating asynchronous tasking and performance portability as first-class design considerations. Expressive programming models should enable runtime flexibility and therefore high performance across different application and platforms. With expressive models, applications can focus on algorithms with runtimes developed by system experts managing execution resources. These sessions explore progress in developing intuitive programming models and abstractions that map to underlying asynchronous, distributed, or multi-threaded runtimes.

Organizer: David S. Hollman  
Sandia National Laboratories, USA

Organizer: Jonathan J. Lifflander

Sandia National Laboratories, USA

#### 2:40-3:00 The C++ Parallel Programming Model

Bryce Lebach, NVIDIA, USA

#### 3:05-3:25 StarPU and KStar, Synergies Between a Task-Based Runtime System and An OpenMP Compiler

Olivier Aumage, Inria, France

#### 3:30-3:50 Task-Based Parallel Programming with PyCOMPSs

Rosa M. Badia and Javier Conejero, Barcelona Supercomputing Center, Spain

#### 3:55-4:15 Optimizing Recursive Programs for Memory Hierarchy Locality

Jonathan J. Lifflander, Sandia National Laboratories, USA

Thursday, March 8

**MS36****Kokkos Enables Productivity and Performance Portability- Part I of II**

2:40 PM-4:20 PM

Room:52-103

**For Part 2 see MS48**

Performance portability for shared memory parallelism across an evolving diversity of manycore architectures (CPUs, Intel Xeon Phi, NVIDIA GPU, AMD GPU) is a major goal for many applications and libraries. Four years after Kokkos' production release ([github.com/kokkos/kokkos](https://github.com/kokkos/kokkos)) numerous applications and libraries are realizing Kokkos' promise for performance portability and productivity. In this minisymposium we present Kokkos' capabilities and "in the field" success stories from application and library developers.

Organizer: H. Carter Edwards  
Sandia National Laboratories, USA

**2:40-3:00 Overview of Kokkos**

H. Carter Edwards, Christian Trott,  
Daniel Sunderland, Daniel Ibanez, and  
Nathan Ellingwood, Sandia National  
Laboratories, USA

**3:05-3:25 KokkosKernels: Math Building Blocks for Performance Portable Applications**

Christian Trott, Siva Rajamanickam,  
Kyungjoo Kim, and Mehmet Deveci,  
Sandia National Laboratories, USA

**3:30-3:50 Implementing the Lattice Quantum Chromodynamics Wilson-Dslash Kernel Using Kokkos**

Balint Joo, Jefferson National Accelerator  
Laboratory, USA; Thorsten Kurth  
and Jack Deslippe, National Energy  
Research Scientific Computing Center,  
USA; Kate Clark, NVIDIA, USA;  
Daniel Ibanez and Daniel Sunderland,  
Sandia National Laboratories, USA

**3:55-4:15 Performance Portability in SPARC – Using Kokkos in a Compressible CFD Code**

James Overfelt, Micah Howard, Steve  
Bova, Travis Fisher, Mark Hoemmen,  
Derek Dinzl, Andrew Bradley, and  
Ross Wagnild, Sandia National  
Laboratories, USA

Thursday, March 8

**MS37****Large-Scale Electronic Structure Calculations - Part II of II**

2:40 PM-4:20 PM

Room:52-104

**For Part 1 see MS25**

Electronic structure calculations and their applications are among the most challenging and computationally demanding science and engineering problems. This minisymposium aims at presenting and discussing new electronic structure calculation algorithms and parallel implementations that are suitable for modern computing architectures. It also highlights recent software development efforts that enable new algorithms to be easily implemented on massively parallel many-core systems and GPUs. The presentations included in this minisymposium cover a variety of topics ranging from linear scaling density functional theory based approaches to fast Hartree-Fock and other accurate wavefunction methods to fast matrix computations required in electronic structure calculations and software infrastructures required to support a user/developer-friendly interface between application software and numerical libraries. They will highlight both the latest success stories and the remaining challenges.

Organizer: Chao Yang

Lawrence Berkeley National Laboratory,  
USA

**2:40-3:00 Quantum Chemistry Methods with Gaussian Basis**

Edoardo Apra, Pacific Northwest National  
Laboratory, USA

**3:05-3:25 100-Nano-Meter-Scale Electronic Structure Calculation for Organic Device Materials**

Takeo Hoshi, Tottori University, Japan

**3:30-3:50 Large Scale Discontinuous Galerkin Density Functional Theory (DGDFT) Method on GPU Supercomputer**

Weile Jia, University of California,  
Berkeley, USA

**3:55-4:15 Applications of Matrix Functions to Electronic Structure Calculations**

William Dawson, RIKEN Advanced  
Institute for Computational Science,  
Japan

Thursday, March 8

**MS38****Deep Learning from HPC Perspective: Opportunities and Challenges - Part II of II**

2:40 PM-4:20 PM

Room:52-201

**For Part 1 see MS26**

Deep learning is becoming an indispensable tool in a wide variety of scientific and engineering domains thanks to its groundbreaking advancement in data classification, recognition, and prediction. For instance, machine learning models such as convolutional networks and recurrent neural networks have been reported to demonstrate competitive or even outperforming accuracy as traditional numerical simulations in various scientific domains. However, training such models is an extremely compute intensive task, taking hours and even days with latest generations of processors. Consequently, parallel and distributed model training has been one of the hot topics in deep learning. In this minisymposium, we will discuss opportunities and challenges in high-performance deep learning from applications and systems perspectives. We will first motivate the importance of high performance deep learning by presenting use cases of deep learning in science and engineering problems, and then present recent efforts in scaling deep learning in large-scale HPC systems, including both systems level as well as algorithmic innovations. Our minisymposium will allow attendees to learn about the state of the art and future directions in large-scale deep learning.

Organizer: Naoya Maruyama  
Lawrence Livermore National Laboratory,  
USA

Organizer: Brian Van Essen  
Lawrence Livermore National Laboratory,  
USA

Organizer: Aleksandr Drozd  
Tokyo Institute of Technology, Japan

Organizer: Mohamed Wahib  
RIKEN, Japan

continued on next page

Thursday, March 8

## MS38

### Deep Learning from HPC Perspective: Opportunities and Challenges - Part II of II

2:40 PM-4:20 PM

continued

#### 2:40-3:00 Scalable Hyperparameter Search for Deep Learning in Scientific Applications

*Prasanna Balaprakash*, Argonne National Laboratory, USA

#### 3:05-3:25 Be Fast & Productive: Accelerating Deep Learning Without Sacrificing Productivity

*Keisuke Fukuda*, Preferred Networks, Japan; *Shuji Suzuki* and *Takuya Akiba*, Preferred Networks, Inc., Japan

#### 3:30-3:50 Faster, Smaller, and More Energy-Efficient Inference Using Codebook-based Quantization and FPGAs

*Mikhail Isaev*, *Jeffrey Young*, and *Rich Vuduc*, Georgia Institute of Technology, USA

#### 3:55-4:15 Trends in Parallel and Distributed Deep Learning

*Tal Ben-Nun*, ETH Zürich, Switzerland

Thursday, March 8

## MS39

### Parallel Lattice Boltzmann Methods - Part II of II

2:40 PM-4:20 PM

Room:52-202

#### For Part 1 see MS27

In the field of computational fluid dynamics lattice Boltzmann methods (LBM) pose an alternative to classical Navier-Stokes solvers for the simulation of time-dependent, incompressible flows. The locality of the LBM makes it an ideal method for distributed memory system and allows for implementations with excellent scalability on GPUs and CPUs. LBM have been applied to a broad range of problems, from complex geometries like porous media to studies of turbulent flows. This Minisymposium offers an overview of the recent advancements in leading LBM implementations. We cover high performance implementations of advanced topics like refinement or particle-coupling methods on CPUs and GPUs.

Organizer: *Christian Godenschwager*

*University of Erlangen-Nuernberg, Germany*

Organizer: *Takayuki Aoki*

*Tokyo Institute of Technology, Japan*

Organizer: *Ulrich Rüde*

*University of Erlangen-Nuernberg, Germany*

#### 2:40-3:00 Development of LBM Code Toward Large-Scale Aerodynamics/Aeroacoustics Simulations

*Takashi Ishida*, Japan Aerospace Exploration Agency, Japan

#### 3:05-3:25 OpenLB: Parallel Open Source Lattice Boltzmann Software

*Mathias J. Krause*, Karlsruhe Institute of Technology, Germany

#### 3:30-3:50 Optimising Massively Parallel Lattice-Boltzmann Codes for Use in Time-Critical Clinical Blood Flow Modelling

*Peter Coveney*, University College London, United Kingdom

#### 3:55-4:15 Large-Scale Lattice Boltzmann Simulations of Particulate Flows

*Christoph Rettinger*, *Sebastian Eibl*, *Christian Godenschwager*, and *Ulrich Rüde*, University of Erlangen-Nuernberg, Germany

*continued in next column*

Thursday, March 8

## MS40

### Scalable Communication-Avoiding and -Hiding Krylov Subspace Methods - Part II of II

2:40 PM-4:20 PM

Room:52-204

#### For Part 1 see MS28

With HPC hardware advancing towards the exascale regime, the gap between computation (i.e. flops) and communication (i.e. internode communication, as well as data movement within the memory hierarchy) keeps steadily increasing. To address the growing time and energy costs of communication, classic numerical algorithms - which were primarily optimized for flop performance - have to be thoroughly revised or even reformulated entirely to be able to benefit from the increasing levels of parallelism. Parallel scalability can be improved by either reducing the number of global synchronization bottlenecks (communication avoiding), and/or overlapping communication latency with computations where possible (communication hiding). Although many of these re-engineered algorithms are equivalent to their classic counterparts in exact arithmetic, their numerical stability in a practical finite precision framework is often reduced. This minisymposium aims to give an overview of recent developments in the design, implementation and numerical analysis of sparse linear algebra solvers, with a strong emphasis on communication-avoiding and -hiding Krylov subspace methods for solving linear systems.

Organizer: Siegfried Cools

University of Antwerp, Belgium

Organizer: Erin C. Carson

New York University, USA

#### 2:40-3:00 Impact of Noise Models on Pipelined Krylov Methods

Hannah Morgan, University of Chicago, USA; Matthew G. Knepley, University of Buffalo, USA; Patrick Sanan, Università della Svizzera italiana, Switzerland; Ridgway Scott, University of Chicago, USA

#### 3:05-3:25 Scalable Krylov Methods for Spectral Graph Partitioning

Pieter Ghysels, Lawrence Berkeley National Laboratory, USA; Siegfried Cools, University of Antwerp, Belgium; Roel Van Beeumen, KU Leuven, Belgium; Mathias Jacquelin and Esmond G. Ng, Lawrence Berkeley National Laboratory, USA

#### 3:30-3:50 Using Non-Blocking Communication to Achieve Scalability for Preconditioned Conjugate Gradient Methods

William D. Gropp and Paul Eller, University of Illinois at Urbana-Champaign, USA

#### 3:55-4:15 Performance of S-Step and Pipelined Krylov Methods

Piotr Luszczek and Ichitaro Yamazaki, University of Tennessee, Knoxville, USA; Mark Hoemmen, Sandia National Laboratories, USA; Jack J. Dongarra, University of Tennessee, ORNL, and University of Manchester, USA

Thursday, March 8

## MS41

### Performance Engineering from the Node Level to the Extreme Scale- Part II of II

2:40 PM-4:20 PM

Room:52-301

#### For Part 1 see MS29

Achieving hardware and energy efficiency is important for current large-scale numerical simulations and will be a key component in the exascale era. In a world of heterogeneous, highly parallel computer architectures with deep memory hierarchies, complex application scenarios, and a broad spectrum of algorithms, software developers can no longer rely on mystic “black-box” performance engineering computer programs. Instead, a thorough analysis and understanding of the complex interaction of software, data structures, algorithms, and hardware features, a.k.a. performance engineering, is required for implementing codes that allow for portable performance on the computer generations to come. The minisymposium addresses a broad range of topics in performance engineering for modern HPC architectures, ranging from recent advances in performance models and tools supporting a “white-box” performance engineering approach to application performance tuning cases studies. The presentations should point out the potentials and limitations of “white-box” performance engineering activities and demonstrate the wide spectrum of performance models used in the PE activities including simple performance expectations, automatic model parameter selections, and analytic models.

Organizer: Gerhard Wellein

Erlangen Regional Computing Center, Germany

Organizer: Georg Hager

Erlangen Regional Computing Center, Germany

continued in next column

continued on next page



Thursday, March 8

## MS41

### Performance Engineering from the Node Level to the Extreme Scale- Part II of II

2:40 PM-4:20 PM

continued

#### 2:40-3:00 Designing an Algorithm with a Tuning Knob that Controls its Power Consumption

Sara Karamati, Jeffrey Young, and Richard Vuduc, Georgia Institute of Technology, USA

#### 3:05-3:25 Extending the Roofline Model: Bottleneck Analysis with Microarchitectural Constraints

Victoria Caparros Cabezas and Markus Pueschel, ETH Zürich, Switzerland

#### 3:30-3:50 Performance Engineering of Emerging Memory Systems

Jeffrey S. Vetter, Oak Ridge National Laboratory, USA

#### 3:55-4:15 Isoefficiency in Practice: Configuring and Understanding the Performance of Task-Based Applications

Sergei Shudler and Alexandru Calotoiu, Technische Universität Darmstadt, Germany; Torsten Hoefler, ETH Zürich, Switzerland; Felix Wolf, Technische Universität Darmstadt, Germany

Thursday, March 8

## MS42

### Resilience for Extreme Scale Computing - Part IV of IV

2:40 PM-4:20 PM

Room:52-302

#### For Part 3 see MS30

The reliability of large scale HPC systems has been a major concern due to the increasing complexity of hardware and system software in addition to the tight power budget for system operations. Despite the reliability improvement in the recent hardware, the users are still responsible for mitigating the failures in their applications. In fact, the US DOE Exascale Computing Project aims at achieving 1 week of meantime between “application” failures, assuming checkpoint and restart is integrated by default. In this minisymposium, we will discuss the recent progress in the resilience for high performance computing systems from a variety of perspectives, including hardware, systems, runtime, algorithm and applications. We will also discuss challenges and opportunities that facilitate resilience-centric HPC system co-design for reducing the operational budget of these systems while maintaining their capability.

Organizer: Keita Teranishi  
Sandia National Laboratories, USA

Organizer: Christian Engelmann  
Oak Ridge National Laboratory, USA

Organizer: George Bosilca  
University of Tennessee, Knoxville, USA

Organizer: Emmanuel Agullo  
Inria, France

Organizer: Luc Giraud  
Inria, France

#### 2:40-3:00 The Case for Resilience Support in MPI

George Bosilca, Aurélien Bouteiller, and Thomas Herault, University of Tennessee, Knoxville, USA

#### 3:05-3:25 Evaluating the Fenix MPI Fault Tolerance Programming Framework

Marc Gamell, Intel Corporation, USA; Keita Teranishi, Sandia National Laboratories, USA; Rob Van Der Wijngaart, Intel Corporation, USA; Manish Parashar, Rutgers University, USA

#### 3:30-3:50 Local Rollback of MPI Applications by Combining Application-Level Checkpointing and Message Logging

Nuria Losada, University of A Coruña, Spain; George Bosilca and Aurélien Bouteiller, University of Tennessee, Knoxville, USA; Patricia González, University of A Coruña, Spain; Thomas Héault, University of Tennessee, Knoxville, USA; María J. Martín, University of A Coruña, Spain

#### 3:55-4:15 GASPI -- Scaling in Dynamic Environments

Mirko Rhan and Franz-Josef Pfreundt, Fraunhofer Institute for Industrial Mathematics, Germany

continued in next column



Thursday, March 8

## MS43

### Parallel Numerical Linear Algebra for Future Extreme-Scale Systems - Part I of II

2:40 PM-4:20 PM

Room:52-303

#### For Part 2 see MS55

We describe recent developments in task-based algorithms for the solution of dense linear systems and the solution of both symmetric and unsymmetric dense eigenproblems. By using directed acyclic graphs (DAGs) we are able to avoid synchronization and obtain high levels of parallelism. Our algorithms also keep communication to a minimum, sometimes achieving a provably theoretical minimum. In the second minisymposium, we will consider the solution of sparse linear systems. For the direct solution of symmetrically structured systems, we express the computation as a DAG and use dense matrix kernels. We study both multifrontal and supernodal methods and show how numerical pivoting can be accommodated without losing much parallelism. We also discuss algorithms based on a Markowitz threshold approach for highly unsymmetric systems. For the iterative solvers, we study robust algebraic preconditioners that are highly scalable and examine a way to expand the Krylov space for better convergence while maintaining parallelism and reducing communication. We then consider hybrid techniques that combine the best features of direct and iterative methods. To aid in portability for a range of computing systems we use the runtime systems, StarPU and PaRSEC and we show the performance of prototype codes on a range of research and industrial applications. These minisymposia present research done in NLA-FET, a Horizon 2020 FET-HPC project funded by the European Union.

Organizer: Iain Duff

Science & Technology Facilities Council,  
United Kingdom and CERFACS, Toulouse,  
France

#### 2:40-3:00 Dense Linear Systems for Extreme Scale

Jack J. Dongarra, University of Tennessee, ORNL, and University of Manchester, USA

#### 3:05-3:25 The Batched Blas

Sven J. Hammarling and Mawussi Zounon, The University of Manchester, UK

#### 3:30-3:50 Parallel Algorithms for Dense Eigenvalue Problems - Progress on Schur Form Computations

Bo T. Kågström, Björn Adlerborn, Mahmoud Eljammaly, Lars Karlsson, Carl Christian K. Mikkelsen, Mirko Myllykoski, and Angelika Schwarz, Umeå University, Sweden

#### 3:55-4:15 Task Based Robust Computation of Eigenvectors

Carl Christian Kjelgaard Mikkelsen, Lars Karlsson, Mirko Myllykoski, and Angelika Schwarz, Umeå University, Sweden

Thursday, March 8

## MS44

### Particle Method Based Applications and Their Parallel Optimizations Towards Exascale- Part II of II

2:40 PM-4:20 PM

Room:52-304

#### For Part 1 see MS32

A large number of industrial problems can be modelled using particle-based methods. The particle method based application software is one of the main development areas in recent funded DOE Exascale Computing Projects (ECP). The reason is that particles based methods provide extremely fine-grained parallelism and allow the exploitation of asynchrony. Efficient parallel particle method applications require the efficient implementation when computing with billions and trillions of particles. The main objective of this proposal is to get together experts to specifically address issues on the following common topics for large scale application simulations, scalable distributed computing (domain decomposition, dynamic load balancing), optimised data mapping (structured and unstructured communication), efficient parallel I/O, nearest neighbour lists searching using tree algorithms (particular for particle refinement) and cell linked lists, particle-to-mesh, and mesh-to-particle interpolation, sparse linear solver for incompressible problems on various modern computing architectures and particle based applications involving complex geometries.

continued in next column

continued on next page

Thursday, March 8

## MS44

### Particle Method Based Applications and Their Parallel Optimizations Towards Exascale-Part II of II

2:40 PM-4:20 PM

continued

Organizer: Xiaohu Guo  
*Science and Technology Facilities Council, United Kingdom*

Organizer: Xiangyu Hu  
*Technical University of Munich, Germany*

Organizer: Tao Cui  
*Chinese Academy of Sciences, China*

Organizer: Petros Koumoutsakos  
*ETH Zürich, Switzerland*

Organizer: Mark Adams  
*Lawrence Berkeley National Laboratory, USA*

#### 2:40-3:00 Parallel Particle-in-Cell Method for Solving Non-Relativistic Vlasov-Maxwell System

*Cui Tao, State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University*

#### 3:05-3:25 A Landau Collision Integral Solver with Adaptivity on Emerging Architectures with Coupling to PIC Vlasov Methods

*Mark Adams, Lawrence Berkeley National Laboratory, USA*

#### 3:30-3:50 Massively Parallel Flow Simulations Using Particles and Grids

*Petros Koumoutsakos, ETH Zürich, Switzerland*

#### 3:55-4:15 Molecular Dynamics Towards ExaScale Age

*Alin Marin Elena, Science and Technology Facilities Council, United Kingdom*

Thursday, March 8

## MS45

### Hierarchical Low Rank Approximation Methods - Part II of VI

2:40 PM-4:20 PM

*Room:57-201*

**For Part 1 see MS33**

**For Part 3 see MS57**

Hierarchical Low-Rank Approximation Methods such as H-matrices, Hierarchical Semi-Separable matrices, and Hierarchical Interpolative Factorization are increasing in popularity. These methods have near linear complexity and can be used as a fast approximate inverse of both dense matrices arising from integral equations and sparse matrices arising from partial differential equations. Accuracy is controllable by increasing the rank of the approximated blocks or by changing the subdivision criteria of the off-diagonal blocks. Therefore, these methods can be used as robust direct solvers with near linear complexity if high accuracy is chosen or they can be used as preconditioners if lower accuracy is used. With the advent of these methods, large dense linear algebra operations are scarcely needed anymore. In terms of parallelism, the hierarchical nature of these methods pose a challenge. Task-based approaches seem to be effective for shared memory parallelism, while load-balance becomes a problem on distributed memory. The recursive block structure also results in the computation of many small dense blocks. The recent development in batched dense linear algebra libraries facilitates the efficient computation of such small dense blocks. This minisymposium consists of cutting edge research on hierarchical low-rank approximation methods, their applications, parallelization, and extension.

Organizer: Rio Yokota  
*Tokyo Institute of Technology, Japan*

Organizer: Akihiro Ida  
*University of Tokyo, Japan*

Organizer: Chao Chen  
*Stanford University, USA*

Organizer: Eric F. Darve  
*Stanford University, USA*

#### 2:40-3:00 Parallel Task Frameworks for FMM

*Simon McIntosh-Smith, University of Bristol, United Kingdom*

#### 3:05-3:25 Performance Evaluations and Optimizations of H-Matrices for Many-Core Processors

*Tetsuya Hoshino, Akihiro Ida, and Toshihiro Hanawa, University of Tokyo, Japan*

#### 3:30-3:50 Low-Rank Matrix Computations on Manycore Systems

*Hatem Ltaief, King Abdul Aziz University, Saudi Arabia; Kadir Akbudak, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; Ali M. Charara, King Abdul Aziz University, Saudi Arabia; Aleksandr Mikhalev and Sameh Abdullah, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; David E. Keyes, King Abdul Aziz University, Saudi Arabia*

#### 3:55-4:15 Accumulator Based Task-Parallel H-Factorization

*Ronald Kriemann, Max Planck Institute for Mathematics in the Sciences, Germany*

*continued in next column*

Thursday, March 8

**MS46****Architecture-Aware Graph Analytics- Part II of II**

2:40 PM-4:20 PM

Room:57-202

**For Part 1 see MS34**

Graphs are used to model, manipulate, distribute, and analyze the complex relations within datasets emerging from various fields. These graphs can have compact but irregular structures, resulting in irregular data access patterns for graph analysis algorithm. Such irregularities make it difficult to achieve efficient parallelism and data access locality, and these difficulties correspondingly increase with the increasing complexity of modern computer architectures. Current architectures provide massive amount of parallelisms through GPUs and KNLs, along with complex memory hierarchies. Achieving high performance in graph analysis requires algorithms to consider both the massive amount of available parallelisms as well as effectively exploit the memory hierarchy. This minisymposium focuses on the effects of such architectural shifts on the design of parallel graph-based algorithms.

Organizer: Mehmet Deveci

Sandia National Laboratories, USA

Organizer: George M. Slota

Rensselaer Polytechnic Institute, USA

**2:40-3:00 Underappreciated Bottlenecks for Graph Algorithm Communication**

Scott Beamer, Lawrence Berkeley National Laboratory, USA

**3:05-3:25 SlimSell: A Vectorized Graph Representation for Breadth-First Search**

Torsten Hoefer, ETH Zürich, Switzerland

**3:30-3:50 Triangle Counting and Truss Decomposition on Modern Parallel Architectures**

Ancy Tom and Shaden Smith, University of Minnesota, USA; George Karypis, University of Minnesota and Army HPC Research Center, USA

**3:55-4:15 Dynamic Load Balancing of Massively Parallel Graphs for Scientific Computing on Many Core and Gpu Based Systems**

Cameron Smith, Gerrett Diamond, George M. Slota, and Mark Shepard, Rensselaer Polytechnic Institute, USA

Thursday, March 8

**CP6****Meshes**

2:40 PM-4:20 PM

Room:52-101

Chair: Kohei Fujita, University of Tokyo, Japan

**2:40-3:00 Optimizations of Implicit Unstructured Aerodynamics Computations for Emerging Many and Multi-Core HPC Architectures**

Mohammed A. Al Farhan, King Abdullah University of Science &amp; Technology (KAUST), Saudi Arabia; David E. Keyes, King Abdul Aziz University, Saudi Arabia

**3:05-3:25 Parallel Efficient Mesh Deformation Method Based on Support Vector Regression**

Xiang Gao, Chuanfu Xu, and Dali Li, National University of Defense Technology, China

**3:30-3:50 Multithreaded Mesh Generation**

Jean-François Remacle, Université Catholique de Louvain, Belgium

**Coffee Break**

4:20 PM-4:50 PM



Room:63-1F

Thursday, March 8

**MS47****Portability Efforts in Modern Fortran**

4:50 PM-6:30 PM

Room:52-102

Fortran codes maintain a significant presence in scientific community. However, most new developments in programming models for next generation systems concentrate on other languages like C++ thus leaving Fortran behind. In this minisymposium, we will discuss the strategies and current efforts to move Fortran codes to next generation systems. The talks will span a wide range, from multi-language approaches including an automatic construction of portable Fortran interfaces to C++, to Fortran-specific approaches for modernizing a large codebase, with many examples of lessons learned.

Organizer: Andrey Prokopenko

Oak Ridge National Laboratory, USA

Organizer: Katherine J. Evans

Oak Ridge National Laboratory, USA

Organizer: Michael Heroux

Sandia National Laboratories, USA

**4:50-5:10 ForTrilinos: Fortran Interface for Trilinos Library**

Andrey Prokopenko and Katherine

J. Evans, Oak Ridge National Laboratory, USA; Michael Heroux, Sandia National Laboratories, USA; Seth Johnson, Oak Ridge National Laboratory, USA

**5:15-5:35 Enamr: XRage's Code Modernization Strategy**

Geoff Womeldorff, Joshua Payne, Zach Medin, Sriram Swaminarayan, and Galen Shipman, Los Alamos National Laboratory, USA

continued on next page

Thursday, March 8

## MS47

### Portability Efforts in Modern Fortran

4:50 PM-6:30 PM

continued

#### 5:40-6:00 Unstructured-Grid CFD Algorithms on Many-Core Architectures

Aaron Walden and Eric Nielsen, NASA Langley Research Center, USA; M Zubair, Old Dominion University, USA; John Linford, Paratools Inc., USA; John Wohlbier, HPCMP PETTT, Engility Corp., USA; Justin Luitjens, NVIDIA, USA; Izaak Beekman, Samuel Khuviz, and Sameer Shende, Paratools Inc., USA

#### 6:05-6:25 Accelerated HPC Considerations for Modern Fortran

Matthew R. Norman, Oak Ridge National Laboratory, USA; Robert Pincus, University of Colorado Boulder, USA

Thursday, March 8

## MS48

### Kokkos Enables Productivity and Performance Portability- Part II of II

4:50 PM-6:30 PM

Room:52-103

#### For Part 1 see MS36

Performance portability for shared memory parallelism across an evolving diversity of manycore architectures (CPUs, Intel Xeon Phi, NVIDIA GPU, AMD GPU) is a major goal for many applications and libraries. Four years after Kokkos' production release ([github.com/kokkos/kokkos](https://github.com/kokkos/kokkos)) numerous applications and libraries are realizing Kokkos' promise for performance portability and productivity. In this minisymposium we present Kokkos' capabilities and "in the field" success stories from application and library developers.

Organizer: H. Carter Edwards  
Sandia National Laboratories, USA

#### 4:50-5:10 Intrepid2: a Performance-Portable Package for Compatible High-Order Finite Element Discretizations

Mauro Perego, Kyungjoo Kim, Nathan Ellingwood, and Kara Peterson, Sandia National Laboratories, USA

#### 5:15-5:35 Kokkos Implementation of Shock Hydrodynamics with Complex Topology Changes and Electric Potential in Alexa

Daniel Ibanez, Glen Hansen, Thomas Voth, Edward Love, and James Overfelt, Sandia National Laboratories, USA

#### 5:40-6:00 Performance Portability of Climate Applications Within the Albany Finite Element Code Using Kokkos

Jerry Watkins and Irina K. Tezaur, Sandia National Laboratories, USA

#### 6:05-6:25 Performant Inline NLTE Calculations for Thermal Radiation Transport

Daniel A. Holladay, Los Alamos National Laboratory, USA

Thursday, March 8

## MS49

### Development of Numerical Computing Software on Emerging Computing Platforms

4:50 PM-6:30 PM

Room:52-104

Technologies of numerical computations based on GPU and XeonPhi has become matured in recent years, and they are no longer special in HPC. Meanwhile, several new architectures and computing platforms have been developed for the next generation supercomputing. In this minisymposium, we will discuss the development of numerical libraries for these computing platforms, which have not been explored enough compared to GPU and XeonPhi, such as Sunway, PEZY-SC/SC2, ARM, and FPGA. In addition, we will cover several case studies to understand the scalability and efficiency of the new platforms and identify the opportunity and challenges for future development.

Organizer: Daichi Mukunoki  
RIKEN, Japan

Organizer: Yusuke Hirota  
RIKEN, Japan

#### 4:50-5:10 Performance Analysis of 2.5D-PDGEMM on the K Computer

Daichi Mukunoki and Toshiyuki Imamura, RIKEN, Japan

#### 5:15-5:35 OpenArray: An Auto-parallel Numerical Library on Sunway TaihuLight

Shixun Zhang and Jiuzhou Tang, National Supercomputing Center, Wuxi, China; Xing Huang, Xiaomeng Huang, and Qi Wu, Tsinghua University, P. R. China

#### 5:40-6:00 Performance Evaluation of Application Kernel using ARM SVE

Tetsuya Odajima, Yuetsu Kodama, Motohiko Matsuda, Miwako Tsuji, Jinpil Lee, and Mitsuhisa Sato, RIKEN Advanced Institute for Computational Science, Japan

#### 6:05-6:25 Implementation and Evaluation of BLAS on PEZY-SC/SC2 Processor

Toshiaki Hishinuma, Ryo Sakamoto, and Hitoshi Ishikawa, PEZY Computing, Japan



Thursday, March 8

## MS50

### Parallel Simulations in Life Sciences

4:50 PM-6:30 PM

Room:52-201

The main objective of this minisymposium is to bring international scientists together working in the area of computational modelling with applications in Life Sciences. Numerical methods include continuum as well as particle-based methods. The goal of minisymposium is, on one hand, to share state-of-the-art results in various applications of parallel simulation methods and, on the other to discuss technical issues of the computational modelling.

Organizer: Igor V. Pivkin

*Università della Svizzera italiana,  
Switzerland*

Organizer: Alexander Alexeev

*Georgia Institute of Technology, USA*

#### 4:50-5:10 Meshfree Simulations of Complex Flows Using General Finite Differences

Yaroslav Vasylyv and Alexander Alexeev,  
*Georgia Institute of Technology, USA*

#### 5:15-5:35 Coarse-Grained Simulations of Biomembranes Using Meshless Membrane Model

*Hiroshi Noguchi, University of Tokyo,  
Japan*

#### 5:40-6:00 Numerical Simulations of Respiratory Aerodynamics in the Lung of Japanese Quail

*Masanori Nakamura, Nagoya University,  
Japan; Masahiro Yano, Saitama  
University, Japan; Naoki Tateishi,  
Osaka University, Japan; Shunichi  
Ishida and Yohsuke Imai, Tohoku  
University, Japan*

#### 6:05-6:25 Coarse-Grained Simulations of Eukaryotic Cell in Flow Using Dissipative Particle Dynamics

*Igor V. Pivkin, Università della Svizzera  
italiana, Switzerland*

Thursday, March 8

## MS51

### Parallel Data Processing in Geophysics - Part I of II

4:50 PM-6:30 PM

Room:52-202

#### For Part 2 see MS63

Modern approaches to the processing and interpretation of seismic data for the reconstruction of 3D geological structures are inconceivable without the use of high-performance computing systems with parallel architecture. One of the key features here is the strong data dependence, which can appear at various stages of processing and interpretation. The simplest situation appears at the level of the model-data upload, when the input model-data should be distributed independently between parallel processes, which needs to the data exchange during the computations. This leads to the necessity of data flow optimization to reduce computational time and RAM requirements. Thus, implementation of such algorithms requires use of the modern efficient approaches to parallel data processing. The objective of this minisymposium is to discuss new advances in parallel seismic data processing and seismic modelling.

Organizer: Vadim Lisitsa

*Institute of Petroleum Geology &  
Geophysics of SB RAS, Russia*

Organizer: Galina Reshetova

*Institute of Computational Mathematics  
and Mathematical Geophysics, Russia*

#### 4:50-5:10 Data Processing in Seismic Imaging

Alexander Pleshkevich, Central  
Geophysical Expedition, Russia;  
*Vadim Lisitsa, Institute of Petroleum  
Geology & Geophysics of SB  
RAS, Russia; Dmitry Vishnevsky,  
Trofimuk Institute of Petroleum  
Geology and Geophysics of SB  
RAS, Russia; Vadim Levchenko,  
Keldysh Institute of Applied  
Mathematics of RAS, Russia and  
Moscow Institute of Physics and  
Technology, Russia*

#### 5:15-5:35 Direct Cluster Solver for 3D Helmholtz Equation with Hss-Based Data Compression

*Sergey Solov'yev and Viktor Kostin,  
Institute of Petroleum Geology &  
Geophysics of SB RAS, Russia;  
Andrey Bakulin and Maxim Dmitriev,  
EXPEC Advanced Research Center,  
Saudi Arabia*

#### 5:40-6:00 Parallel Multi-Scale Multi-Physics Time-Domain Solver for Seismic Wave Propagation

*Galina Reshetova, Institute of  
Computational Mathematics and  
Mathematical Geophysics, Russia;  
Vladimir A. Tcheverda, Vadim  
Lisitsa, and Viktor Kostin, Institute of  
Petroleum Geology & Geophysics of  
SB RAS, Russia*

*continued in next column*



Thursday, March 8

## MS52

### Disruptive Technologies for Future Science and Engineering - Part I of III

4:50 PM-6:30 PM

Room:52-204

#### For Part 2 see MS64

The recent end of Dennard scaling is necessitating diverse and innovative changes in emerging high-end computing systems, for attaining ever-increasing performance while simultaneously addressing energy efficiency constraints. In this minisymposium we discuss disruptive technologies for next generation computing systems. We begin by presenting solutions for transforming conventional architectural approaches for sustained exascale performance. Next we discuss the potential of quantum computing to deliver unprecedented computing capability for key application drivers. Finally we present neuromorphic computing technologies that hope to leverage brain-inspired systems to alleviate some of science's big data challenges.

Organizer: Rupak Biswas

NASA Ames Research Center, USA

Organizer: Jonathan Carter

Lawrence Berkeley National Laboratory, USA

Organizer: Leonid Oliker

Lawrence Berkeley National Laboratory, USA

#### 4:50-5:10 Open Hardware: How Open Source Designs Will Drive the Next Generation of HPC Systems

David Donofrio, Lawrence Berkeley National Laboratory, USA

#### 5:15-5:35 A Partitioned Approach for Highly Integrated Exascale Processors

Yasuko Eckert, AMD, USA

#### 5:40-6:00 Architectural Innovations for Applications of the Next Decades

Eng-Lim Goh, Hewlett Packard Corporation, USA

#### 6:05-6:25 Disruption in Memory

Bob Lucas, University of Southern California, USA; Richard Murphy, Micron, USA

Thursday, March 8

## MS53

### High Performance Communication with Lots of Threads and Tasks

4:50 PM-6:30 PM

Room:52-301

Today's high performance computers are increasingly relying on large intra-node parallelism to draw their performance. This creates a challenge of how threads efficiently communicate with each other across nodes. MPI + OpenMP is by far the dominant programming model for such environments. Many MPI implementations, however, do not perform very well when many threads concurrently invoke MPI functions. Many applications therefore need to manually aggregate communication of threads within a node. This puts a heavy burden on the programmer. Moreover, it creates a potential performance problem by introducing frequent barrier synchronizations within a node and thus losing an opportunity to overlap communication and computation. Researchers are working to advance this state of the art towards various directions. On the implementation side, they include efficient multithreaded MPIs and lower level communication substrates optimized for fine grain communication. On the programmability side, they include MPI + tasks and global address space models that unify intra-/inter-node communication. In this mini symposium, we like to gather researchers working in this area and discuss issues and solutions.

Organizer: Kenjiro Taura

University of Tokyo, Japan

#### 4:50-5:10 A Communication Layer for Fine Grain Communication and Task Parallelism

Kenjiro Taura and Wataru Endo, University of Tokyo, Japan

#### 5:15-5:35 Communication Throughput in HPC Applications

Costin Iancu, Lawrence Berkeley National Laboratory, USA

#### 5:40-6:00 Chapel: Task-Based Communication in a Productive Language

Elliot Ronaghan, Cray Inc., USA; Bradford L. Chamberlain, Cray, Inc., USA

#### 6:05-6:25 From OpenSHMEM to UCX - Co-design Programming Models and Communication Libraries with Network Interconnects

Pavel Shamis, ARM Research, USA

continued in next column

Thursday, March 8

## MS54

### Reliable Numerical Computations in HPC Environments- Part I of II

4:50 PM-6:30 PM

Room:52-302

#### For Part 2 see MS66

This minisymposium is devoted to reliable numerical computations to be performed in high-performance computing (HPC) environments. The main objective of the minisymposium is to bridge the gap between theories and algorithms in numerical analysis and practical applications in HPC. For this purpose, the speakers are assembled from a variety of research field in computational science. The theories and algorithms in numerical analysis include forward/backward stability, rounding error analysis, reproducible algorithms, high-precision computations, and verified numerical computations. The practical applications with HPC include computational material physics and electromagnetics.

Organizer: Takeshi Ogita

Tokyo Woman's Christian University,  
Japan

Organizer: Stef Graillat

University Pierre and Marie Curie  
(UPMC), France

Organizer: Takeo Hoshi

Tottori University, Japan

#### 4:50-5:10 Accurate and Verified Numerical Computations with HPC

Takeshi Ogita, Tokyo Woman's Christian  
University, Japan

#### 5:15-5:35 Iterative Refinement in Three Precisions for Fast and Accurate Solution of Ill-Conditioned Sparse Linear Systems

Nicholas J. Higham, University of  
Manchester, United Kingdom; Erin  
C. Carson, New York University,  
USA

#### 5:40-6:00 Faithfully Rounded Floating-Point Computations

Marko Lange, Waseda University,  
Japan; Siegfried M. Rump, Technische  
Universität, Hamburg-Harburg,  
Germany

#### 6:05-6:25 Open-Source Software Development That Bridges the Gap Between Numerical Analysis and Material Simulation on Supercomputers

Kazuyoshi Yoshimi, Mitsuaki Kawamura,  
and Youhei Yamaji, University of  
Tokyo, Japan; Tomohiro Sogabe,  
Nagoya University, Japan; Takeo  
Hoshi, Tottori University, Japan

Thursday, March 8

## MS55

### Parallel Numerical Linear Algebra for Future Extreme-Scale Systems - Part II of II

4:50 PM-6:30 PM

Room:52-303

#### For Part 1 see MS43

We describe recent developments in task-based algorithms for the solution of dense linear systems and the solution of both symmetric and unsymmetric dense eigenproblems. By using directed acyclic graphs (DAGs) we are able to avoid synchronization and obtain high levels of parallelism. Our algorithms also keep communication to a minimum, sometimes achieving a provably theoretical minimum. In the second minisymposium, we will consider the solution of sparse linear systems. For the direct solution of symmetrically structured systems, we express the computation as a DAG and use dense matrix kernels. We study both multifrontal and supernodal methods and show how numerical pivoting can be accommodated without losing much parallelism. We also discuss algorithms based on a Markowitz threshold approach for highly unsymmetric systems. For the iterative solvers, we study robust algebraic preconditioners that are highly scalable and examine a way to expand the Krylov space for better convergence while maintaining parallelism and reducing communication. We then consider hybrid techniques that combine the best features of direct and iterative methods. To aid in portability for a range of computing systems we use the runtime systems, StarPU and PaRSEC and we show the performance of prototype codes on a range of research and industrial applications. These minisymposia present research done in NLAFET, a Horizon 2020 FET-HPC project funded by the European Union.

Organizer: Iain Duff

Science & Technology Facilities Council,  
United Kingdom and CERFACS, Toulouse,  
France

*continued in next column*

*continued on next page*

Thursday, March 8

## MS55

### Parallel Numerical Linear Algebra for Future Extreme-Scale Systems - Part II of II

4:50 PM-6:30 PM

#### 4:50-5:10 Sparse Solvers on Extreme-Scale Systems

Iain Duff, Science & Technology Facilities Council, United Kingdom and CERFACS, Toulouse, France

#### 5:15-5:35 Task-Based Sparse Direct Solver for Symmetric Indefinite Systems

Florent Lopez, Rutherford Appleton Laboratory, United Kingdom; Iain Duff, Science & Technology Facilities Council, United Kingdom and CERFACS, Toulouse, France

#### 5:40-6:00 Robust Algebraic Preconditioners for Large Scale Applications

Laura Grigori, Simplicite Donfack, and Olivier Tissot, Inria, France

#### 6:05-6:25 Enlarged Conjugate Gradient Method for Reducing Communication

Olivier Tissot and Laura Grigori, Inria, France

Thursday, March 8

## MS56

### Tensor Decomposition for High Performance Data Analytics - Part III of III

4:50 PM-6:30 PM

Room:52-304

#### For Part 2 see MS20

With the explosion of Big Data, finding fast and scalable solutions for mining and analyzing large amounts of data is becoming increasingly important. In this regard, tensor decomposition has recently received much attention due to its ability to identify latent properties in high-order data. Unfortunately, the sparseness and the high-order nature of real-world data makes decomposing tensors slow and inefficient. This minisymposium will present research into new and improved methods for tensor decomposition for both shared-memory and distributed systems, as well as their application for analyzing real data.

Organizer: Jee Choi

IBM T.J. Watson Research Center, USA

Organizer: Keita Teranishi

Sandia National Laboratories, USA

Organizer: Richard Vuduc

Georgia Institute of Technology, USA

#### 4:50-5:10 Analysis of Performance and Portability of Sparse Tensor Decompositions on CPU/MIC/GPU Architectures

Christopher J. Forster, Keita Teranishi, Daniel M. Dunlavy, and Tamara G. Kolda, Sandia National Laboratories, USA

#### 5:15-5:35 Tensor Decompositions on Emerging Manycore Hardware with Ganten and Kokkos

Eric Phipps and Tamara G. Kolda, Sandia National Laboratories, USA

#### 5:40-6:00 Sparse Tensor Computations on Graphic Processing Units

Maryam M. Mehri Dehnavi and Bangtian Liu, Rutgers University, USA

#### 6:05-6:25 A Set of Building Blocks for Tensor Operations: Transposition, Summation, and Contraction

Paolo Bientinesi, RWTH-Aachen, Germany; Paul Springer, RWTH Aachen University, Germany

Thursday, March 8

## MS57

### Hierarchical Low Rank Approximation Methods - Part III of VI

4:50 PM-6:30 PM

Room:57-201

#### For Part 2 see MS45

#### For Part 4 see MS69

Hierarchical Low-Rank Approximation Methods such as H-matrices, Hierarchical Semi-Separable matrices, and Hierarchical Interpolative Factorization are increasing in popularity. These methods have near linear complexity and can be used as a fast approximate inverse of both dense matrices arising from integral equations and sparse matrices arising from partial differential equations. Accuracy is controllable by increasing the rank of the approximated blocks or by changing the subdivision criteria of the off-diagonal blocks. Therefore, these methods can be used as robust direct solvers with near linear complexity if high accuracy is chosen or they can be used as preconditioners if lower accuracy is used. With the advent of these methods, large dense linear algebra operations are scarcely needed anymore. In terms of parallelism, the hierarchical nature of these methods pose a challenge. Task-based approaches seem to be effective for shared memory parallelism, while load-balance becomes a problem on distributed memory. The recursive block structure also results in the computation of many small dense blocks. The recent development in batched dense linear algebra libraries facilitates the efficient computation of such small dense blocks. This minisymposium consists of cutting edge research on hierarchical low-rank approximation methods, their applications, parallelization, and extension.

*continued on next page*

Organizer: Rio Yokota  
*Tokyo Institute of Technology, Japan*

Organizer: Akihiro Ida  
*University of Tokyo, Japan*

Organizer: Chao Chen  
*Stanford University, USA*

Organizer: Eric F. Darve  
*Stanford University, USA*

**4:50-5:10 Dynamic Load Balancing for Construction and Arithmetic of Hierarchical Matrices**

*Tasuku Hiraishi, Kyoto University, Japan*

**5:15-5:35 Highly Optimized H<sup>2</sup>-Matrix Implementation on Multiple GPUs**

*George Turkiyyah, American University of Beirut, Lebanon; Wajih Boukaram, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; David E. Keyes, King Abdul Aziz University, Saudi Arabia*

**5:40-6:00 Hierarchical Matrix Preconditioner on a Distributed Memory Environment**

*Gustavo Chavez, Lawrence Berkeley National Laboratory, USA; George M Turkiyyah, American University of Beirut, Lebanon and King Abdullah University of Science & Technology (KAUST), Saudi Arabia; David E. Keyes, King Abdul Aziz University, Saudi Arabia*

**6:05-6:25 Distributed-Memory Block Low-Rank Factorization for Large-Scale Systems and Applications**

*Patrick Amestoy, ENSEEIHT-IRIT, France; Alfredo Buttari, CNRS, France; Jean-Yves L'Excellent, Inria-LIP-ENS Lyon, France; Theo Mary, Universite de Toulouse, France*

Thursday, March 8

**MS58**

**Scalable and Dynamic Graph Algorithms**

*4:50 PM-6:30 PM*

*Room:57-202*

Graph analysis provides tools for analyzing the irregular data sets common in health informatics, computational biology, climate science, sociology, security, finance, and many other fields. In recent years these graphs have exploded in size and are changing at increasing rates. This minisymposium will include talks on large scale graph analytics, with a focus on dynamic graph algorithms. It will also focus on graph frameworks for novel accelerators as well as real world applications.

Organizer: Oded Green  
*Georgia Institute of Technology, USA*

Organizer: David A. Bader  
*Georgia Institute of Technology, USA*

Organizer: Jason Riedy  
*Georgia Institute of Technology, USA*

**4:50-5:10 Title Not Available At Time Of Publication**

*Robert S. Schreiber, Hewlett-Packard Enterprise, USA*

**5:15-5:35 Massive-Scale Streaming Analytics**

*David A. Bader, Georgia Institute of Technology, USA*

**5:40-6:00 Dynamic Algorithms for Network Analysis at Scale**

*Eugenio Angriman, Karlsruhe Institute of Technology, Germany*

**6:05-6:25 An Adaptive Graph Processing Library on GPUs**

*Guangming Tan, Chinese Academy of Sciences, China*

Thursday, March 8

**CP7**

**Domain Decomposition**

*4:50 PM-6:30 PM*

*Room:52-101*

*Chair: Masha Sosonkina, Old Dominion University, USA*

**4:50-5:10 Comparison of Different Inexact BDDC Variants for Systems of PDEs**

*Martin Lanser and Axel Klawonn, Universitaet zu Koeln, Germany; Oliver Rheinbach, Technische Universität Bergakademie Freiberg, Germany*

**5:15-5:35 Robust Domain Decomposition Methods for High Performance Computing in Aerospace Industrial Context**

*Augustin Parret-Fréaud, Safran Tech, France; Christophe Bovet, ONERA, France; Pierre Gosselet, LMT-Cachan, France; Nicole Spillane, CMAP, Ecole Polytechnique, France*

**5:40-6:00 Multi-Level Domain Decomposition Solvers for Intrusive Spectral Stochastic Fem Versus Non-Intrusive Sparse Grid Based Solvers for Large Random Dimensions**

*Abhijit Sarkar and Ajit Desai, Carleton University, Canada; Mohammad Khalil, Sandia National Laboratories, USA; Chris Pettit, United States Naval Academy, USA; Dominique Poirel, Royal Military College, Canada*

**6:05-6:25 Simulation of Wind Flows in Large Urban Areas Using a Scalable Domain Decomposition Method**

*Zhengzheng Yan, Shenzhen Institute of Advanced Technology, China; Tangzheng Du and Rongliang Chen, Chinese Academy of Sciences, China; Xiao-Chuan Cai, University of Colorado Boulder, USA*

**Business Meeting**

*6:30 PM-7:15 PM*

*Room:63-201/202*





Thursday, March 8

## PP1

### Poster Session and Reception

7:15 PM-9:15 PM

Room:63-1F

#### Application of Game Theory to Correct Economics Decision

Saeed Seyed Agha Banihashemi, School of International Relations, Iran

#### A User-Oriented Parallel Wave Propagation Solver for Geophysical Applications

Regis Cottreau and Lucio De Abreu Corrêa, CNRS, CentraleSupélec, Université Paris-Saclay, France; José Camata and Alvaro Coutinho, COPPE/Universidade Federal do Rio de Janeiro, Brazil

#### A Parallel Implementation Technique of HOTRG for the 3D Cubic Lattice Ising Model

Haruka Yamada and Akira Imakura, University of Tsukuba, Japan; Toshiyuki Imamura, RIKEN, Japan; Tetsuya Sakurai, University of Tsukuba, Japan

#### The Schwarz Alternating Method for Concurrent Multiscale Coupling in Solid Mechanics

Irina K. Tezaur, Sandia National Laboratories, USA

#### Principal Component Analysis with Large-Scale Electronic State Calculation for Ultra-Flexible Device Materials

Takeo Hoshi, Hiroto Imachi, Yukiya Abe, and Kentaro Oohira, Tottori University, Japan; Koji Hukusima, University of Tokyo, Japan

#### A Proposal for a Nonlinear Semi-NMF Based Method with Bias Vectors and Regularization for Deep Neural Networks

Ryosuke Arai, Akira Imakura, and Tetsuya Sakurai, University of Tsukuba, Japan

#### Dataset Characterization of Data Mining Algorithms

Sayaka Akioka, Meiji University, Japan; Suzanne M. Shontz, University of Kansas, USA

#### Effect of Algebraic Block Multi-Color Ordering for Multi-Threaded ILU-GMRES Solver

Senxi Li, Takeshi Fukaya, and Takeshi Iwashita, Hokkaido University, Japan

#### A Fast and Efficient Preconditioning Method for Solving Ill-Conditioned Linear Systems by Partly Using LU Factors

Yuka Kobayashi and Takeshi Ogita, Tokyo Woman's Christian University, Japan

#### Open Distributive Interoperable Executive Library (opendiel) for Multi-Systems Simulation and Analysis

Kwai L. Wong, University of Tennessee and Oak Ridge National Laboratory, USA

#### Efficient Reduction of the Generalized HPD Eigenproblems

Valeriy Manin and Bruno Lang, Bergische Universität Wuppertal, Germany

#### Machine Learning Assisted Multiscale Modeling

Ansel Blumers, Brown University, USA

#### Locally Recursive Non-Locally Asynchronous Algorithms for FDTD Method on New GPGPU Cluster Architectures

Vadim Levchenko, Anastasia Perepelkina, and Anastasia Perepelkina, Keldysh Institute for Applied Mathematics, RAS, Russia; Andrey Zakirov, Kintech Lab Ltd., Russia; Yasunari Zempo, Hosei University, Japan

#### Parallel Symmetric Tridiagonal Eigensolver Based on Bisection and Inverse Iteration Algorithms with the Blocked Classical Gram-Schmidt Reorthogonalization and the Shifted CholeskyQR Decomposition

Kinji Kimura, Masayuki Osawa, and Yoshimasa Nakamura, Kyoto University, Japan

#### A Rational Function Algorithm of Blind Source Separation Based on ICA

Lei Du, Jintao Zhang, and Jinwei Xu, Dalian University of Technology, China

#### Adoption of Less Synchronous Modes of Computation in Nonlinear Solvers

Amani Alonazi and Lulu Liu, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; Hatem Ltaief and David E. Keyes, King Abdul Aziz University, Saudi Arabia

#### Using the Moment to Reduce Linear System Size

Sarah Huber, University of Wuppertal, Germany; Yasunori Futamura, University of Tsukuba, Japan; Martin Galgon, University of Wuppertal, Germany; Bruno Lang, Bergische Universität Wuppertal, Germany; Tetsuya Sakurai, University of Tsukuba, Japan

#### pFEM-CRAFT: A Library for Application-Level Fault-Resilience Based on the CRAFT Framework

Tatsuya Fukasawa, University of Tokyo, Japan; Faisal Shahzad, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany; Kengo Nakajima, University of Tokyo, Japan; Gerhard Wellein, University of Erlangen-Nürnberg, Germany

#### Avoiding Communication in Proximal Methods for Convex Optimization

Saeed Soori, Rachit Shah, and Maryam Dehnavi, Rutgers University, USA

#### Accelerating Simulations of Cerebrovascular Blood Flow Through Parallelization in Time

Daniel Ruprecht, University of Leeds, United Kingdom; Derek Groen, Brunel University, United Kingdom; Rupert Nash and David Scott, University of Edinburgh, United Kingdom

#### Toward Fault-Tolerant Parallel-in-Time Integration with Pfasst

Robert Speck, Jülich Supercomputing Centre, Germany; Daniel Ruprecht, University of Leeds, United Kingdom

#### Iterative Construction of Non-Symmetric Factored Sparse Approximate Preconditioners

Andrea Franceschini, Università di Padova, Italy; Carlo Janna and Massimiliano Ferronato, University of Padova, Italy

#### Fast Crustal Deformation Computation Method Using Openacc for Stochastic Inversion Analysis

Takuma Yamaguchi, Kohei Fujita, Tsuyoshi Ichimura, Muneo Hori, and Lalith Wijerathne, University of Tokyo, Japan

#### Efficient Robust Multi-Shift Triangular Solves



*Angelika Schwarz*, Mirko Myllykoski,  
Carl Christian Kjelgaard Mikkelsen,  
and Lars Karlsson, Umeå University,  
Sweden

#### **Large-Scale Dynamic Graph Processing on HPC Systems**

*Keita Iwabuchi* and Roger Pearce,  
Lawrence Livermore National  
Laboratory, USA

#### **Geodynamo Simulations and Visualizations Using Calypso**

*Hiroaki Matsui*, University of California,  
Davis, USA

#### **A Performance Spectrum for Parallel Computational Frameworks that Solve PDEs**

*Justin Chang*, Rice University, USA;  
Kalyana Nakshatrala, University of  
Houston, USA; Matthew G. Knepley,  
University of Buffalo, USA; Lennart  
Johnsson, University of Houston,  
USA

#### **Dynamical Interaction Between Earth's Core and Mantle with Massively Parallel Computing**

*Takashi Nakagawa*, Japan Agency  
for Marine-Earth Science and  
Technology, Japan

#### **Pseudo-Assembly Programming for Batched Matrix Factorization**

*Yaohung M. Tsai*, Piotr Luszczek,  
Jakub Kurzak, and Jack Dongarra,  
University of Tennessee, USA

#### **Fast Anomaly Diagnosis in HPC Systems Using Abnormal Subspace Sparse PCA Models**

*Wenming Jiang* and Ying Zhao, Tsinghua  
University, P. R. China

#### **Recent Progress on Low-Order Unstructured Implicit Finite-Element Earthquake Simulation on Oakforest-Pacs**

*Kohei Fujita* and Tsuyoshi Ichimura,  
University of Tokyo, Japan; Masashi  
Horikoshi and Larry Meadows,  
Intel Corporation, USA; Kengo  
Nakajima, Muneo Hori, and Lalith  
Maddegadara, University of Tokyo,  
Japan

#### **Dynamic Load Balancing in Electric Propulsion Plume Simulations**

*Samuel Araki*, Air Force Research  
Laboratory, USA

#### **Contour Integral-Based Verified Computing for Partial Eigenvalues**

*Akitoshi Takayasu*, Akira Imakura, and  
Keiichi Morikuni, University of  
Tsukuba, Japan

#### **Development of Multi-Scale, Multi-Physics Methods for Multi-Fluid Plasma Simulations with Application to In-Space Propulsion**

*Eder M. Sousa*, ERC Inc. and Air  
Force Research Laboratory, USA;  
Robert Martin, Air Force Research  
Laboratory, USA

#### **Task-Parallel Factorizations of Hierarchical Matrices Using OmpSs and Openmp**

*Rocío Carratalá-Sáez* and José I. Aliaga,  
Universitat Jaume I, Spain; Enrique  
S. Quintana-Ortí, Universidad Jaume  
I, Spain

#### **Dynamic Task Scheduling Implementation of Tile QR Decomposition on CPU/GPU Heterogeneous Cluster System**

*Masatoshi Takayanagi* and Tomohiro  
Suzuki, University of Yamanashi,  
Japan

#### **Research of Finding Formula That Express Data Without Any Background Information**

*Issei Koga* and Kenji Ono, Kyushu  
University, Japan

#### **Performance Evaluation of Hierarchical Matrix Computation on Various Modern Architectures**

*Satoshi Ohshima*, University of Tokyo,  
Japan; Ichitaro Yamazaki, University  
of Tennessee, Knoxville, USA;  
Akihiro Ida, University of Tokyo,  
Japan; Rio Yokota, Tokyo Institute of  
Technology, Japan

#### **Sparse Iterative Solvers Review on Large Seismic Tomography**

*Lionel Boillot*, Total E&P, France

#### **Batched Factorization and Inversion for Iterative Solvers**

*Goran Flegar*, Universitat Jaume I,  
Spain; Hartwig Anzt, University of  
Tennessee, USA; Jack J. Dongarra,  
University of Tennessee, ORNL,

and University of Manchester,  
USA; Enrique S. Quintana-Ortí,  
Universidad Jaume I, Spain

#### **Consideration on the Optimal Balance Between Space Parallelism and Time Parallelism**

*Ryo Yoda*, Akihiro Fujii, and Teruo  
Tanaka, Kogakuin University, Japan;  
Kengo Nakajima, University of  
Tokyo, Japan

#### **Efficient Parallel Implementation of Fill-Reducing Ordering Using Spectral Nested Dissection for Large-Scale Sparse Linear Systems**

*Yuta Inagawa*, Yasunori Futamura,  
and Tetsuya Sakurai, University of  
Tsukuba, Japan

#### **Performance Evaluation of Large-Scale Deep Learning Framework ChainerMN on Oakforest-PACS**

*Kohei Tamura* and Toshihiro Hanawa,  
University of Tokyo, Japan

#### **Convergence Property and Accuracy Improvement of Block Bicgstab Class Solvers for Linear Systems with Many Right-Hand Sides**

*Hirotto Tadano*, University of Tsukuba,  
Japan

#### **Parallel Implementation of Hall-effect Thruster Current-driven Instabilities Simulations**

*Jonathan Tran*, University of California,  
Los Angeles, USA

#### **Large-Scale Estimation of Sparse Diagonally Dominant Inverse Covariance Matrices**

*Aryan Eftekhari*, Università della Svizzera  
italiana, Switzerland; Olaf Schenk,  
Università della Svizzera italiana,  
Switzerland

#### **Numerical Simulation of Density-Driven Groundwater Flows with a Free Surface in Complicated Geometries**

*Dmitry Logashenko*, King Abdullah  
University of Science & Technology  
(KAUST), Saudi Arabia; Sebastian  
Reiter, Goethe University Frankfurt,  
Germany; Gabriel Wittum, King  
Abdullah University of Science &  
Technology (KAUST), Saudi Arabia

## Friday, March 9

### Registration

8:00 AM-5:00 PM

Room:63-1F

Friday, March 9

### IP4

#### Enabling Multi-Peta-Scale Fully Implicit Simulations of Atmospheric Dynamics on Sunway TaihuLight

8:30 AM-9:15 AM

Room:63-201/202

Chair: Jeffrey S. Vetter, Oak Ridge National Laboratory, USA

This joint talk is comprised of two parts. In the first part, we will introduce the technical details and the underlying design philosophy of Sunway TaihuLight, which is the world's first system with a peak performance greater than 100 PFlops, and a parallel scale of over 10 million cores. In contrast with other existing heterogeneous supercomputers, which include both CPU processors and PCIe-connected many-core accelerators (NVIDIA GPU or Intel MIC), the computing power of TaihuLight is provided by a homegrown many-core SW26010 CPU that includes two types of processing elements in one chip. With a comparison between other leadership systems, we would highlight the major programming challenges for utilizing the system for extreme-scale scientific simulations. In the second part of the talk, we will present some technical details of our recent work that won the ACM Gordon Bell Prize in 2016. In the work, we have designed an ultra-scalable fully implicit solver for solving the atmospheric dynamics at the nonhydrostatic scale. The fully implicit solver can scale to the entire system of Sunway TaihuLight and sustain an aggregate performance of 8 PFlops. Remarkably, the simulation capability of the fully implicit solver at the full system scale is nearly two orders of magnitude better than a highly optimized explicit solver with an aggregate performance of 23.66 PFlops.

Haohuan Fu

National Supercomputing Center, Wuxi, China and Tsinghua University, China

Chao Yang

Chinese Academy of Sciences, China

Friday, March 9

### Intermission

9:15 AM-9:25 AM

### MS59

#### High Level High Performance Programming Abstractions for CSE Applications

9:25 AM-11:05 AM

Room:52-102

Parallel programming is hard, and high performance parallel programming is even harder, especially when it comes to CSE domain experts. Therefore, the high performance computing community has long been devoting significant efforts on high level programming abstractions, with the hope of both easing programming and improving performance. This minisymposium will discuss recent progresses on this topic, including operator-alike DSLs, pattern-driven frameworks, and domain specific libraries. It brings together researchers worldwide to share experiences and exchange ideas on designing, architecting, and optimizing these abstractions.

Organizer: Ulrich J. Ruede

University of Erlangen-Nuernberg, Germany

Organizer: Xiaowen Xu

Institute of Applied Physics and Computational Mathematics, China

Organizer: Zhang Yang

Institute of Applied Physics and Computational Mathematics, China

#### 9:25-9:45 A Pattern-Driven Implicit Parallel Computing Abstraction for Mesh-Based CSE Applications

Zeyao Mo, CAEP Software Center for High Performance Numerical

Simulations, China; Zhang Yang and Xu Liu, Institute of Applied Physics and Computational Mathematics, China

**9:50-10:10 Sustainability and Efficiency for Simulation Software in the Exascale Era***Ulrich J. Ruede, University of Erlangen-Nuernberg, Germany***10:15-10:35 Wasatch: A Use Case of High Level Abstractions for Computational Multiphysics***Tony Saad and James C. Sutherland, University of Utah, USA***10:40-11:00 From Serial to Parallel: Visual Programming for High Performance Numerical Simulation Development***Li Liao, Institute of Applied Physics and Computational Mathematics, China*

Friday, March 9

**MS60****Tools for Implicitly Solving Large-Scale Nonsymmetric Systems***9:25 AM-11:05 AM**Room:52-103*

Nonsymmetric systems are often encountered in computational sciences and engineering and require proper techniques to be solved efficiently. The lack of theory for preconditioned iterative methods with nonsymmetric systems should not limit the scalability of the applications. We will discuss in this minisymposium about efficient strategies for discretizing, preconditioning and solving nonsymmetric systems.

Organizer: Pierre Jolivet  
*CNRS, France*

Organizer: Atsushi Suzuki  
*Osaka University, Japan*

**9:25-9:45 A Parallel Direct Factorization with Symmetric Pivoting for Nonsymmetric and Indefinite Matrices in a Semi-Conductor Problem***Atsushi Suzuki, Osaka University, Japan***9:50-10:10 A Parallelized Fill-in Controlled ILU Preconditioner for Multiphysics Fem Applications***Takumi Washio, University of Tokyo, Japan***10:15-10:35 Domain Decomposition Approach for the Iterative Solution of Strongly Coupled Vibro-Acoustic Problems***François-Xavier Roux, ONERA, France***10:40-11:00 High Performance Computation of Radiative Transfer Equation Using the Finite Element Method***Pierre Jolivet, CNRS, France*

Friday, March 9

**MS61****Theory Meets Practice for High Performance Computing - Part I of II***9:25 AM-11:05 AM**Room:52-104***For Part 2 see MS73**

Algorithms are often evaluated in terms of the number of arithmetic operations they performed. However, on today's machines, communication, i.e., moving data through memory hierarchies and among processors, often requires much more time (and energy) than performing computations. Hardware trends suggest that the relative costs of such communication will only increase. In this minisymposium we will review several recent theoretical advances for reducing both arithmetic and communication costs, fostering better performance.

Organizer: Oded Schwartz  
*Hebrew University of Jerusalem, Israel*

Organizer: Sivan A. Toledo  
*Tel Aviv University, Israel*

**9:25-9:45 A Bridging Model for High Performance Cloud Computing***Bill McColl, Huawei Technologies, France***9:50-10:10 Communication-Optimal Loop Nests***Nicholas Knight, New York University, USA***10:15-10:35 S-Step Methods in Machine Learning**

*Aditya Devarakonda, Kimon Fountoulakis, James W. Demmel, and Michael Mahoney, University of California, Berkeley, USA*

**10:40-11:00 A Geometric Partitioning Method for Distributed Tomographic Reconstruction**

*Jan-Willem Buurlage and Rob H. Bisseling, Utrecht University, The Netherlands; Joost Batenburg, Centrum voor Wiskunde en Informatica (CWI), Netherlands*

Friday, March 9

## MS62

### Emerging Programming Models for Extreme-Scale Computing - Part I of II

9:25 AM-11:05 AM

Room:52-201

#### For Part 2 see MS74

For rather long years, Message Passing Interface (MPI) has been widely used to program parallel supercomputers. However, now that supercomputer systems become larger and larger and more and more complicated as well as application programs, the degradation of productivity due to the programming in the style of "MPI+X" is one of the biggest concern in this area. To resolve this problem, various programming models have been proposed that could provide programmers with the capabilities of higher-level abstraction, higher portability, higher parallelism, etc. for better performance and productivity of applications. In this minisymposium, we have eight speakers from all over the world. They will talk about a variety of topics about programming models for extreme-scale computing, ranging from the one actually exploited for the world's fastest supercomputer Sunway TaihuLight, to many ongoing research projects each of which focuses on a particular target such as accelerators, task parallelism, PGAS, workflow, and hybrid parallelism.

Organizer: Hitoshi Murai

RIKEN, Japan

#### 9:25-9:45 Trends and Challenges on Programming Models for Extreme-Scale Computing

Mitsuhsa Sato, RIKEN Advanced Institute for Computational Science, Japan

#### 9:50-10:10 UPC++: a PGAS Library for Exascale Computing

Scott B. Baden, Lawrence Berkeley National Laboratory, USA

#### 10:15-10:35 Opportunities for Data-Centric Programming in the PGAS Model: Experience with DASH

Karl Fuerlinger, Ludwig-Maximilians-Universität München, Germany

#### 10:40-11:00 The OmpSs Programming Model and Its Runtime Support

Jesus Labarta, Barcelona Supercomputing Center, Spain

Friday, March 9

## MS63

### Parallel Data Processing in Geophysics - Part II of II

9:25 AM-10:40 AM

Room:52-202

#### For Part 1 see MS51

Modern approaches to the processing and interpretation of seismic data for the reconstruction of 3D geological structures are inconceivable without the use of high-performance computing systems with parallel architecture. One of the key features here is the strong data dependence, which can appear at various stages of processing and interpretation. The simplest situation appears at the level of the model-data upload, when the input model-data should be distributed independently between parallel processes, which needs to the data exchange during the computations. This leads to the necessity of data flow optimization to reduce computational time and RAM requirements. Thus, implementation of such algorithms requires use of the modern efficient approaches to parallel data processing. The objective of this minisymposium is to discuss new advances in parallel seismic data processing and seismic modelling.

Organizer: Vadim Lisitsa

Institute of Petroleum Geology & Geophysics of SB RAS, Russia

Organizer: Galina Reshetova

Institute of Computational Mathematics and Mathematical Geophysics, Russia

#### 9:25-9:45 High-Performance DTgeo4 Software for 3D Wave Modeling of Seismic Fields

Anastasia Perepelkina, and Vadim Levchenko, Keldysh Institute for Applied Mathematics, RAS, Russia; Andrey Zakirov, Kintech Lab Ltd., Russia; Anton Ivanov, Keldysh Institute for Applied Mathematics, RAS, Russia; Tatiana Levchenko and Vladimir Rok, VNIGNI, Russia

#### 9:50-10:10 Highly-Scalable Algorithm for Numerical Simulation of Sonic Logs

Galina Reshetova, Institute of Computational Mathematics and Mathematical Geophysics, Russia; Vladimir A. Tcheverda and Vadim Lisitsa, Institute of Petroleum Geology & Geophysics of SB RAS, Russia

#### 10:15-10:35 Semi-Analytical Highly-Scalable Preconditioner for Helmholtz Equation

Mikhail Belonosov, Saudi Aramco Oil Company, Saudi Arabia

continued in next column



Friday, March 9

**MS64****Disruptive Technologies for Future Science and Engineering - Part II of III**

9:25 AM-11:05 AM

*Room: 52-204***For Part 1 see MS52****For Part 3 see MS76**

The recent end of Dennard scaling is necessitating diverse and innovative changes in emerging high-end computing systems, for attaining ever-increasing performance while simultaneously addressing energy efficiency constraints. In this minisymposium we discuss disruptive technologies for next generation computing systems. We begin by presenting solutions for transforming conventional architectural approaches for sustained exascale performance. Next we discuss the potential of quantum computing to deliver unprecedented computing capability for key application drivers. Finally we present neuromorphic computing technologies that hope to leverage brain-inspired systems to alleviate some of science's big data challenges.

**Organizer: Rupak Biswas***NASA Ames Research Center, USA***Organizer: Jonathan Carter***Lawrence Berkeley National Laboratory, USA***Organizer: Leonid Oliker***Lawrence Berkeley National Laboratory, USA***9:25-9:45 Quantum Computing by Quantum Annealing***Hidetoshi Nishimori, Tokyo Institute of Technology, Japan***9:50-10:10 Quantum Machine Learning***Rupak Biswas, NASA Ames Research Center, USA***10:15-10:35 Quantum Supremacy in Near-Term Devices***Sergio Boixo, Google, Inc., USA***10:40-11:00 Computing Molecular Energies on a Quantum Processor***Irfan Siddiqi, University of California, Berkeley, USA; Jonathan Carter, Lawrence Berkeley National Laboratory, USA*

Friday, March 9

**MS65****Hazard and Disaster Simulation of Earthquake and Tsunami with HPC- Part I of II**

9:25 AM-11:05 AM

*Room: 52-301***For Part 2 see MS77**

The scope of the minisymposium is basically to discuss developments and directions for large scale earthquake and tsunami simulation with high performance computing technology. Broadband aspects from seismology to earthquake engineering are expected with a special emphasis on use of super computers (e.g. earthquake generation, seismic wave propagation, crustal deformation, structural response, soil amplification, city and social response, evacuation, recovery, fluid-structure coupling, global tsunami propagation, local tsunami run-up and the other related issues in earthquake and tsunami).

**Organizer: Takane Hori***Japan Agency for Marine-Earth Science and Technology, Japan***Organizer: Takamasa Iyo***Kobe University, Japan***9:25-9:45 A Proposal of Monitoring and Forecasting Method for Crustal Activity in and Around Japan with 3-Dimensional Heterogeneous Medium Using a Large-Scale High-Fidelity Finite Element Simulation***Takane Hori, Japan Agency for Marine-Earth Science and Technology, Japan; Tsuyoshi Ichimura and Kohei Fujita, University of Tokyo, Japan; Ryoichiro Agata, Japan Agency for Marine-Earth Science and Technology, Japan; Takuma Yamaguchi, University of Tokyo, Japan***9:50-10:10 Towards Exascale Seismology: Challenges and Opportunities***Jeroen Tromp, Princeton University, USA***10:15-10:35 Simulation Software for the Simulation of the Sumatra 2004 Earthquake and Tsunami***Carsten Uphoff, Technical University of Munich, Germany; Leonhard Rannabauer and Michael Bader, Technische Universität München, Germany; Alice A. Gabriel, Ludwig-Maximilians-Universität München, Germany***10:40-11:00 Scalable Solvers and Preconditioners for Nonlinear Stokes Equations with Yielding***Johann Rudi, University of Texas at Austin, USA; Georg Stadler, Courant Institute of Mathematical Sciences, New York University, USA; Omar Ghattas, University of Texas at Austin, USA*



Friday, March 9

## MS66

### Reliable Numerical Computations in HPC Environments- Part II of II

9:25 AM-11:05 AM

Room:52-302

#### For Part 1 see MS54

This minisymposium is devoted to reliable numerical computations to be performed in high-performance computing (HPC) environments. The main objective of the minisymposium is to bridge the gap between theories and algorithms in numerical analysis and practical applications in HPC. For this purpose, the speakers are assembled from a variety of research field in computational science. The theories and algorithms in numerical analysis include forward/backward stability, rounding error analysis, reproducible algorithms, high-precision computations, and verified numerical computations. The practical applications with HPC include computational material physics and electromagnetics.

Organizer: Stef Graillat

University Pierre and Marie Curie (UPMC), France

Organizer: Takeshi Ogita

Tokyo Woman's Christian University, Japan

Organizer: Takeo Hoshi

Tottori University, Japan

#### 9:25-9:45 Parallel Accurate Algorithms for Polynomial and Rational Function Evaluation

Stef Graillat, University Pierre and Marie Curie (UPMC), France

#### 9:50-10:10 A Reproducible Solution of Linear Systems

Roman Iakymchuk, KTH Royal Institute of Technology, Sweden

#### 10:15-10:35 Verified Numerical Computations for Large-Scale Linear Systems

Katsuhisa Ozaki, Shibaura Institute of Technology, Japan; Takeshi Ogita, Tokyo Woman's Christian University, Japan

#### 10:40-11:00 Eddy Current Analysis by Iterative Domain Decomposition Method Using Pseudo-Quadruple Precision

Amane Takei, University of Miyazaki, Japan

Friday, March 9

## MS67

### Approaches Towards Exascale Computational Fluid Dynamics

9:25 AM-11:05 AM

Room:52-303

The complex nature of turbulent fluid flows implies that the computational resources needed to accurately model problems of industrial and academic relevance is virtually unbounded. Computational Fluid Dynamics (CFD) is therefore a natural driver for exascale computing and has the potential for substantial societal impact, like reduced energy consumption, alternative sources of energy, improved health care, and improved climate models. Extreme-scale CFD possesses several cross disciplinary challenges e.g. algorithmic issues in scalable solver design, handling of extreme sized data with compression and in-situ analysis, resilience and energy awareness in both hardware and algorithm design. The wide range of topics makes exascale CFD relevant to a great HPC audience, extending outside the traditional fluid dynamics community. This proposed minisymposium will be organized by the EU funded Horizon 2020 project ExaFLOW together with RIKEN AICS, and will feature presentations showcasing their work on addressing key algorithmic challenges in CFD in order to facilitate simulations at exascale, e.g. accurate and scalable solvers, data reduction methods and strategies for accuracy and error control. This session aims at bringing together the CFD community as a whole, from HPC experts to domain scientists, discussing current and future challenges towards exascale fluid dynamics simulations and facilitating international collaboration.

Organizer: Niclas Jansson

KTH Royal Institute of Technology, Sweden

Organizer: Niclas Jansson

KTH Royal Institute of Technology, Sweden

Organizer: Makoto Tsubokura

Kobe University, Japan

#### 9:25-9:45 Efficient Gather Scatter Operations in Nek5000 Using PGAS

Niclas Jansson, KTH Royal Institute of Technology, Sweden

#### 9:50-10:10 A Parallel Algorithm of Pre and Post Processing for Industrial CFD Applications

Keiji Onishi, RIKEN Advanced Institute for Computational Science, Japan; Niclas Jansson and Niclas Jansson, KTH Royal Institute of Technology, Sweden; Rahul Bale, RIKEN Advanced Institute for Computational Science, Japan; Makoto Tsubokura, Kobe University, Japan

#### 10:15-10:35 Adaptive Meshes and Error Control in Nek5000

Philipp Schlatter, KTH Stockholm, Sweden

#### 10:40-11:00 A Stencil Penalty Approach For Improving Accuracy of Constraint Immersed Boundary Method

Rahul Bale, RIKEN Advanced Institute for Computational Science, Japan; Niclas Jansson and Niclas Jansson, KTH Royal Institute of Technology, Sweden; Keiji Onishi, RIKEN Advanced Institute for Computational Science, Japan; Makoto Tsubokura, Kobe University, Japan; Neelesh Patankar, Northwestern University, USA

continued in next column

Friday, March 9

## MS68

### Challenges in Parallel Adaptive Mesh Refinement - Part I of III

9:25 AM-11:05 AM

Room:52-304

#### For Part 2 see MS80

Parallel adaptive mesh refinement (AMR) is a key technique when simulations are required to capture time-dependent and/or multiscale features. Frequent re-adaptation and repartitioning of the mesh during the simulation can impose significant overhead, particularly in large-scale parallel environments. Further challenges arise due to the availability of accelerated or special-purpose hardware, and the trend toward hierarchical and hybrid compute architectures. Our minisymposium addresses algorithms, scalability, and software issues of parallel AMR.

Organizer: Carsten Burstedde  
*Universität Bonn, Germany*

Organizer: Michael Bader  
*Technische Universität München, Germany*

Organizer: Martin Berzins  
*University of Utah, USA*

#### 9:25-9:45 Nonlocal Algorithms on Adaptive Meshes

Carsten Burstedde, Universität Bonn, Germany

#### 9:50-10:10 Scalable Algorithms for Tree-Based AMR with General Element Types

Carsten Burstedde and Johannes Holke, Universität Bonn, Germany

#### 10:15-10:35 Parallel Unstructured Mesh (Re)partitioning Based on Sfc

Ricard Borrell, Guillaume Houzeaux, and Juan Carlos Cajas, Barcelona Supercomputing Center, Spain

#### 10:40-11:00 Structured and Unstructured Adaptivity in PETSc

Tobin Isaac, Georgia Institute of Technology, USA; Matthew G. Knepley, University of Buffalo, USA

Friday, March 9

## MS69

### Hierarchical Low Rank Approximation Methods - Part IV of VI

9:25 AM-11:05 AM

Room:57-201

#### For Part 3 see MS57

#### For Part 5 see MS81

Hierarchical Low-Rank Approximation Methods such as H-matrices, Hierarchical Semi-Seperable matrices, and Hierarchical Interpolative Factorization are increasing in popularity. These methods have near linear complexity and can be used as a fast approximate inverse of both dense matrices arising from integral equations and sparse matrices arising from partial differential equations. Accuracy is controllable by increasing the rank of the approximated blocks or by changing the subdivision criteria of the off-diagonal blocks. Therefore, these methods can be used as robust direct solvers with near linear complexity if high accuracy is chosen or they can be used as preconditioners if lower accuracy is used. With the advent of these methods, large dense linear algebra operations are scarcely needed anymore. In terms of parallelism, the hierarchical nature of these methods pose a challenge. Task-based approaches seem to be effective for shared memory parallelism, while load-balance becomes a problem on distributed memory. The recursive block structure also results in the computation of many small dense blocks. The recent development in batched dense linear algebra libraries facilitates the efficient computation of such small dense blocks. This minisymposium consists of cutting edge research on hierarchical low-rank approximation methods, their applications, parallelization, and extension.

Organizer: Rio Yokota  
*Tokyo Institute of Technology, Japan*

Organizer: Akihiro Ida  
*University of Tokyo, Japan*

Organizer: Chao Chen  
*Stanford University, USA*

Organizer: Eric F. Darve  
*Stanford University, USA*

#### 9:25-9:45 Fast Numerical Methods for Space-Time Fractional Partial Differential Equations

Hong Wang, University of South Carolina, USA

#### 9:50-10:10 Fast Iterative Solvers for Fractional Partial Differential Equations

Michael K. Ng and Lin Xuelei, Hong Kong Baptist University, Hong Kong

#### 10:15-10:35 Application of Hierarchical Matrices to Adaptive Finite Element Method for the Multi-Term Fractional Differential Equations

Xuan Zhao, Southeast University, China

#### 10:40-11:00 O(N) Method for Spatiotemporal Boundary Integral Equation Method

Daisuke Sato, University of Tokyo, Japan

Friday, March 9

## MS70

### Parallel Programming Models, Algorithms and Frameworks for Extreme Computing - Part I of III

9:25 AM-11:05 AM

Room:57-202

#### For Part 2 see MS82

Many-core and GPU processors are the foundation for almost all leadership computing platforms. Application and library software developers may often effectively use these processors and some general approaches have emerged. It is widely recognized that careful design of software and data structures, with effective memory management, are the most critical challenges to obtaining scalable optimized performance on those systems. Furthermore, investment in formal software techniques for robust environments is needed for advancing simulation capabilities. In these minisymposia we discuss current experiences and development of applications, libraries and frameworks using a variety of hardware. Speakers will address performance results and software design.

Organizer: Michael Heroux

*Sandia National Laboratories, USA*

Organizer: Kengo Nakajima

*University of Tokyo, Japan*

Organizer: Serge G. Petiton

*Université Lille 1 and CNRS, France*

#### 9:25-9:45 An Overview of Progress and Challenges for Scalable Manycore and Accelerator Systems

*Michael Heroux, Sandia National Laboratories, USA*

#### 9:50-10:10 Role of Framework Development in Multiphysics High Performance Computing Software

*Anshu Dubey and Lois Curfman McInnes, Argonne National Laboratory, USA*

#### 10:15-10:35 Towards Performance Portable Assembly Tools for Multi-Fluid Plasma Simulations

*Roger Pawlowski, Eric C. Cyr, Sean Miller, Matthew Bettencourt, Eric Phipps, Christian Trott, and John Shadid, Sandia National Laboratories, USA*

#### 10:40-11:00 Preparing Sundials for Exascale Computing Platforms

*Carol S. Woodward, Lawrence Livermore National Laboratory, USA; Daniel R. Reynolds, Southern Methodist University, USA; David J. Gardner, Slaven Peles, and John Loffeld, Lawrence Livermore National Laboratory, USA*

Friday, March 9

## CP8

### Performance Tuning

9:25 AM-11:05 AM

Room:52-101

*Chair: Michaela Taufer, University of Delaware, USA*

#### 9:25-9:45 Parallelization of Recursive Inverse Factorization using the Chunks and Tasks Programming Model

*Anton Artemov, Elias Rudberg, and Emanuel H. Rubensson, Uppsala University, Sweden*

#### 9:50-10:10 A Distributed Shared Memory Library with Global-View Tasks on High-Performance Interconnects

*Wataru Endo and Kenjiro Taura, University of Tokyo, Japan*

#### 10:15-10:35 Efficient Multiple Precision Dense Matrix Multiplication Library with Automatic Turning Tool

*Tomonori Kouya, Shizuoka Institute of Science and Technology, Japan*

#### 10:40-11:00 Block Red-Black Milu(0) Preconditioner with Relaxation on GPU

*Akemi Shioya and Yusaku Yamamoto, University of Electro-Communications, Japan*

### Coffee Break

11:05 AM-11:35 AM



Room:63-1F

Friday, March 9

**MS71****Applied Graph Theory in Interconnection Network Design and Operation**

11:35 AM-1:15 PM

*Room:52-102*

Large-scale interconnection networks are omnipresent in high performance computing systems and data centers, and are on the rise in network-on-chip architectures. Optimizing the topology and management of these interconnects, as well as efficiently utilizing the right technology, are the key challenges, especially while balancing multiple and partially conflicting requirements, such as minimizing hardware costs vs. achievable global throughput or topological structure vs. routability of the network. In fact, graph analysis and graph algorithms play a major role in all of these aspects of interconnect design and management. This minisymposium addresses the state-of-the-art approaches to reduce the hardware costs and latency via low-diameter topologies, and deepens the understanding of packet-based routing and circuit-based switching of these networks.

Organizer: Jens Domke

*Tokyo Institute of Technology, Japan***11:35-11:55 Optimizing TSUBAME2.5 Network with Sub-Optimal Infrastructure***Akihiro Nomura, Tokyo Institute of Technology, Japan***12:00-12:20 Cable-Geometric Error-Prone Approach for Low-Latency Interconnection Networks***Nguyen Truong, The Graduate University for Advanced Studies, Sokendai, Japan***12:25-12:45 Routing on the Channel Dependency Graph***Jens Domke, Tokyo Institute of Technology, Japan***12:50-1:10 Circuit-Switched Interconnects Using Limited Number of Slots***Yao Hu, National Institute of Informatics, Japan*

Friday, March 9

**MS72****Performance Portability of Scientific HPC Applications**

11:35 AM-1:15 PM

*Room:52-103*

Modernizing scientific codes to run performantly on multiple emerging architectures is a task of great complexity and challenge for computational scientists today. Typically the set of performance necessary optimizations for a code on a given platform will be distinct from those for another. To ameliorate this difficulty, we turn to abstraction layers, mapping software, compiler tools, and alternative programming models. This minisymposium includes studies of applications using performance tools running on multiple architectures. We will present how these tools ease the cross platform application development challenge and the performance tradeoffs of using them.

Organizer: Geoff Womeldorff

*Los Alamos National Laboratory, USA*

Organizer: Ian Karlin

*Lawrence Livermore National Laboratory, USA***11:35-11:55 The Flexible Computational Science Infrastructure (FLECSI) Programming System: Backend Interfaces and Applications***Marc Charest, Los Alamos National Laboratory, USA***12:00-12:20 Enhancing Runtime and Applications with Non-Blocking Collective Operations***Julien Jaeger, CEA, France***12:25-12:45 Integrating Existing Fortran Codes with Modern Runtime Systems***Joshua Payne, Los Alamos National Laboratory, USA***12:50-1:10 Title Not Available At Time Of Publication***Rob Hoekstra, Sandia National Laboratories, USA*

Friday, March 9

**MS73****Theory Meets Practice for High Performance Computing - Part II of II**

11:35 AM-1:15 PM

*Room:52-104***For Part 1 see MS61**

Algorithms are often evaluated in terms of the number of arithmetic operations they performed. However, on today's machines, communication, i.e., moving data through memory hierarchies and among processors, often requires much more time (and energy) than performing computations. Hardware trends suggest that the relative costs of such communication will only increase. In this minisymposium we will review several recent theoretical advances for reducing both arithmetic and communication costs, fostering better performance.

Organizer: Oded Schwartz

*Hebrew University of Jerusalem, Israel*

Organizer: Sivan A. Toledo

*Tel Aviv University, Israel***11:35-11:55 Matrix Multiplication, A Little Faster***Elaye E. Karstadt and Oded Schwartz, Hebrew University of Jerusalem, Israel***12:00-12:20 Parallel 3D CholeskyQR2 for Rectangular Matrices***Edward Hutter, University of Illinois, Urbana-Champaign, USA***12:25-12:45 Faster Parallel Tensor Compression using Randomization***Casey Battaglin, Georgia Institute of Technology, USA***12:50-1:10 Title Not Available At Time Of Publication***Gianfranco Bilardi, University of Padova, Italy*



Friday, March 9

## MS74

### Emerging Programming Models for Extreme-Scale Computing - Part II of II

11:35 AM-1:15 PM

Room:52-201

#### For Part 1 see MS62

For rather long years, Message Passing Interface (MPI) has been widely used to program parallel supercomputers. However, now that supercomputer systems become larger and larger and more and more complicated as well as application programs, the degradation of productivity due to the programming in the style of “MPI+X” is one of the biggest concern in this area. To resolve this problem, various programming models have been proposed that could provide programmers with the capabilities of higher-level abstraction, higher portability, higher parallelism, etc. for better performance and productivity of applications. In this minisymposium, we have eight speakers from all over the world. They will talk about a variety of topics about programming models for extreme-scale computing, ranging from the one actually exploited for the world’s fastest supercomputer Sunway TaihuLight, to many ongoing research projects each of which focuses on a particular target such as accelerators, task parallelism, PGAS, workflow, and hybrid parallelism.

Organizer: Hitoshi Murai

RIKEN, Japan

#### 11:35-11:55 Graph of Component Languages for Extreme Computational and Data Science; the YML Example

Serge Petiton, University of Lille, France;  
Nahid Emad, University of Versailles, France; Laurent Bobelin, Université François Rabelais, France

#### 12:00-12:20 Xcalableacc: Highly Productive Accelerated Programming Language for Extreme-Scale Computing

Taisuke Boku, University of Tsukuba, Japan; Hitoshi Murai, RIKEN, Japan; Masahiro Nakao, RIKEN Advanced Institute for Computational Science, Japan; Akihiro Tabuchi, University of Tsukuba, Japan; Mitsuhsa Sato, RIKEN Advanced Institute for Computational Science, Japan

#### 12:25-12:45 MPC: A Runtime Suited for Hybrid Programming

Patrick Carribault, CEA, France

#### 12:50-1:10 Sunway Openacc: Extended Programming Support for the Sunway Many-Core Architecture

Haohuan Fu, National Supercomputing Center, Wuxi, China and Tsinghua University, China

Friday, March 9

## MS75

### State-of-the-Art Auto-Tuning: New Approaches and Numerical Algorithmic Innovations Towards Exascale Computing - Part I of II

11:35 AM-1:15 PM

Room:52-202

#### For Part 2 see MS86

Exascale computers are expected to be deployed by 2022, and their architectures will be complex, building upon many-cores (such as the Intel Xeon Phi, Knights Landing) and GPUs (such as the NVIDIA Volta, V100). Auto-tuning (AT) research and technology will continue building upon its proven success for delivering high performance on a variety of computer architectures, and will seek to provide optimized, high performance implementations of specific computations for those challenging architectures. In concert with the hardware evolution and complexity, algorithmic innovations will need to be devised for numerical computations that are essential to many applications, such as eigenvalue solvers and linear equation solvers. For example, reducing communications together with high performance implementations will be essential. Exascale computers will enable the solution of problems with unprecedented levels of details for many applications of interest. However, the accuracy of these computations cannot be taken for granted: for example, codes that perform floating point sums of distributed data can produce results that are difficult to be reproduced. Given potentially conflicting goals for performance and reproducibility, this minisymposium will discuss technology trends for AT and their interplay with AT frameworks, numerical algorithms, and accuracy assurance towards the exascale computing era.



Organizer: Takahiro Katagiri  
*University of Tokyo, Japan*

Organizer: Osni A. Marques  
*Lawrence Berkeley National Laboratory, USA*

Organizer: Toshiyuki Imamura  
*RIKEN, Japan*

**11:35-11:55 High Precision Computing of Matrix-Matrix Multiplications and a New Approach of Auto-Tuning to Numerical Libraries by Deep Learning**

*Takahiro Katagiri, University of Tokyo, Japan; Shuntaro Ichimura and Kenya Yamada, Nagoya University, Japan*

**12:00-12:20 Use of Synthetic Benchmarks for Machine Learning-Based Performance Auto-Tuning**

*Tarek S. Abdelrahman, University of Toronto, Canada*

**12:25-12:45 Compiler-based Software Autotuning**

*Protonu Basu, Lawrence Berkeley National Laboratory, USA*

**12:50-1:10 Leveraging Formal Methods for Algorithmic Design**

*Tze Meng Low, Carnegie Mellon University, USA*

Friday, March 9

## MS76

**Disruptive Technologies for Future Science and Engineering - Part III of III**

*11:35 AM-1:15 PM*

*Room:52-204*

**For Part 2 see MS64**

The recent end of Dennard scaling is necessitating diverse and innovative changes in emerging high-end computing systems, for attaining ever-increasing performance while simultaneously addressing energy efficiency constraints. In this minisymposium we discuss disruptive technologies for next generation computing systems. We begin by presenting solutions for transforming conventional architectural approaches for sustained exascale performance. Next we discuss the potential of quantum computing to deliver unprecedented computing capability for key application drivers. Finally we present neuromorphic computing technologies that hope to leverage brain-inspired systems to alleviate some of science's big data challenges.

Organizer: Rupak Biswas  
*NASA Ames Research Center, USA*

Organizer: Jonathan Carter  
*Lawrence Berkeley National Laboratory, USA*

Organizer: Leonid Oliker

*Lawrence Berkeley National Laboratory, USA*

**11:35-11:55 Programmable Neuromorphic Learning Machines**

*Mike Davies and Narayan Srinivasa, Intel Corporation, USA*

**12:00-12:20 Neuromorphic Computing at Lawrence Livermore National Laboratory**

*Katie Lewis, Lawrence Livermore National Laboratory, USA*

**12:25-12:45 Neuromorphic Computing for Scientific Applications**

*Catherine Schuman, Oak Ridge National Laboratory, USA*

**12:50-1:10 Neural-Inspired Technologies for Data Processing and Scientific Computing**

*Conrad James and Craig Vineyard, Sandia National Laboratories, USA*

Friday, March 9

## MS77

**Hazard and Disaster Simulation of Earthquake and Tsunami with HPC- Part II of II**

*11:35 AM-1:15 PM*

*Room:52-301*

**For Part 1 see MS65**

The scope of the minisymposium is basically to discuss developments and directions for large scale earthquake and tsunami simulation with high performance computing technology. Broadband aspects from seismology to earthquake engineering are expected with a special emphasis on use of super computers (e.g. earthquake generation, seismic wave propagation, crustal deformation, structural response, soil amplification, city and social response, evacuation, recovery, fluid-structure coupling, global tsunami propagation, local tsunami run-up and the other related issues in earthquake and tsunami).

Organizer: Takane Hori

*Japan Agency for Marine-Earth Science and Technology, Japan*

Organizer: Takamasa Iryo

*Kobe University, Japan*

**11:35-11:55 Simulating Traffic Flow and Demand Pattern by a Parallel Computer**

*Takamasa Iryo, Junji Urata, Genaro Peque Jr., Kazuki Fukuda, and Riki Kawase, Kobe University, Japan; Lalith Wijerathne and Wasuwat Petprak, University of Tokyo, Japan*

**12:00-12:20 High Performance Computing Methods for Scheduling Transportation Network Maintenance**

*Hillel Bar-Gera, Ben-Gurion University, Israel; David Rey, University of New South Wales, Australia*

**12:25-12:45 Cyber-Physical System and Industrial Applications of Large-Scale Graph Analysis and Optimization Problem**

*Katsuki Fujisawa, Kyushu University, Japan*

**12:50-1:10 Dynamic Programming of Firms' Activities and Interactions in Disaster Recovery Process**

*Muneta Yokomatsu, Kyoto University, Japan; Yuki Akiyama, Yoshiki Ogawa, and Ryosuke Shibasaki, University of Tokyo, Japan*

Friday, March 9

## MS78

### Scalable Eigenvalue Computation - Part I of II

11:35 AM-1:15 PM

Room:52-302

#### For Part 2 see MS89

The eigenvalue computation of large sparse but also dense matrices is a central element in the process of finding solutions for many new applications related to social networks and “Big Data” in addition to the more classical ones of numerical simulation. This minisymposium focuses on solving this problem on new and emerging supercomputers. More specifically, the mathematic, algorithmic and implementation techniques allowing to realize efficient scalable eigensolver including optimization of communication, autotuning ... are targeted.

Organizer: Tetsuya Sakurai  
*University of Tsukuba, Japan*

Organizer: Nahid Emad  
*University of Versailles, France*

#### 11:35-11:55 Scalable Eigensolver for Large-scale Electronic Structure Calculations

Chao Yang and Eugene Vecharynski,  
Lawrence Berkeley National  
Laboratory, USA; Yousef Saad,  
University of Minnesota, USA;  
Ruipeng Li, Lawrence Livermore  
National Laboratory, USA; Yuanzhe  
Xi, University of Minnesota, USA

#### 12:00-12:20 Data Analysis with SVD for Physical Experiments, Application to the Cherenkov Telescope Array

Pierre Aubert, Thomas Vuillaume,  
Florian Gaté, Gilles Maurin, and  
Jean Jacquemier, CNRS, France;  
Nahid Emad, University of Versailles,  
France; Giovanni Lamanna, CNRS,  
France

#### 12:25-12:45 Scalable Computation of Many Eigenpairs of a Large Hermitian Matrix

Luan Nguyen and Zhaojun Bai,  
University of California, Davis, USA

#### 12:50-1:10 Contour Integral Method to Evaluate Self-Energy Matrices for Large-Scale Electron Transport Calculations

Shigeru Iwase, Yasunori Futamura,  
Akira Imakura, and Tetsuya Sakurai,  
University of Tsukuba, Japan; Shigeru  
Tsukamoto, Forschungszentrum  
Jülich, Germany; Tomoya Ono,  
University of Tsukuba, Japan

Friday, March 9

## MS79

### Towards Graph Algorithms and Analytics for Exascale Applications - Part I of II

11:35 AM-1:15 PM

Room:52-303

#### For Part 2 see MS90

The scientific community has a number of ongoing efforts to solve challenging applications on exascale computers, which are being co-designed with algorithms and software. Graph algorithms play a critical enabling role in many of these applications in the sciences, engineering, data science, etc. The increasing sizes of graphs deployed in these applications make highly concurrent algorithms essential for solving these problems. However, the irregular memory access inherent in graph algorithms makes them some of the hardest algorithmic kernels to implement in parallel. The prevalence of dynamic graphs and semantic graphs intensify these algorithmic challenges. The eight speakers in this minisymposium will discuss a number of parallel graph algorithms that play key enabling roles in applications, including graph traversals, matching, edge covers, coloring, clustering, partitioning, sparse matrix multiplication, etc. Some of the talks will also discuss software frameworks for efficient implementation of these algorithms on hierarchical distributed-memory architectures representative of potential exascale platforms.

Organizer: Alex Pothen  
*Purdue University, USA*

#### 11:35-11:55 Parallel Algorithm Design Via Approximation

Alex Pothen, Purdue University, USA;  
Arif Khan, Pacific Northwest National  
Laboratory, USA

#### 12:00-12:20 Connected Components and Spanning Trees on Multi-GPU Systems

Fredrik Manne and Md Naim, University  
of Bergen, Norway; Jean Blair, United  
States Military Academy, USA

**12:25-12:45 Mining for Communities in Large-scale Graphs**

Ananth Kalyanaraman, Washington State University, USA; Mahantesh Halappanavar, Pacific Northwest National Laboratory, USA; Hao Lu, Oak Ridge National Laboratory, USA

**12:50-1:10 Productive Graph Analytics Enabled by GraphBLAS**

Ariful Azad and Aydin Buluc, Lawrence Berkeley National Laboratory, USA

Friday, March 9

**MS80****Challenges in Parallel Adaptive Mesh Refinement - Part II of III**

11:35 AM-1:15 PM

Room: 52-304

**For Part 1 see MS68**

**For Part 3 see MS91**

Parallel adaptive mesh refinement (AMR) is a key technique when simulations are required to capture time-dependent and/or multiscale features. Frequent re-adaptation and repartitioning of the mesh during the simulation can impose significant overhead, particularly in large-scale parallel environments. Further challenges arise due to the availability of accelerated or special-purpose hardware, and the trend toward hierarchical and hybrid compute architectures. Our minisymposium addresses algorithms, scalability, and software issues of parallel AMR.

Organizer: Carsten Burstedde  
*Universität Bonn, Germany*

Organizer: Michael Bader  
*Technische Universität München, Germany*

Organizer: Martin Berzins  
*University of Utah, USA*

**11:35-11:55 Parallel Adaptive Mesh Refinement in  $\text{Sam}(\text{oa})^2$  -- Load Balancing vs. Work Stealing**

Philipp Samfass, Technical University of Munich, Germany

**12:00-12:20 The Triumvirate of Adaptivity in the Exahype Hyperbolic Equation Solvers: Mesh, Discretisation and Parameter Space Adaption**

Tobias Weinzierl, Durham University, United Kingdom

**12:25-12:45 Comparison Between Different Versions of AMR**

Klaus Weide, University of Chicago, USA

**12:50-1:10 Phase Asynchronous AMR Algorithm**

Didem Unat, Lawrence Berkeley National Laboratory, USA; Nufail Farooqi, Koc University, Turkey; Tan Nguyen, Weiqun Zhang, and Ann S. Almgren, Lawrence Berkeley National Laboratory, USA

Friday, March 9

**MS81****Hierarchical Low Rank Approximation Methods - Part V of VI**

11:35 AM-1:15 PM

Room: 57-201

**For Part 4 see MS69**

**For Part 6 see MS92**

Hierarchical Low-Rank Approximation Methods such as H-matrices, Hierarchical Semi-Seperable matrices, and Hierarchical Interpolative Factorization are increasing in popularity. These methods have near linear complexity and can be used as a fast approximate inverse of both dense matrices arising from integral equations and sparse matrices arising from partial differential equations. Accuracy is controllable by increasing the rank of the approximated blocks or by changing the subdivision criteria of the off-diagonal blocks. Therefore, these methods can be used as robust direct solvers with near linear complexity if high accuracy is chosen or they can be used as preconditioners if lower accuracy is used. With the advent of these methods, large dense linear algebra operations are scarcely needed anymore. In terms of parallelism, the hierarchical nature of these methods pose a challenge. Task-based approaches seem to be effective for shared memory parallelism, while load-balance becomes a problem on distributed memory. The recursive block structure also results in the computation of many small dense blocks. The recent development in batched dense linear algebra libraries facilitates the efficient computation of such small dense blocks. This minisymposium consists of cutting edge research on hierarchical low-rank approximation methods, their applications, parallelization, and extension.

Friday, March 9

## MS81

### Hierarchical Low Rank Approximation Methods - Part V of VI

11:35 AM-1:15 PM

continued

Organizer: Rio Yokota  
*Tokyo Institute of Technology, Japan*

Organizer: Akihiro Ida  
*University of Tokyo, Japan*

Organizer: Chao Chen  
*Stanford University, USA*

Organizer: Eric F. Darve  
*Stanford University, USA*

**11:35-11:55 Boundary Integral Equations for Calculating Complex Eigenvalues for Open Domains**  
*Naoshi Nishimura, Kyoto University, Japan*

**12:00-12:20 Generalizations to the Superfast Divide-and-Conquer Eigenvalue Algorithm**  
*James Vogel, Purdue University, USA*

**12:25-12:45 Electrostatic Field Analyses of Voxel-based Biological Models by Indirect-BEM and Static-MoM Combined with FMM**  
*Shoji Hamada, Kyoto University, Japan*

**12:50-1:10 Accelerating Hierarchical-Matrix Based Linear Solver on a GPU Cluster**  
*Ichitaro Yamazaki, University of Tennessee, Knoxville, USA; Satoshi Ohshima, University of Tokyo, Japan; Akihiro Ida, University of Tokyo, Japan; Rio Yokota, Tokyo Institute of Technology, Japan; Jack J. Dongarra, University of Tennessee, ORNL, and University of Manchester, USA*

Friday, March 9

## MS82

### Parallel Programming Models, Algorithms and Frameworks for Extreme Computing - Part II of III

11:35 AM-1:15 PM

*Room: 57-202*

**For Part 1 see MS70**

**For Part 3 see MS93**

Many-core and GPU processors are the foundation for almost all leadership computing platforms. Application and library software developers may often effectively use these processors and some general approaches have emerged. It is widely recognized that careful design of software and data structures, with effective memory management, are the most critical challenges to obtaining scalable optimized performance on those systems. Furthermore, investment in formal software techniques for robust environments is needed for advancing simulation capabilities. In these minisymposia we discuss current experiences and development of applications, libraries and frameworks using a variety of hardware. Speakers will address performance results and software design.

Organizer: Michael Heroux  
*Sandia National Laboratories, USA*

Organizer: Kengo Nakajima  
*University of Tokyo, Japan*

Organizer: Serge G. Petiton  
*Université Lille 1 and CNRS, France*

**11:35-11:55 Development of Large-Scale Scientific & Engineering Applications on Post-Peta/Exascale Systems Using ppOpen-HPC**  
*Kengo Nakajima, University of Tokyo, Japan*

**12:00-12:20 Spectral Graph Analysis with Unite and Conquer Approach**  
*Nahid Emad, University of Versailles, France; Serge Petiton, University of Lille, France; Alexandre Fender, Joe Eaton, and Maxim Naumov, NVIDIA, USA*

**12:25-12:45 Production Implementations of Pipelined and Communication-Avoiding Iterative Linear Solvers**

*Mark Hoemmen, Sandia National Laboratories, USA; Ichitaro Yamazaki, University of Tennessee, Knoxville, USA*

**12:50-1:10 Algebraic Multigrid Methods, Data Structures and Performance**

*Ulrike Meier Yang, Lawrence Livermore National Laboratory, USA*



Friday, March 9

## CP9

### Neural Networks and Machine Learning

11:35 AM-1:15 PM

Room:52-101

Chair: Sarah Knepper, Intel Corporation, USA

#### 11:35-11:55 A Waveform-Relaxation Method for the Simulation of Rate-Based Models in a Distributed Spiking Neural Network Simulator

Jan Hahne and Matthias Bolten, University of Wuppertal, Germany

#### 12:00-12:20 Feature Scaling Method for Supervised Spectral Clustering

Momo Matsuda, Keiichi Morikuni, and Tetsuya Sakurai, University of Tsukuba, Japan

#### 12:25-12:45 Neural Network Architectures for Kalman Filters

Elizabeth C. Ramirez, Columbia University, USA

#### 12:50-1:10 Predicting Biological Cell Aggregation Using Scalable Random Forest Decision Trees

Sulimon Sattari and Tamiki Komatsuzaki, Hokkaido University, Japan

## Lunch Break

1:15 PM-2:30 PM

Attendees On Their Own

Friday, March 9

## MS83

### Fully Parallel, Scalable Solution Algorithms for Fluids and Astrodynamics

2:30 PM-4:10 PM

Room:52-103

In recent years, with the rapid advance of high-performance computers, the numerical simulation for 3D complex biomedical or engineering problems within a reasonable time becomes possible. The current trend is to replace some algorithm that is sequential in nature, such as operator-splitting algorithm, time-marching approach or reduced space method by another fully parallel algorithm to enhance the scalability of a kernel solver in the simulator. This minisymposium is dedicated to the discussion of most recent developments in fully parallel, scalable state-of-the-art solution algorithms and their applications in space missions, multiphase flows in porous media, and patient-specific blood flow simulation.

Organizer: Feng-Nan Hwang  
National Central University, Taiwan

#### 2:30-2:50 A Parallel Full Space Lagrange-Newton-Krylov Algorithm for Trajectory Optimization Problems in Space Missions

Feng-Nan Hwang, National Central University, Taiwan

#### 2:55-3:15 A Fully Implicit Constraint-Preserving Simulator for Multiphase Flow in Porous Media

Haijian Yang, Hunan University, China

#### 3:20-3:40 A Semi-Implicit Solver for Two Phase Flow with Moving Contact Line on Rough Surface

Jizu Huang, Chinese Academy of Sciences, China

#### 3:45-4:05 Parallel Monolithic Algorithm for Simulation of Patient-Specific Coronary Artery

Wen-Shin Shiu, Shenzhen Institute of Advanced Technology, China; Rongliang Chen, Chinese Academy of Sciences, China; Zhengzheng Yan and Jia Liu, Shenzhen Institute of Advanced Technology, China; Xiao-Chuan Cai, University of Colorado Boulder, USA

Friday, March 9

## MS84

### Graph Algorithms Related to Sparse Matrix Computations

2:30 PM-4:10 PM

Room:52-104

In this minisymposium, we will discuss various graph problems related to sparse matrix computations. Ranging from fill-reducing ordering problems (of interest to sparse direct linear solvers), graph matching problems, to reordering techniques for low-rank based fast solvers, graph algorithms are often critical to sparse matrix computations.

Organizer: Mathias Jacquelin  
Lawrence Berkeley National Laboratory, USA

Organizer: Esmond G. Ng  
Lawrence Berkeley National Laboratory, USA

#### 2:30-2:50 Enhancing Performance of Sparse Matrix Factorizations Via Ordering Refinements

Esmond G. Ng, Lawrence Berkeley National Laboratory, USA

#### 2:55-3:15 Variable Clustering Techniques for Low-rank Factorization-based Linear Solvers and Preconditioners

Francois-Henry Rouet, Livermore Software Technology Corporation, USA

#### 3:20-3:40 Multi-Level Spectral Nested-Dissection Graph Ordering

Mathias Jacquelin, Lawrence Berkeley National Laboratory, USA

#### 3:45-4:05 Title Not Available At Time Of Publication

Yousef Saad, University of Minnesota, USA

continued in next column



Friday, March 9

## MS85

**Performance-Aware Programming -- Performance Engineering and Modeling in Practice**  
2:30 PM-4:10 PM

*Room: 52-201*

As the system complexity increases, it is often said that the programmers will have to invest more time and effort to achieve “reasonable” sustained-performance on actual scientific simulations. It is probably true. However, how can we define the reasonable performance for a given application and/or a given system? How much time and effort should we invest on improving the performance of applications? There are various open problems in the area of performance-aware programming of scientific simulations. This minisymposium focuses on the importance of performance engineering and modeling in the extreme-scale computing era, and introduces several approaches to the open problems.

Organizer: Hiroyuki Takizawa  
*Tohoku University, Japan*

Organizer: Ritu Arora  
*Texas Advanced Computing Center, USA*  
**2:30-2:50 User-Defined Code Transformation for Separation of Performance-Awareness from Application Codes**

*Hiroyuki Takizawa, Tohoku University, Japan*

**2:55-3:15 Raising the Level of Abstraction of Developing Optimized Parallel Applications**  
*Ritu Arora, Texas Advanced Computing Center, USA*

**3:20-3:40 Performance Engineering for Sparse Linear Algebra Kernels: Navigating Between Models and Expectations**

*Gerhard Wellein, Erlangen Regional Computing Center, Germany*

**3:45-4:05 Performance Engineering Scientific Applications Using the Tau Performance System**

*Sameer Shende, University of Oregon, USA*

Friday, March 9

## MS86

**State-of-the-Art Auto-Tuning: New Approaches and Numerical Algorithmic Innovations Towards Exascale Computing - Part II of II**  
2:30 PM-4:10 PM

*Room: 52-202*

**For Part I see MS75**

Exascale computers are expected to be deployed around 2022 -- they architectures will be complex, building upon many-cores (such as the Intel Xeon Phi, Knights Landing)) and GPUs (such as the NVIDIA Volta, V100). Auto-tuning (AT) technology will built upon its proven success for delivering high performance on a variety of computer architectures, and will seek to provide optimized, high performance implementations of specific computations for the challenging upcoming architectures. In concert with the hardware evolution, algorithmic innovations will have to be devised for numerical computations that are essential to many applications, such as eigenvalue solvers and linear equation solvers. In particular, reducing communications together with high performance implementations will be essential. Exascale computers will enable the solution of problems with unprecedented levels of details for many applications of interest, implying very large problems. However, the accuracy of these computations cannot be taken for granted: usually, codes that perform floating point sums of distributed data can produce results that are difficult to be reproduced. Given these conflicting goals, this minisymposium will discuss technologies for AT and its interplay with AT frameworks, numerical algorithms, and accuracy assurance towards the exascale computing era.

Organizer: Toshiyuki Imamura  
*RIKEN, Japan*

Organizer: Osni A. Marques  
*Lawrence Berkeley National Laboratory, USA*

Organizer: Takahiro Katagiri  
*University of Tokyo, Japan*

**2:30-2:50 Communication Avoiding Approach for Reduction to Tri-Diagonal, Bi-Diagonal, and Hessenberg Forms**

*Toshiyuki Imamura and Yusuke Hirota, RIKEN, Japan; Susumu Yamada and Masahiko Machida, Japan Atomic Energy Agency, Japan*

**2:55-3:15 Integrating Multiple Random Sketches for Singular Value Decomposition**

*Ting-Li Chen, Academia Sinica, Taiwan; Dawei D. Chang, National Taiwan University, Taiwan; Su-Yun Huang, National Institute of Statistical Sciences, USA; Hung Chen, National Taiwan University, Taiwan; Chienyao Lin, Academia Sinica, Taiwan; Weichung Wang, National Taiwan University, Taiwan*

**3:20-3:40 Validated Solution of Linear Systems for Real Symmetric and Positive Definite Matrices**

*Takeshi Terao and Katsuhisa Ozaki, Shibaura Institute of Technology, Japan*

**3:45-4:05 Unconstrained Functionals for Efficient Parallel Scaling of Conjugate Gradient Eigensolvers**

*Osni A. Marques, Andrew M. Canning, and Mauro Del ben, Lawrence Berkeley National Laboratory, USA*

Friday, March 9

## MS87

### Innovative Methods for High Performance Iterative Solvers - Part I of II

2:30 PM-4:10 PM

Room:52-204

#### For Part 2 see MS98

Iterative methods are commonly used in HPC applications to solve large size linear systems, eigenvalue or least squares problems. To be able to take advantage of the latest generations of processors and to be scalable on future Exascale platforms, such iterative methods must be adapted or sometimes rethought. The speakers of this mini symposium describe recent advances in iterative linear solvers including communication-avoiding Krylov subspace methods, optimized implementations for SpMV and preconditioners, but also new challenges in domain decomposition methods and fault tolerance.

Organizer: Marc Baboulin

*University of Paris-Sud, France*

Organizer: Takeshi Fukaya

*Hokkaido University, Japan*

Organizer: Takeshi Iwashita

*Hokkaido University, Japan*

#### 2:30-2:50 Enhancing a Parallel Iterative Solver Through Randomization and GPU Computing

Marc Baboulin, University of Paris-Sud, France; Evan Coleman, Old Dominion University, USA; Aygul Jamal and Amal Khabou, University of Paris-Sud, France; Masha Sosonkina, Old Dominion University, USA

#### 2:55-3:15 Performance Property of Preconditioned Chebyshev Basis CG Solver for Multiphase CFD Simulations

Akie Mayumi, Yasuhiro Idomura, Takuya Ina, and Susumu Yamada, Japan Atomic Energy Agency, Japan; Toshiyuki Imamura, RIKEN, Japan

#### 3:20-3:40 ParILUT - A New Parallel Threshold ILU

Hartwig Anzt, University of Tennessee, USA; Edmond Chow, Georgia Institute of Technology, USA; Jack J. Dongarra, University of Tennessee, ORNL, and University of Manchester, USA

#### 3:45-4:05 Shifted Cholesky QR for Computing the QR Factorization for Ill-conditioned Matrices

Yuka Yanagisawa, Waseda University, Japan; Takeshi Fukaya, Hokkaido University, Japan; Yuji Nakatsukasa, University of Oxford, United Kingdom; Yusaku Yamamoto, University of Electro-Communications, Japan; Ramseshan Kannan, University of Manchester, United Kingdom

Friday, March 9

## MS88

### Scientific Accomplishments on Massively Parallel Computing of Earth and Planetary Sciences - Part I of II

2:30 PM-4:10 PM

Room:52-301

#### For Part 2 see MS99

With greatly progressing on technologies for both numerical scheme and computational capability, many fundamental issues in solid Earth geosciences have been tackled with massively parallel computing, which is known as generating plate tectonics in mantle convection, secular variations of geomagnetic field in geodynamo and early warning of occurrence of large-scale earthquakes. In this mini symposium, we introduce important scientific accomplishments using massively parallel computing on seismic wave propagations, mantle convection and geo- and stellar-dynamo simulations and first principle computations of physical properties of Earth's material, which have been applied for fundamental topics, but still unresolved, in Earth and Planetary Sciences. This mini symposium is tied up with 'Computational Infrastructure of Geodynamics' running in the Geoscience community in the United States, which is successfully organized for providing 'community code' on various fundamental topics on Earth and Planetary Sciences with massively parallel computing, computational geodynamics community in Europe, and collaborated with applied mathematics and computational sciences, and with FLAGSHIP 2020 project, MEXT within CBSM2 Project "Structure and Properties of Materials in deep Earth and planets".

Friday, March 9

## MS88

### Scientific Accomplishments on Massively Parallel Computing of Earth and Planetary Sciences - Part I of II

2:30 PM-4:10 PM

continued

Organizer: Takashi Nakagawa  
*Japan Agency for Marine-Earth Science and Technology, Japan*

Organizer: Hiroaki Matsui  
*University of California, Davis, USA*

Organizer: Koichiro Umemoto  
*Tokyo Institute of Technology, Japan*

#### 2:30-2:50 Frontiers of Scientific Computation in Geodynamics

*Louise Kellogg, University of California, Davis, USA*

#### 2:55-3:15 Imaging Earth's Mantle: Global Full-Waveform Inversion Based on Spectral-Element Simulations & Adjoint Methods

*Ebru Bozdag, Colorado School of Mines, USA; Matthieu Lefebvre and Wenjie Lei, Princeton University, USA; Ridvan Orsvuran, University of Cote d'Azur, France; Daniel Peter, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; Youyi Ruan and James Smith, Princeton University, USA; Dimitri Komatitsch, CNRS & Universite de Marseille, France; Jeroen Tromp, Princeton University, USA*

#### 3:20-3:40 Advances in Mantle Convection Modelling: Nonlinear Solvers, Multiphysics, Linking Scales

*Rene Gassmoeller, Wolfgang Bangerth, and Juliane Dannberg, Colorado State University, USA; Timo Heister, Clemson University, USA*

#### 3:45-4:05 Exploring Planetary and Stellar Convection Using the Rayleigh Code

*Nicholas A. Featherstone, University of Colorado Boulder, USA*

Friday, March 9

## MS89

### Scalable Eigenvalue Computation - Part II of II

2:30 PM-4:10 PM

*Room:52-302*

#### For Part 1 see MS78

The eigenvalue computation of large sparse but also dense matrices is a central element in the process of finding solutions for many new applications related to social networks and "Big Data" in addition to the more classical ones of numerical simulation. This minisymposium focuses on solving this problem on new and emerging supercomputers. More specifically, the mathematic, algorithmic and implementation techniques allowing to realize efficient scalable eigensolver including optimization of communication, autotuning ... are targeted.

Organizer: Tetsuya Sakurai  
*University of Tsukuba, Japan*

Organizer: Nahid Emad  
*University of Versailles, France*

#### 2:30-2:50 A Parallel Generator of Non-Hermitian Matrices Computed from Known Given Spectra

*Xinzhe Wu, CNRS, France; Serge Petiton, University of Lille, France; Hervé Galicher, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; Christophe Calvin, CEA Saclay, France*

#### 2:55-3:15 Addressing Scalability in SLEPc Eigensolvers

*Jose E. Roman, Universidad Politecnica de Valencia, Spain*

#### 3:20-3:40 Distributed Parallel Implementation of the Sakurai-Sugiura Method for Large Dense Generalized Eigenvalue Problems

*Takahiro Yano, Yasunori Futamura, and Tetsuya Sakurai, University of Tsukuba, Japan*

#### 3:45-4:05 Multi-GPU Scaling of Eigenvalue Computations for Spectral Partitioning

*Joe Eaton, Alex Fender, and Maxim Naumov, NVIDIA, USA; Nahid Emad, University of Versailles, France; Serge Petiton, University of Paris, France*

Friday, March 9

## MS90

### Towards Graph Algorithms and Analytics for Exascale Applications - Part II of II

2:30 PM-4:10 PM

*Room:52-303*

#### For Part 1 see MS79

The scientific community has a number of ongoing efforts to solve challenging applications on exascale computers, which are being co-designed with algorithms and software. Graph algorithms play a critical enabling role in many of these applications in the sciences, engineering, data science, etc. The increasing sizes of graphs deployed in these applications make highly concurrent algorithms essential for solving these problems. However, the irregular memory access inherent in graph algorithms makes them some of the hardest algorithmic kernels to implement in parallel. The prevalence of dynamic graphs and semantic graphs intensify these algorithmic challenges. The eight speakers in this minisymposium will discuss a number of parallel graph algorithms that play key enabling roles in applications, including graph traversals, matching, edge covers, coloring, clustering, partitioning, sparse matrix multiplication, etc. Some of the talks will also discuss software frameworks for efficient implementation of these algorithms on hierarchical distributed-memory architectures representative of potential exascale platforms.

Organizer: Alex Pothen  
*Purdue University, USA*

#### 2:30-2:50 Performance of a Scalable Programming Environment for Graph Algorithms

*John T. Feo, Pacific Northwest National Laboratory, USA*

*continued on next page*

**2:55-3:15 Asynchronous, One-Sided Communication: Programming Support for Distributed-Memory Scientific Applications**

Assefaw Gebremedhin and Sayan Ghosh,  
Washington State University, USA;  
Yanfei Guo and Pavan Balaji,  
Argonne National Laboratory, USA

**3:20-3:40 Performance Portable Sparse Matrix Matrix Multiplication with Applications in Scientific Computing and Graph Analytics**

Mehmet Deveci, Michael M.  
Wolf, Jonathan W. Berry, Siva  
Rajamanickam, Erik G. Boman,  
and Christian Trott, Sandia National  
Laboratories, USA; Neil Butcher,  
University of Notre Dame, USA;  
Simon D. Hammond and Stephen  
Olivier, Sandia National Laboratories,  
USA

**3:45-4:05 Balanced Partition Refinement with the Graph P-Laplacian**

Toby Simpson and Olaf Schenk,  
Università della Svizzera italiana,  
Switzerland

Friday, March 9

**MS91**

**Challenges in Parallel Adaptive Mesh Refinement - Part III of III**

2:30 PM-4:10 PM

Room:52-304

**For Part 2 see MS80**

Parallel adaptive mesh refinement (AMR) is a key technique when simulations are required to capture time-dependent and/or multiscale features. Frequent re-adaptation and repartitioning of the mesh during the simulation can impose significant overhead, particularly in large-scale parallel environments. Further challenges arise due to the availability of accelerated or special-purpose hardware, and the trend toward hierarchical and hybrid compute architectures. Our minisymposium addresses algorithms, scalability, and software issues of parallel AMR.

Organizer: Carsten Burstedde  
Universität Bonn, Germany

Organizer: Michael Bader  
Technische Universität München, Germany

Organizer: Martin Berzins  
University of Utah, USA

**2:30-2:50 Extreme-Scale Block-Structured Adaptive Mesh Refinement**

Florian Schornbaum, University of  
Erlangen-Nuernberg, Germany

**2:55-3:15 Implementation of Massively Parallel Hyperbolic Multiresolution Methods in the Block-structured Mesh Refinement Framework AMROC**

Ralf Deiterding, University of  
Southampton, United Kingdom;  
Margarete Domingues, Instituto  
Nacional de Pesquisa Espaciais, Brazil

**3:20-3:40 Performance Modeling of Adaptive Mesh Refinement on Modern Supercomputers**

Mohamed Wahib, RIKEN, Japan

**3:45-4:05 An AMR Framework for Realizing Effective High-Resolution Simulations on Multiple GPUs**

Takashi Shimokawabe, University of  
Tokyo, Japan; Takayuki Aoki, Tokyo  
Institute of Technology, Japan;  
Naoyuki Onodera, Japan Atomic  
Energy Agency, Japan

Friday, March 9

**MS92**

**Hierarchical Low Rank Approximation Methods - Part VI of VI**

2:30 PM-4:10 PM

Room:57-201

**For Part 5 see MS81**

Hierarchical Low-Rank Approximation Methods such as H-matrices, Hierarchical Semi-Separable matrices, and Hierarchical Interpolative Factorization are increasing in popularity. These methods have near linear complexity and can be used as a fast approximate inverse of both dense matrices arising from integral equations and sparse matrices arising from partial differential equations. Accuracy is controllable by increasing the rank of the approximated blocks or by changing the subdivision criteria of the off-diagonal blocks. Therefore, these methods can be used as robust direct solvers with near linear complexity if high accuracy is chosen or they can be used as preconditioners if lower accuracy is used. With the advent of these methods, large dense linear algebra operations are scarcely needed anymore. In terms of parallelism, the hierarchical nature of these methods pose a challenge. Task-based approaches seem to be effective for shared memory parallelism, while load-balance becomes a problem on distributed memory. The recursive block structure also results in the computation of many small dense blocks. The recent development in batched dense linear algebra libraries facilitates the efficient computation of such small dense blocks. This minisymposium consists of cutting edge research on hierarchical low-rank approximation methods, their applications, parallelization, and extension.



Friday, March 9

## MS92

### Hierarchical Low Rank Approximation Methods - Part VI of VI

2:30 PM-4:10 PM

continued

Organizer: Rio Yokota  
*Tokyo Institute of Technology, Japan*

Organizer: Akihiro Ida  
*University of Tokyo, Japan*

Organizer: Chao Chen  
*Stanford University, USA*

Organizer: Eric F. Darve  
*Stanford University, USA*

#### 2:30-2:50 Parallel Butterfly-Based Direct Solvers for Highly Oscillatory Problems

*Yang Liu, Lawrence Berkeley National Laboratory, USA*

#### 2:55-3:15 Integer Matrix Approximation and Its Applications

*Matthew M. Lin, National Cheng Kung University, Taiwan*

#### 3:20-3:40 Parallelization of the IFMM-based Preconditioner for 3D Helmholtz BEM

*Toru Takahashi, Nagoya University, Japan; Chao Chen and Eric F. Darve, Stanford University, USA*

#### 3:45-4:05 Asynchronous Communication in Hierarchical Low-Rank Solvers

*Rio Yokota, Tokyo Institute of Technology, Japan*

Friday, March 9

## MS93

### Parallel Programming Models, Algorithms and Frameworks for Extreme Computing - Part III of III

2:30 PM-4:10 PM

*Room: 57-202*

#### For Part 2 see MS82

Many-core and GPU processors are the foundation for almost all leadership computing platforms. Application and library software developers may often effectively use these processors and some general approaches have emerged. It is widely recognized that careful design of software and data structures, with effective memory management, are the most critical challenges to obtaining scalable optimized performance on those systems. Furthermore, investment in formal software techniques for robust environments is needed for advancing simulation capabilities. In these minisymposia we discuss current experiences and development of applications, libraries and frameworks using a variety of hardware. Speakers will address performance results and software design.

Organizer: Michael Heroux  
*Sandia National Laboratories, USA*

Organizer: Kengo Nakajima  
*University of Tokyo, Japan*

Organizer: Serge G. Petiton  
*Université Lille 1 and CNRS, France*

#### 2:30-2:50 Efficient Load Balancing for High Performance Multi-Physics Computations

*Cédric Chevalier and Remi Barat, CEA/DAM, France; Francois Pellegrini, University of Bordeaux, France*

#### 2:55-3:15 Fast Linear Solvers Exploiting Hierarchical Matrix Structures and Hardware Organization

*X. Sherry Li, Gustavo Chavez, and Pieter Ghysels, Lawrence Berkeley National Laboratory, USA; Chris Gorman, University of California, Santa Barbara, USA; Yang Liu, Lawrence Berkeley National Laboratory, USA*

#### 3:20-3:40 Efficient Monte Carlo Transport Methods on Manycore Architecture

*Christophe Calvin, CEA Saclay, France; Yunsong Wang, CEA, France; Emeric Brun and Fausot Malvagi, CEA, DEN, SRMP, France*

#### 3:45-4:05 Efficient Sparse General Matrix-Matrix Multiplication Algorithms for Many-Core Processors

*Yusuke Nagasaka and Akira Nukada, Tokyo Institute of Technology, Japan; Ariful Azad and Aydin Buluc, Lawrence Berkeley National Laboratory, USA; Satoshi Matsuoka, Tokyo Institute of Technology, Japan*

Friday, March 9

**CP10****GPU Computing and Application**

2:30 PM-4:10 PM

Room:52-101

*Chair: Matthew G. Knepley, University of Buffalo, USA***2:30-2:50 Scalable Krylov-FFT-Type Algorithms for Subsurface Scattering Problems***Yury A. Gryazin, Ron Gonzales, and Yun Teck Lee, Idaho State University, USA***3:55-3:15 A Randomized, Communication-Efficient Algorithm for Computing a Rank-Revealing UTV Matrix Decomposition***Nathan Heavner, University of Colorado, USA; Per-Gunnar Martinsson, University of Oxford, United Kingdom; Gregorio Quintana-Orti, Universidad Jaume I, Spain***3:20-3:40 An Effective GPU Implementation of the Adaptive Pattern FSAI Preconditioner for SPD Problems***Massimo Bernaschi and Mauro Carrozzo, Istituto Applicazioni del Calcolo, Italy; Ferronato Massimiliano and Carlo Janna, University of Padova, Italy*

Friday, March 9

**CP11****Algorithms**

2:30 PM-4:10 PM

Room:52-102

*Chair: Ümit Çatalyürek, Georgia Institute of Technology, USA***2:30-2:50 Scalable Spectral Clustering***Andrew Knyazev, Mitsubishi Electric Research Laboratories, USA***2:55-3:15 Low Communication Neighbor Discovery for Matrix Migration***Christopher Siefert, Sandia National Laboratories, USA; Christopher Luchini, Sci Tac LLC, USA***3:20-3:40 A Molecular Fingerprint Algorithm for Learning Force Fields from Ab Initio Calculations***Yu-Hang Tang, Lawrence Berkeley National Laboratory, USA***Coffee Break**

4:10 PM-4:40 PM

Room:63-1F

Friday, March 9

**IP5****Fast Scalable Implicit Solver for Low-Ordered Unstructured Finite Element Analysis and Its Application**

4:40 PM-5:25 PM

Room:63-201/202

*Chair: Kengo Nakajima, University of Tokyo, Japan*

We developed a fast scalable implicit solver that uses low-order unstructured finite element (FE) analysis to estimate static/dynamic and nonlinear/linear responses of a solid body. As this type of FE analysis uses sparse matrix-vector multiplication with random data access, it does not generally perform well on modern computer architectures. We overcame this bottleneck by developing special algorithms that are better suited to current computer architectures. These algorithms have many characteristics, including multi-grid methods, adaptive conjugate gradient methods, multi-precision arithmetic, element-by-element methods, predictors based on Adams-Bashforth methods and grid-structuring methods. Our solver performs well in terms of scalability, time-to-solution and peak performance. The superiority of this solver on the K computer led to it being selected as a finalist for the SC14 and SC15 Gordon Bell Prize, and for the SC16 Best Poster Award. In this talk, we describe the solver in detail with reference to its applications to earthquake problems, which we were able to analyze using nonlinear wave FE analysis with a trillion degrees of freedom.

*Tsuyoshi Ichimura**University of Tokyo, Japan*

Friday, March 9

**Dinner Banquet**

(A separate fee of 3,000 JPY is required to attend)

7:00 PM-9:00 PM

Room:Hotel Chinzanso Tokyo

## Saturday, March 10

### Registration

8:00 AM-5:00 PM

Room:63-1F

## IP6

### ChainerMN: Scalable Distributed Deep Learning with Chainer

8:30 AM-9:15 AM

Room:63-201/202

Chair: Satoshi Matsuoka, Tokyo Institute of Technology, Japan

We'll present ChainerMN, a multi-node distributed deep learning framework, together with the basics of distributed deep learning. Even though GPUs are continuously gaining more computation throughput, it is still very time-consuming to train state-of-the-art deep neural network models. For better scalability and productivity, it is paramount to accelerate the training process by using multiple GPUs. To enable high-performance and flexible distributed training, we developed ChainerMN, built on top of Chainer. We'll first introduce the basic approaches to distributed deep learning. Then, we'll explain the design choice, basic usage, and implementation details of Chainer and ChainerMN. We'll report benchmark results using hundreds of GPUs and discuss the future directions of distributed deep learning.

Takuya Akiba

Preferred Networks, Inc., Japan

### Intermission

9:15 AM-9:25 AM

Saturday, March 10

## MS94

### Data Analytics in HPC: An Applications' Perspective

9:25 AM-11:05 AM

Room:52-103

Everybody speaks about data analytics in HPC but when it is time to concretely integrate any analysis into traditional workflows, scientists still face both platform- and simulation-specific challenges. As we move to exascale platforms, the concurrency level of our platforms increases by several orders of magnitude and multiple high-end clusters will share congested, and too often old, parallel file system (PFS), as funding is preferably invested in cheaper computing nodes than more expensive file systems. Traditional centralized, post-simulation data analyses are no longer feasible as I/O and data movement becomes too expensive. This session takes a close look to how applications in four key fields (i.e., molecular dynamics, bioinformatics, combustion and astronomy) are tackling the big data challenges in an increasingly concurrent computational world. Presenters outline both general approaches and specific challenges in data analytics for four representative applications in the four fields of interests. Particular attention will be paid to the parallel middleware packages enabling large-scale data analysis for these applications' data and how these packages have transformed the centralized nature of data analytics into a distributed approach including \textit{in situ} and \textit{in transit} processing, data management, and I/O optimization.

Organizer: Michela Taufer

University of Delaware, USA

### 9:25-9:45 Transitioning Data Analytics of Md Simulations Toward the Exascale Era

Michela Taufer, University of Delaware, USA

### 9:50-10:10 Towards Petascale Genome Analytics

Pavan Balaji, Argonne National Laboratory, USA

### 10:15-10:35 Using Workflows to Automate Big Data Processing

Ewa Deelman, University of Southern California, USA

### 10:40-11:00 Experiences with In-Situ Analytics at Extreme Scale

Manish Parashar, Rutgers University, USA

Saturday, March 10

## MS95

### Matrix Computations with Applications - Part I of II

9:25 AM-11:05 AM

Room: 52-104

#### For Part 2 see MS106

Matrix computations are ubiquitous and important in scientific computing. Thus, the development of efficient matrix computation algorithms is still a very active area of research. This minisymposium seeks to bring together interdisciplinary researchers, share ideas and present research on the state of the art in all the fields of computational science and engineering. In this minisymposium, eight contributed presentations will discuss the topics on algorithms, implementations and applications for solving linear and nonlinear equations, nonnegative matrix factorization, (inverse) eigenvalue problems by exploiting modern computer architectures.

Organizer: Akira Imakura

University of Tsukuba, Japan

Organizer: Lei Du

Dalian University of Technology, China

#### 9:25-9:45 A Nonlinear Semi-NMF Based Method and Its Parallel Implementation for Deep Neural Networks

Akira Imakura, Ryosuke Arai, and Tetsuya Sakurai, University of Tsukuba, Japan

#### 9:50-10:10 Fast Solution of Nonnegative Matrix Factorization Via a Matrix-Based Active Set Method

Ning Zheng, Ken Hayami, and Nobutaka Ono, National Institute of Informatics, Japan

#### 10:15-10:35 A New High Performance and Scalable SVD Algorithm on Distributed Memory Systems

Shengguo Li, Jie Liu, and Xinbiao Gan, National University of Defense Technology, China

#### 10:40-11:00 Matrix Multiplication Based Algorithm for Inverse Eigenvalue Problems and Its Quadratic Convergence

Kensuke Aishima, University of Tokyo, Japan

Saturday, March 10

## MS96

### Highly Scalable Solvers for Computational PDEs - Part I of III

9:25 AM-11:05 AM

Room: 52-201

#### For Part 2 see MS107

Efficient numerical simulation of high-fidelity large-scale problems in science and engineering is still challenging. Complex applications include difficulties such as transient problems with widely varying time and spatial scales, strongly-coupled multiphysics, heterogenous media, nonlinearities, etc. Developing efficient linear system solvers for these problems has many challenges, especially for large-scale simulations, which makes proper preconditioning critical. This minisymposium will focus on highly scalable preconditioners, e.g. multigrid or domain decomposition approaches, multiphysics solvers (e.g. based on block factorization techniques), nonlinear preconditioning, multiscale solvers for heterogenous problems or space-time solvers. Contributions discussing algorithms that can exploit many-core processors and accelerators are also welcomed.

Organizer: Paul Lin

Sandia National Laboratories, USA

Organizer: Santiago Badia

Universitat Politecnica de Catalunya, Spain

Organizer: Hiroshi Kawai

Tokyo University, Japan

Organizer: Masao Ogino

Nagoya University, Japan

Organizer: Ryuji Shioya

Tokyo University, Japan

#### 9:25-9:45 Towards Large Scale Unfitted Adaptive Finite Element Simulations

Santiago Badia, Universitat Politecnica de Catalunya, Spain; Alberto Martin, International Center for Numerical Methods in Engineering, Spain; Francesc Verdugo, CIMNE, Spain

#### 9:50-10:10 New Aspects on Nonlinear Domain Decomposition Solvers on the Extreme Scale

Axel Klawonn and Martin Lanser, Universitaet zu Koeln, Germany; Oliver Rheinbach, Technische Universität Bergakademie Freiberg, Germany

#### 10:15-10:35 Large-Scale Nonlinear Solid Mechanics Analysis Using Domain Decomposition Method Based on Nonlinear Solver

Yasunori Yusa and Hiroshi Okada, Tokyo University of Science, Japan; Tomonori Yamada and Shinobu Yoshimura, University of Tokyo, Japan

#### 10:40-11:00 Performance Tuning of DDM Subdomain Local FE Solvers for Many-Core Architectures

Hiroshi Kawai, Tokyo University, Japan; Masao Ogino, Nagoya University, Japan; Ryuji Shioya, Tokyo University, Japan; Tomonori Yamada and Shinobu Yoshimura, University of Tokyo, Japan



Saturday, March 10

## MS97

### Emerging Architectural Support for Scientific Kernels - Part I of II

9:25 AM-11:05 AM

Room:52-202

#### For Part 2 see MS108

There has been a consensus that no single type of processor is able to be both fast and energy-efficient for any algorithm. As a result, today's hardware architectures grow ever more complicated by composing various heterogeneous compute units, and adding or reducing memory hierarchies for different purposes. This brings opportunities and challenges to a wide range of scientific kernels. In this proposed minisymposium, we will discuss actual impact of a variety of emerging architectures and hardware, such as on-package high bandwidth memory, memory-side processing, FPGA, SW many-core processor and CPU-GPU heterogeneous processor, on a wide range of scientific kernels from compute-bound dense problems to memory-bound sparse problems as well as FFT and stencil computations in the middle. The objective is to explore how emerging architectural support can help next-generation scientific algorithm design.

Organizer: Guangming Tan  
*Chinese Academy of Sciences, China*

Organizer: Weifeng Liu  
*Norwegian University of Science and Technology, Norway*

#### 9:25-9:45 Exploring and Analyzing the Real Impact of Modern On-Package Memory on HPC Scientific Kernels

Shuaiwen Song, Pacific Northwest  
National Laboratory, USA

#### 9:50-10:10 Accelerating Scientific Kernels on Energy-Efficient FPGAs

Yun (Eric) Liang, Peking University,  
China

#### 10:15-10:35 Sparse Tensor Decomposition on EMU Platform

Srinivas Eswar, Jiajia Li, Richard Vuduc,  
Patrick Lavin, and Young Jeffrey,  
Georgia Institute of Technology, USA

#### 10:40-11:00 Design and Implementation of Deep Learning Kernel Library on APU

Haoqiang Guo, Chinese Academy of  
Sciences, China

Saturday, March 10

## MS98

### Innovative Methods for High Performance Iterative Solvers - Part II of II

9:25 AM-11:05 AM

Room:52-204

#### For Part 1 see MS87

Iterative methods are commonly used in HPC applications to solve large size linear systems, eigenvalue or least squares problems. To be able to take advantage of the latest generations of processors and to be scalable on future Exascale platforms, such iterative methods must be adapted or sometimes rethought. The speakers of this mini symposium describe recent advances in iterative linear solvers including communication-avoiding Krylov subspace methods, optimized implementations for SpMV and preconditioners, but also new challenges in domain decomposition methods and fault tolerance.

Organizer: Marc Baboulin  
*University of Paris-Sud, France*

Organizer: Takeshi Fukaya  
*Hokkaido University, Japan*

Organizer: Takeshi Iwashita  
*Hokkaido University, Japan*

#### 9:25-9:45 An Approach to Accelerating the SpMV Kernel by Exploiting Specific Sparse Structures

Takeshi Fukaya, Koki Ishida, Akie  
Miura, and Takeshi Iwashita,  
Hokkaido University, Japan; Hiroshi  
Nakashima, Kyoto University, Japan

#### 9:50-10:10 Resilience for Fine-Grained Preconditioning Methods

Masha Sosonkina and Evan Coleman, Old  
Dominion University, USA

#### 10:15-10:35 Adaptive Multipreconditioned Domain Decomposition Methods

Nicole Spillane, Centre de Mathématiques  
et Informatique, France; Christophe  
Bovet, CNRS, Université Paris-  
Saclay, France; Augustin Parret-  
Fréaud, Safran Tech, France; Pierre  
Gosselet, LMT-Cachan, France

#### 10:40-11:00 Solving the $k$ -th Eigenvalue Problem in Large-Scale Electronic Structure Calculations

Dongjin Lee, Nagoya University, Japan;  
Takeo Hoshi, Tottori University,  
Japan; Tomohiro Sogabe, Yuto  
Miyatake, and Shao-Liang Zhang,  
Nagoya University, Japan

*continued in next column*

Saturday, March 10

## MS99

### Scientific Accomplishments on Massively Parallel Computing of Earth and Planetary Sciences - Part II of II

9:25 AM-11:05 AM

Room:52-301

#### For Part 1 see MS88

With greatly progressing on technologies for both numerical scheme and computational capability, many fundamental issues in solid Earth geosciences have been tackled with massively parallel computing, which is known as generating plate tectonics in mantle convection, secular variations of geomagnetic field in geodynamo and early warning of occurrence of large-scale earthquakes. In this minisymposium, we introduce important scientific accomplishments using massively parallel computing on seismic wave propagations, mantle convection and geo- and stellar-dynamo simulations and first principle computations of physical properties of Earth's material, which have been applied for fundamental topics, but still unresolved, in Earth and Planetary Sciences. This mini symposium is tied up with 'Computational Infrastructure of Geodynamics' running in the Geoscience community in the United States, which is successfully organized for providing 'community code' on various fundamental topics on Earth and Planetary Sciences with massively parallel computing, computational geodynamics community in Europe, and collaborated with applied mathematics and computational sciences, and with FLAGSHIP 2020 project, MEXT within CBSM2 Project "Structure and Properties of Materials in deep Earth and planets".

Organizer: Takashi Nakagawa  
*Japan Agency for Marine-Earth Science and Technology, Japan*

Organizer: Hiroaki Matsui  
*University of California, Davis, USA*

Organizer: Koichiro Umemoto  
*Tokyo Institute of Technology, Japan*

#### 9:25-9:45 Large Scale Seismic Simulations with SW4

*Anders Petersson, Bjorn Sjogreen, and Arthur Rodgers, Lawrence Livermore National Laboratory, USA*

#### 9:50-10:10 Global Simulations of Mantle Convection and Plate Tectonics on Massively Parallel Computers using StagYY

*Paul J. Tackley, ETH Zürich, Switzerland; Patrick Sanan, Università della Svizzera italiana, Switzerland*

#### 10:15-10:35 Solar Convection Simulation with Reduced Speed of Sound Technique

*Hideyuki Hotta, Chiba University, Japan*

#### 10:40-11:00 First-Principles Investigation of Thermal Transport Property of Earth's Lowermost Mantle

*Haruhiko Dekura, Ehime University, Japan*

Saturday, March 10

## MS100

### Modeling and Simulation of HPC Architectures and Applications - Part I of III

9:25 AM-11:05 AM

Room:52-302

#### For Part 2 see MS111

High performance computing architectures and applications are significantly complex that pen and paper models do not suffice for system design and performance prediction. A wide variety of modeling and simulation tools have been developed in the high performance computing community to study architectural designs (CPU, memory, GPGPUs and networks among other things), and to predict the performance of parallel codes on future architectures. In this multi-part minisymposium, we bring together researchers from the on-node and off-node modeling/simulation communities, discuss recent advances to simulation tools of varying levels of fidelity, and present recent validation and simulation studies performed using these tools.

Organizer: Abhinav Bhatele

*Lawrence Livermore National Laboratory, USA*

Organizer: Rob Ross

*Argonne National Laboratory, USA*

Organizer: Jeremiah Wilke

*Sandia National Laboratories, USA*

#### 9:25-9:45 Flexible and Scalable Applications Models for the Structural Simulation Toolkit (SST)

*Jeremiah Wilke, Sandia National Laboratories, USA*

#### 9:50-10:10 Workload Simulations with TraceR-CODES

*Nikhil Jain, Lawrence Berkeley National Laboratory, USA; Misbah Mubarak, Argonne National Laboratory, USA; Abhinav Bhatele, Lawrence Livermore National Laboratory, USA; Christopher Carothers, Rensselaer Polytechnic Institute, USA; Rob Ross, Argonne National Laboratory, USA*

Saturday, March 10

## MS100

### Modeling and Simulation of HPC Architectures and Applications - Part I of III

9:25 AM-11:05 AM

continued

#### 10:15-10:35 Simulating MPI

##### Applications: the SMPI Approach

*Frédéric Suter*, CNRS, France; *Arnaud Legrand*, Inria Grenoble, France; *Martin Quinson*, ENS, France

#### 10:40-11:00 Implementation

##### and Evaluation of NSIM-ACE: An Interconnect Simulator for Remote Direct Memory Access

*Ryutaro Susukita* and *Yoshiyuki Morie*, Kyushu University, Japan; *Takeshi Nanri* and *Hidetomo Shibamura*, JST CREST, Japan

Saturday, March 10

## MS101

### HPC Benchmarking ~Past, Present, and Future~

9:25 AM-11:05 AM

Room: 52-303

Toward true hardware/software co-design of future HPC (High-Performance Computing) systems, this minisymposium focuses on performance evaluation of HPC systems through benchmarking. Since HPC systems/architectures and applications executed on them have rapidly diversified, it is getting hard to reproduce performance behaviors of all real scientific applications using only existing benchmark programs. Therefore, it is necessary to reconsider the metric for evaluating HPC systems. Besides, the future HPC system is required to achieve a certain level of sustained performance with limited power and hardware resources; it is essential to have benchmark programs capable of reproducing the sustained performance of real applications. To this end, cooperation between computational and computer scientists is imperative. In this minisymposium, we first look back on research activities related to HPC benchmarking. Especially the advantages and limitations of HPL and HPCG currently used complementarily are discussed. Based on these arguments, we will introduce research activities on the latest benchmark development such as HPGMG, which will fill the gap between the both benchmarks. Then, future direction for benchmarking of HPC system will be presented by performance tuner/benchmarkers. Based on these discussions, this workshop will provide an opportunity for both computational and computer scientists to discuss prospects for the future HPC benchmarking.

Organizer: Ryusuke Egawa

*Tohoku University, Japan*

Organizer: Jose Gracia

*High Performance Computing Center  
Stuttgart, Germany*

#### 9:25-9:45 HPC Benchmarking: Perspective from Power and Sustained Performance

*Ryusuke Egawa*, Tohoku University, Japan

#### 9:50-10:10 Comparing Signatures of Common Supercomputer Benchmarks and Their Utility for the Future

*Erich Strohmaier*, Lawrence Berkeley National Laboratory, USA

#### 10:15-10:35 Expectation to Supercomputer Benchmarks from the Viewpoint of Performance Optimization

*Kazuo Minami*, RIKEN, Japan

#### 10:40-11:00 HPC Benchmarking: Problem Size and Other Pitfalls

*Jose Gracia* and *Vladimir Marjanovic*, High Performance Computing Center Stuttgart, Germany

Saturday, March 10

## MS102

### Large-Scale Simulation in Geodynamics - Part I of II

9:25 AM-11:05 AM

Room:52-304

#### For Part 2 see MS113

This minisymposium focuses on numerical methods and software for large-scale parallel simulations of geodynamical processes. Due to the enormous spatial and time scales of geodynamic processes, studying these processes requires sophisticated computer simulations. To maintain good parallel efficiency of the corresponding implementations on future exascale systems, a performance driven co-design is necessary, involving a systematic complexity analysis of mathematical methods and the design of physics-aware approaches. The minisymposium will also show the utilization of highly optimized and parallel methods in geophysical applications. We will bring together experts of different disciplines to discuss scalable computational methods for large-scale geophysical simulations while highlighting recent advances.

Organizer: Dominik Bartuschat  
*University of Erlangen-Nuernberg, Germany*

Organizer: Ulrich J. Ruede  
*University of Erlangen-Nuernberg, Germany*

Organizer: Omar Ghattas  
*University of Texas at Austin, USA*

#### 9:25-9:45 Simulation Studies on Mantle Dynamics of Terrestrial Planets: Theoretical Backgrounds, Tools and Outcrops

*Masanori Kameyama, Ehime University, Japan*

#### 9:50-10:10 Thermal Inversion in Subduction Zones

*Matthew G. Knepley and Margarete Jadamec, University of Buffalo, USA; Tobin Isaac, Georgia Institute of Technology, USA*

#### 10:15-10:35 An Extreme-Scale Implicit Solver for Highly Nonlinear and Heterogeneous Flow in Earth's Mantle

*Johann Rudi, University of Texas at Austin, USA; Georg Stadler, Courant Institute of Mathematical Sciences, New York University, USA; Omar Ghattas, University of Texas at Austin, USA*

#### 10:40-11:00 TerraNeo - A Finite Element Multigrid Framework for Extreme-Scale Earth Mantle Convection Simulations

*Dominik Bartuschat and Ulrich J. Ruede, University of Erlangen-Nuernberg, Germany*

Saturday, March 10

## MS103

### High Performance Computing in Optimization - Part I of III

9:25 AM-11:05 AM

Room:57-201

#### For Part 2 see MS114

High performance computing (HPC) becomes increasingly relevant in solving large-scale optimization problems in many applications areas such as operations research, energy systems, industrial engineering, advanced manufacturing, and others. Despite being in its infancy, HPC optimization was successfully used to solve problems of unprecedented sizes. The mainstream computational approaches are parallelization of linear algebra, optimization decomposition algorithms and branch-and-bound searches. This minisymposium on HPC optimization will host a series of presentations on parallel methods and software implementations and is aimed at fostering interdisciplinary collaborations between the optimization and parallel processing communities.

Organizer: Katsuki Fujisawa  
*Kyushu University, Japan*

Organizer: Yuji Shinano  
*Zuse Institute Berlin, Germany*

Organizer: Olaf Schenk  
*Università della Svizzera italiana, Switzerland*

Organizer: Kibaek Kim  
*Argonne National Laboratory, USA*

Organizer: Cosmin G. Petra  
*Lawrence Livermore National Laboratory, USA*

#### 9:25-9:45 Solving Steiner Tree Problems and Variants to Optimality on a Supercomputer

*Thorsten Koch, Daniel Rehfeldt, and Yuji Shinano, Zuse Institute Berlin, Germany*



Saturday, March 10

## MS103

### High Performance Computing in Optimization - Part I of III

9:25 AM-11:05 AM

continued

#### 9:50-10:10 Parallel Computing in a Hierarchical Mixed-Integer Linear Programming Method for Optimal Design of Distributed Energy Systems

*Ryohei Yokoyama*, Osaka Prefecture University, Japan; *Yuji Shinano*, Zuse Institute Berlin, Germany

#### 10:15-10:35 Security Constrained Optimization of Large Scale Energy Systems on High Performance Computers

*Juraj Kardos*, Drosos Kourounis, and *Olaf Schenk*, Università della Svizzera italiana, Switzerland

#### 10:40-11:00 Towards the New Generation of Multiperiod Optimal Power Flow Solvers

*Drosos Kourounis* and *Juraj Kardos*, Università della Svizzera italiana, Switzerland; *Alexander Fuchs*, Swiss Federal Institute of Technology, Switzerland; *Olaf Schenk*, Università della Svizzera italiana, Switzerland

Saturday, March 10

## MS104

### Parallel-in-Time Integration Methods for High-Performance Computing - Part I of III

9:25 AM-11:05 AM

*Room: 57-202*

#### For Part 2 see MS115

With million-way concurrency at hand, the efficient use of modern HPC systems has become one of the key challenges in computational science and engineering. New mathematical concepts are needed to fully exploit massively parallel architectures.

For the numerical solution of time-dependent PDEs, recent developments in the field of time-parallel methods have opened new ways of overcoming both strong and weak scaling limits. This minisymposium brings together scientists from the fields of parallel-in-time integration, multigrid methods, applications and more. Talks will cover a broad range of topics, from implementation and performance analysis to convergence acceleration and novel algorithms.

Organizer: *Mikio Iizuka*

*Kyushu University, Japan*

Organizer: *Kenji Ono*

*Kyushu University, Japan*

Organizer: *Daniel Ruprecht*

*University of Leeds, United Kingdom*

Organizer: *Robert Speck*

*Jülich Supercomputing Centre, Germany*

#### 9:25-9:45 On Implementations of Parallel-in-Time Integration Methods

*Ruth Schöbel*, Jülich Aachen Research Alliance, Forschungszentrum Jülich, Germany; *Robert Speck*, Jülich Supercomputing Centre, Germany

#### 9:50-10:10 Exploring Intrusiveness of Parallel-in-Time Methods

*Stephanie Friedhoff*, University of Wuppertal, Germany; *Robert D. Falgout* and *Jacob B. Schroder*, Lawrence Livermore National Laboratory, USA

#### 10:15-10:35 Performance Analysis of the PinT Computation of the Large-Scale Simulation for the Diffusion Equation

*Yoshiharu Ohi*, RIKEN, Japan

#### 10:40-11:00 Communication-Aware Adaptive Parareal (CAAP) with Application to a Hyperbolic PDE System

*Allan Nielsen* and *Jan S. Hesthaven*, École Polytechnique Fédérale de Lausanne, Switzerland

Saturday, March 10

**CP12****Extreme-Scale Applications**

9:25 AM-11:05 AM

Room:52-101

*Chair: Hatem Ltaief, King Abdul Aziz University, Saudi Arabia***9:25-9:45 FMM-Accelerated Wave Scattering Solver on Emerging Many and Multi-Core High-Performance Computing Architectures***Mustafa Abduljabbar, Noha Al-Harathi, Mohammed A. Al Farhan, Rui Chen, and Huda Ibeid, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; Rio Yokota, Tokyo Institute of Technology, Japan; Hakan Bagci, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; David E. Keyes, King Abdul Aziz University, Saudi Arabia***9:50-10:10 Parallel 3d Ffts for Scientific Application Codes on Exascale Architectures***Andrew M. Canning, Lawrence Berkeley National Laboratory, USA***10:15-10:35 Application of Parallel Hierarchical Matrices in Spatial Statistics and Parameter Identification***Alexander Litvinenko, King Abdullah University of Science & Technology (KAUST), Saudi Arabia***10:40-11:00 Large-Scale Parallel Finite-Element Analysis of Seismic Response Considering Soil-Structure Interaction***Yoshiyuki Takahashi, Kajima Technical Research Institute, Japan; Naoki Morita, University of Tokyo, Japan; Toshihide Saka, Kajima Technical Research Institute, Japan; Gaku Hashimoto and Hiroshi Okuda, University of Tokyo, Japan; Toshihiro Koiso and Kazuhiko Yamada, Kajima Technical Research Institute, Japan*

Saturday, March 10

**CP13****Simulations and Scientific Workflows**

9:25 AM-11:05 AM

Room:52-102

*Chair: Carol S. Woodward, Lawrence Livermore National Laboratory, USA***9:25-9:45 Event Detection in Multi-Variate Scientific Simulations Using Feature Anomaly Metrics***Aditya Konduri, Hemanth Kolla, Julia Ling, Philip Kegelmeyer, Timothy Shead, Daniel M. Dunlavy, and Warren Davis, Sandia National Laboratories, USA***9:50-10:10 The Journey from On-Premises to Serverless in Scientific Computing***Germán Moltó, Miguel Caballer, Carlos de Alfonso, Alfonso Pérez, Amanda Calatrava, and Ignacio Blanquer, Universitat Politècnica de Valencia, Spain***10:15-10:35 BIND: Platform for Integrating Tasks of Bigdata Application in Cloud***Yoonsu Shin and songkuk kim, Yonsei University, South Korea***10:40-11:00 Tuple Space Implementation in a Parallel Workflow Engine, Opendiel***Zachary Trzil, University of Tennessee, USA; Kwai L. Wong, University of Tennessee and Oak Ridge National Laboratory, USA***Coffee Break**

11:05 AM-11:35 AM

Room:63-1F



Saturday, March 10

**MS105****Modern Techniques to Bridge Flexibility and Performance for Parallel Numerical Codes**

11:35 AM-1:15 PM

Room:52-103

The complexity of numerical simulations keeps growing, both at the mathematical level with advanced techniques like adaptivity, high-order discretizations and scalable, robust solvers, and at the implementation level, where we are confronted with challenges like massive (multi-level) parallelism, accelerators and fault tolerance. In this environment, sustainable software development increasingly relies on frameworks that abstract away hardware details, provide composable components with clear interfaces and use techniques like domain specific languages and code generation to enable a separation of concerns and to combine expert knowledge from different fields of science. We bring together scientists from the fields of numerical methods, compiler architecture and scientific computing to discuss current challenges at different levels of the simulation stack and their influence on software development and this type of software framework.

*Organizer: Steffen Müthing  
Heidelberg University, Germany**Organizer: Christian Engwer  
University of Münster, Germany***11:35-11:55 Efficient Matrix-Free Discontinuous Galerkin Assembly: To Hand-Write Or to Generate?***Steffen Müthing, Peter Bastian, René Heß, Dominic Kempf, and Marian Piatkowski, Heidelberg University, Germany*

Saturday, March 10

## MS105

### Modern Techniques to Bridge Flexibility and Performance for Parallel Numerical Codes

11:35 AM-1:15 PM

continued

#### 12:00-12:20 Achieving Portable Performance Across Architectures with PETSc

*Karl Rupp*, Freelance Computational Scientist, Austria; *Richard T. Mills*, Intel Corporation, USA; *Hong Zhang*, Hong Zhang, and *Barry F. Smith*, Argonne National Laboratory, USA

#### 12:25-12:45 A DSL for Hybridization and Static Condensation Techniques in Firedrake

*Thomas H. Gibson*, Imperial College, United Kingdom

#### 12:50-1:10 Optimizing Auto-Generating Kernel-Code for Low Order FEM on Locally Structured Meshes

*Marcel Koch*, University of Münster, Germany

Saturday, March 10

## MS106

### Matrix Computations with Applications - Part II of II

11:35 AM-1:15 PM

Room:52-104

#### For Part 1 see MS95

Matrix computations are ubiquitous and important in scientific computing. Thus, the development of efficient matrix computation algorithms is still a very active area of research. This minisymposium brings together interdisciplinary researchers in order for the researchers to share a variety of studies on the state of the art in all the fields of computational science and engineering. In this minisymposium, eight contributed presentations will discuss the topics on algorithms, implementations and applications for solving linear and nonlinear equations, nonnegative matrix factorization, (inverse) eigenvalue problems by exploiting modern computer architectures.

Organizer: *Akira Imakura*

*University of Tsukuba, Japan*

Organizer: *Lei Du*

*Dalian University of Technology, China*

#### 11:35-11:55 A Parallel Algorithm for Tridiagonal Quasi-Toeplitz Linear Systems

*Jintao Zhang* and *Lei Du*, Dalian University of Technology, China

#### 12:00-12:20 Block GMREs Methods for Linear Systems with Multiple Right-Hand Sides

*Yanfei Jing*, University of Electronic Science and Technology of China, China

#### 12:25-12:45 Solving a Newton Equation on the Stiefel Manifold with Matrix-Free Krylov Subspace Methods

*Kensuke Aihara*, Tokyo City University, Japan; *Hiroyuki Sato*, Kyoto University, Japan

#### 12:50-1:10 Parallel Solution for Multi-Group Radiation Transfer Equations

*Hengbin An*, Institute of Applied Physics and Computational Mathematics, China

Saturday, March 10

## MS107

### Highly Scalable Solvers for Computational PDEs - Part II of III

11:35 AM-1:15 PM

Room:52-201

#### For Part 1 see MS96

#### For Part 3 see MS118

Efficient numerical simulation of high-fidelity large-scale problems in science and engineering is still challenging. Complex applications include difficulties such as transient problems with widely varying time and spatial scales, strongly-coupled multiphysics, heterogeneous media, nonlinearities, etc. Developing efficient linear system solvers for these problems has many challenges, especially for large-scale simulations, which makes proper preconditioning critical. This minisymposium will focus on highly scalable preconditioners, e.g. multigrid or domain decomposition approaches, multiphysics solvers (e.g. based on block factorization techniques), nonlinear preconditioning, multiscale solvers for heterogeneous problems or space-time solvers. Contributions discussing algorithms that can exploit many-core processors and accelerators are also welcomed.

Organizer: *Santiago Badia*

*Universitat Politecnica de Catalunya, Spain*

Organizer: *Hiroshi Kawai*

*Tokyo University, Japan*

Organizer: *Paul Lin*

*Sandia National Laboratories, USA*

Organizer: *Masao Ogino*

*Nagoya University, Japan*

Organizer: *Ryuji Shioya*

*Tokyo University, Japan*

#### 11:35-11:55 Matrix-Free Parallel Multigrid for Fe Systems with a Trillion Unknowns

*Ulrich J. Ruede*, *Dominik Thoennes*, and *Nils Kohl*, University of Erlangen-Nuernberg, Germany

continued on next page

**12:00-12:20 An Implicit, Multiscale, Conservative Hybrid Kinetic-Ion/Fluid-Electron Algorithm for Plasma Simulation**

*Luis Chacón* and Adam Stanier, Los Alamos National Laboratory, USA

**12:25-12:45 A Partitioned Coupling Analysis of FEM and SPH for Large-Scale Fluid-Structure Interaction Problems**

*Masao Ogino* and Takuya Iwama, Nagoya University, Japan; Mitsuteru Asai, Kyushu University, Japan

**12:50-1:10 Leveraging Application Structure Within Next Generation Multigrid Solvers**

*Raymond Tuminaro*, Luc Berger-Vergiat, and Matthias Mayr, Sandia National Laboratories, USA

Saturday, March 10

## MS108

### Emerging Architectural Support for Scientific Kernels - Part II of II

11:35 AM-1:15 PM

Room:52-202

**For Part 1 see MS97**

There has been a consensus that no single type of processor is able to be both fast and energy-efficient for any algorithm. As a result, today's hardware architectures grow ever more complicated by composing various heterogeneous compute units, and adding or reducing memory hierarchies for different purposes. This brings opportunities and challenges to a wide range of scientific kernels. In this proposed minisymposium, we will discuss actual impact of a variety of emerging architectures and hardware, such as on-package high bandwidth memory, memory-side processing, FPGA, SW many-core processor and CPU-GPU heterogeneous processor, on a wide range of scientific kernels from compute-bound dense problems to memory-bound sparse problems as well as FFT and stencil computations in the middle. The objective is to explore how emerging architectural support can help next-generation scientific algorithm design.

Organizer: Guangming Tan  
*Chinese Academy of Sciences, China*

Organizer: Weifeng Liu  
*Norwegian University of Science and Technology, Norway*

**11:35-11:55 Refactoring Several Scientific Kernels on SW26010 Many-core Processor**

*Wei Xue*, Tsinghua University, P. R. China

**12:00-12:20 A Novel Parallel SPGEMM Algorithm on Sunway TaihuLight**

*Xin He*, Chinese Academy of Sciences, China

**12:25-12:45 Accelerating Irregular Applications on Heterogeneous Computer Systems**

*Shuai Che*, AMD Research, USA

**12:50-1:10 When Sparse Matrices Meet Heterogeneous Processors: Opportunities and Challenges**

*Weifeng Liu*, Norwegian University of Science and Technology, Norway

Saturday, March 10

## MS109

### Techniques for Developing Massively-Parallel Linear Solvers - Part I of II

11:35 AM-1:15 PM

Room:52-204

**For Part 2 see MS120**

This minisymposium highlights four techniques to address the challenges of developing linear solvers at extreme scale. These techniques are (1) hierarchical low-rank algorithms (e.g., the fast multipole method) (2) asynchronous-communication based algorithms (3) multi-level algorithms (e.g., domain decomposition and multi-grid methods) and (4) task-based parallel algorithms. Both scalable algorithms and approaches that can exploit many-core and massively parallel architectures will be discussed.

Organizer: Yingzhou Li  
*Duke University, USA*

Organizer: Yingzhou Li  
*Duke University, USA*

Organizer: Chao Chen  
*Stanford University, USA*

**11:35-11:55 A Distributed-memory Hierarchical Solver for General Sparse Linear Systems**

*Chao Chen*, Stanford University, USA; Siva Rajamanickam and Erik G. Boman, Sandia National Laboratories, USA; Eric F. Darve, Stanford University, USA

**12:00-12:20 ACHILES : An Asynchronous Iterative Sparse Linear Solver**

*Paritosh P. Ramanan* and Edmond Chow, Georgia Institute of Technology, USA; Siva Rajamanickam and Erik G. Boman, Sandia National Laboratories, USA

**12:25-12:45 Multilevel Iterative Methods and Their Application in Field-Scale Petroleum Reservoir Simulation**

*Chensong Zhang*, Chinese Academy of Sciences, China

**12:50-1:10 Tacho: Task Parallel Sparse Direct Cholesky Factorization with Bounded Memory Constraint**

*Kyungjoo Kim*, Siva Rajamanickam, and H. Carter Edwards, Sandia National Laboratories, USA



Saturday, March 10

## MS110

### FPGAs in High Performance Computing (Part I of II)

11:35 AM-1:15 PM

Room:52-301

#### For Part 2 see MS121

Field-Programmable Gate-Array (FPGAs) belong to the class of fine-grained reconfigurable architecture. Historically incepted to simulate ASIC designs, FPGAs have today transcended their initial purpose and are being considered to accelerate general-purpose computations. Modern FPGAs are equipped with millions of reconfigurable cells, thousands of hard compute-blocks and large amount of on-chip memory; characteristics that allows FPGAs to compete with general purpose processors (CPUs) and graphics processors (GPUs). With Dennard scaling failing and with the threats of dark-silicon standing around the corner, FPGAs offer a unique way of tuning the hardware resources to the application, leaving no silicon unused. This property allows FPGAs to potentially be more power-efficient and still perform on par with CPUs and GPUs. This minisymposium is divided into two sessions, covering important parts of current and future FPGAs and their role in High-Performance Computing. The first session focuses on FPGA systems and includes concepts such as in-switch computing and multi-FPGA architectures. The second session focus on how to leverage FPGAs to accelerate HPC applications.

Organizer: Artur M. Podobas  
*Tokyo Institute of Technology, Japan*

Organizer: Toshihiro Hanawa  
*University of Tokyo, Japan*

Organizer: Taisuke Boku  
*University of Tsukuba, Japan*

Organizer: Kentaro Sano  
*Tohoku University, Japan*

11:35-11:55 **Data-Flow HPC with Custom Hardware on FPGA Cluster**  
*Kentaro Sano, Tohoku University, Japan*

12:00-12:20 **Scalable Inter-FPGA Direct Communication for Parallel FPGA Applications**

*Ryohei Kobayashi, University of Tsukuba, Japan*

12:25-12:45 **Design Experience on Intel HARP2 Platform using OpenCL**  
*Toshihiro Hanawa, University of Tokyo, Japan; Taisuke Boku, University of Tsukuba, Japan*

12:50-1:10 **Towards Compute Everywhere: Computing in the Network with FPGA-Accelerated Clouds and Clusters**  
*Martin Herbordt, Boston University, USA*

Saturday, March 10

## MS111

### Modeling and Simulation of HPC Architectures and Applications - Part II of III

11:35 AM-1:15 PM

Room:52-302

#### For Part 1 see MS100

#### For Part 3 see MS122

High performance computing architectures and applications are significantly complex that pen and paper models do not suffice for system design and performance prediction. A wide variety of modeling and simulation tools have been developed in the high performance computing community to study architectural designs (CPU, memory, GPGPUs and networks among other things), and to predict the performance of parallel codes on future architectures. In this multi-part minisymposium, we bring together researchers from the on-node and off-node modeling/simulation communities, discuss recent advances to simulation tools of varying levels of fidelity, and present recent validation and simulation studies performed using these tools.

Organizer: Abhinav Bhatele  
*Lawrence Livermore National Laboratory, USA*

Organizer: Rob Ross  
*Argonne National Laboratory, USA*

Organizer: Jeremiah Wilke  
*Sandia National Laboratories, USA*

11:35-11:55 **Exploring Extreme Heterogeneity with Analytical Modeling**

*Shirley Moore and Jeffrey S. Vetter, Oak Ridge National Laboratory, USA*

12:00-12:20 **Unifying Software Models with Hardware Emulation**

*Farzad Fatollahi-Fard, David Donofrio, George Michelogiannakis, and John Shalf, Lawrence Berkeley National Laboratory, USA*

12:25-12:45 **Modeling Massively Multithreaded GPUs with GPGPU-Sim**  
*Tim Rogers, Purdue University, USA*

12:50-1:10 **Analyzing Exascale Memory Architectures Using the SST Toolkit**

*Clayton Hughes, Simon D. Hammond, and Arun Rodrigues, Sandia National Laboratories, USA*

Saturday, March 10

**MS112****Parallel Phase-Field Simulation**

11:35 AM-1:15 PM

*Room: 52-303*

Phase-field method has been emerged as a powerful numerical tool to simulate complicated material microstructures. Meanwhile, one drawback of the phase-field method is a large computational cost due to the diffuse interface model. Therefore, the high performance computing (HPC) in the phase-field simulation is required. In this minisymposium, we will give a presentation regarding the state-of-the-art HPC techniques in the phase-field simulation and discuss deeply the further development in the future.

Organizer: Tomohiro Takaki

*Kyoto Institute of Technology, Japan*

Organizer: Akinori Yamanaka

*Tokyo University of Agriculture and Technology, Japan***11:35-11:55 Parallel GPU Phase-Field Simulations of Dendrite Competitive Growth**

Tomohiro Takaki, Kyoto Institute of Technology, Japan; Mune kazu Ohno, Hokkaido University, Japan; Yasushi Shibuta, University of Tokyo, Japan; Shinji Sakane, Kyoto Institute of Technology, Japan; Takashi Shimokawabe, University of Tokyo, Japan; Takayuki Aoki, Tokyo Institute of Technology, Japan

**12:00-12:20 A Parallel-Adaptive Mesh Refinement Numerical Approach for Generalized Phase Field Simulations**

Zipeng Guo, Tsinghua University, P. R. China

**12:25-12:45 An Optimized and Parallelized Phase-Field Solver for Sintering Using the Pace3d Framework**

Johannes Hötzer, Hochschule Karlsruhe Technik und Wirtschaft, Germany

**12:50-1:10 Multiple-Gpu Computing of Phase-Field Crystal Simulation of Grain Growth**

Akinori Yamanaka, Tokyo University of Agriculture and Technology, Japan

Saturday, March 10

**MS113****Large-Scale Simulation in Geodynamics - Part II of II**

11:35 AM-1:15 PM

*Room: 52-304***For Part 1 see MS102**

This minisymposium focuses on numerical methods and software for large-scale parallel simulations of geodynamical processes. Due to the enormous spatial and time scales of geodynamic processes, studying these processes requires sophisticated computer simulations. To maintain good parallel efficiency of the corresponding implementations on future exascale systems, a performance driven co-design is necessary, involving a systematic complexity analysis of mathematical methods and the design of physics-aware approaches. The minisymposium will also show the utilization of highly optimized and parallel methods in geophysical applications. We will bring together experts of different disciplines to discuss scalable computational methods for large-scale geophysical simulations while highlighting recent advances.

Organizer: Dominik Bartuschat

*University of Erlangen-Nuernberg, Germany*

Organizer: Ulrich J. Ruede

*University of Erlangen-Nuernberg, Germany*

Organizer: Omar Ghattas

*University of Texas at Austin, USA***11:35-11:55 Coupling Field and Particle Methods for Massively Parallel Simulations of the Earth Mantle**

Wolfgang Bangerth and Rene Gassmoeller, Colorado State University, USA

**12:00-12:20 A Stencil Scaling Approach for Accelerating Matrix-Free Finite Element Implementations**

Daniel Drzisga, Simon Bauer, and Marcus Mohr, Technical University of Munich, Germany; Ulrich J. Ruede, University of Erlangen-Nuernberg, Germany; Christian Waluga and Barbara Wohlmuth, Technical University of Munich, Germany

**12:25-12:45 Efficient, Matrix-based Approaches to Solving Landscape Evolution Models in Parallel**

Louis Moresi, University of Melbourne, Australia

**12:50-1:10 Hybrid Operators and Composable Software within Lithospheric Dynamic Simulation**

Patrick Sanan, Università della Svizzera italiana, Switzerland; Dave A. May, University of Oxford, United Kingdom; Karl Rupp, Freelance Computational Scientist, Austria; Guillaume Duclaux, University of Bergen, Norway

Saturday, March 10

## MS114

### High Performance Computing in Optimization - Part II of II

11:35 AM-1:15 PM

Room:57-201

**For Part 1 see MS103**

**For Part 3 see MS125**

High performance computing (HPC) becomes increasingly relevant in solving large-scale optimization problems in many applications areas such as operations research, energy systems, industrial engineering, advanced manufacturing, and others. Despite being in its infancy, HPC optimization was successfully used to solve problems of unprecedented sizes. The mainstream computational approaches are parallelization of linear algebra, optimization decomposition algorithms and branch-and-bound searches. This minisymposium on HPC optimization will host a series of presentations on parallel methods and software implementations and is aimed at fostering interdisciplinary collaborations between the optimization and parallel processing communities.

Organizer: Katsuki Fujisawa  
*Kyushu University, Japan*

Organizer: Yuji Shinano  
*Zuse Institute Berlin, Germany*

Organizer: Olaf Schenk  
*Università della Svizzera italiana, Switzerland*

Organizer: Kibaek Kim  
*Argonne National Laboratory, USA*

Organizer: Cosmin G. Petra  
*Lawrence Livermore National Laboratory, USA*

**11:35-11:55 Optimization Problems Arising in Programming Environments Based on Data-Driven Tasking**  
*Utz-Uwe Haus, Cray, Inc., USA*

### 12:00-12:20 Realizing Extremely Large-Scale Scientific Applications Using Deeper Memory Hierarchy

*Toshio Endo, Tokyo Institute of Technology, Japan*

### 12:25-12:45 Toward Multiperiod AC Security Constrained Power Flow Optimization at Exascale

*Michel Schanen and Mihai Anitescu, Argonne National Laboratory, USA; Cosmin G. Petra, Lawrence Livermore National Laboratory, USA*

### 12:50-1:10 Distributed Solving of Large-scale LPs and MIPs from Energy System Models

*Daniel Rehfeldt and Yuji Shinano, Zuse Institute Berlin, Germany*

Saturday, March 10

## MS115

### Parallel-in-Time Integration Methods for High-Performance Computing - Part II of III

11:35 AM-1:15 PM

Room:57-202

**For Part 1 see MS104**

**For Part 3 see MS126**

With million-way concurrency at hand, the efficient use of modern HPC systems has become one of the key challenges in computational science and engineering. New mathematical concepts are needed to fully exploit massively parallel architectures. For the numerical solution of time-dependent PDEs, recent developments in the field of time-parallel methods have opened new ways of overcoming both strong and weak scaling limits. This minisymposium brings together scientists from the fields of parallel-in-time integration, multigrid methods, applications and more. Talks will cover a broad range of topics, from implementation and performance analysis to convergence acceleration and novel algorithms.

Organizer: Mikio Iizuka  
*Kyushu University, Japan*

Organizer: Kenji Ono  
*Kyushu University, Japan*

Organizer: Daniel Ruprecht  
*University of Leeds, United Kingdom*

Organizer: Robert Speck  
*Jülich Supercomputing Centre, Germany*

### 11:35-11:55 Convergence Acceleration of PinT Integration of Advection Equation Using Accurate Phase Calculation Method

*Mikio Iizuka and Kenji Ono, Kyushu University, Japan*

*continued on next page*

*continued in next column*

**12:00-12:20 Towards Parallel in Time Methods for Numerical Weather Prediction**

*Jemma Shipton*, Imperial College, United Kingdom

**12:25-12:45 Accelerating Simulations of Fast Ion Trajectories in a Fusion Reactor**

*Daniel Ruprecht*, University of Leeds, United Kingdom; *Rob Akers*, Culham Centre For Fusion Energy, United Kingdom; *Debasmita Samaddar*, UK Atomic Energy Authority, United Kingdom; *Krasymyr Tretiak*, Kharkiv Institute of Physics and Technology, Ukraine

**12:50-1:10 Preliminary Application of Parareal for Two Phase Fluid Flow Simulation in Geological Media with Tough2 Code**

*Mitsuhiro Miyagi* and *Hajime Yamamoto*, Taisei Corporation, Japan; *Toshiya Takami*, Oita University, Japan; *Mikio Iizuka* and *Kenji Ono*, Kyushu University, Japan; *Kengo Nakajima*, University of Tokyo, Japan

Saturday, March 10

**CP14****Advanced Simulation - Part I of II**

11:35 AM-1:15 PM

Room:52-102

*Chair: Harald Koestler*, University of Erlangen-Nuernberg, Germany

**11:35-11:55 Scaling Machine Learning Models to Optimize Parallel Applications Using Feature Space Partitioning**

*Murali Emani*, Lawrence Livermore National Laboratory, USA

**12:00-12:20 Tensor Decomposition to Perform Change of Basis in Multi-Variate Hpc Data to Preserve Higher Order Statistical Moments**

*Hemanth Kolla*, *Aditya Konduri*, *Prashant Rai*, *Tammy Kolda*, and *Warren Davis*, Sandia National Laboratories, USA

**12:25-12:45 Assessing and Improving the Numerical Solution of Atmospheric Physics in E3SM**

*David J. Gardner* and *Carol S. Woodward*, Lawrence Livermore National Laboratory, USA; *Hui Wan*, Pacific Northwest National Laboratory, USA; *Chris Vogl*, Lawrence Livermore National Laboratory, USA; *Philip Rasch*, Pacific Northwest National Laboratory, USA

**12:50-1:10 Towards a General Purpose Multirate Time Integration Software Suite for Multiphysics Problems**

*John Loffeld*, *Slaven Peles*, and *Carol S. Woodward*, Lawrence Livermore National Laboratory, USA

**Lunch Break**

1:15 PM-2:30 PM

Attendees On Their Own

Saturday, March 10

**MS116****A Survey of Applications Using Tasking Runtimes**

2:30 PM-4:10 PM

Room:52-103

The need to solve complex science and engineering problems on rapidly evolving HPC systems poses a considerable challenge for modern simulation frameworks. It is likely that fundamentally new approaches are needed in order to succeed. Solutions must address issues such as scalability, performance portability, and programmability. Asynchronous, task-parallel programming models show great promise in addressing these issues. In this minisymposium, we will look at several application codes that are using tasking runtimes for performing large-scale scientific simulations, followed by a discussion of their strengths and weaknesses.

Organizer: *Irina Demeshko*  
Los Alamos National Laboratory, USA

Organizer: *Ben Bergen*  
Los Alamos National Laboratory, USA

**2:30-2:50 Flecsale: A Task-Based, Multi-Dimensional Eulerian-Lagrangian Solver Built with the Flexible Computational Science Infrastructure (FleCSI)**

*Irina Demeshko* and *Ben Bergen*, Los Alamos National Laboratory, USA

**2:55-3:15 Enzo-P / Cello: A Charm++ Adaptive Mesh Refinement Astrophysics Application**

*James Bordner*, University of California, San Diego, USA

**3:20-3:40 Harnessing {HPX} to Run Billions of Tasks for a Scalable Portable Hydrodynamics Simulation of the Merger of Two Stars**

*Hartmut Kaiser*, Louisiana State University, USA

**3:45-4:05 Amt Runtime Scalability and Portability of Large-Scale Parallel Applications Using Lessons from the Uintah Project**

*Martin Berzins* and *Alan Humphrey*, University of Utah, USA; *Dan Sunderland*, Sandia National Laboratories, USA; *Damodar Sahasbarude*, University of Utah, USA; *Zhang Yang*, IAPM Beijing, China



Saturday, March 10

## MS117

### HPC Applications and Algorithms

2:30 PM-4:10 PM

Room:52-104

The development of numerical methods for the solution of partial differential equations relies on both mathematical and algorithmic technique and innovation. When computational efficiency at large scale is the target, the algorithmic aspect becomes critical to realize the performance metrics derived from theory. In this minisymposium, we discuss parallel solvers that are compatible with adaptive mesh refinement and demonstrate their role in enabling a wide range of numerical applications.

Organizer: Frederic Gibou

University of California, Santa Barbara, USA

Organizer: Carsten Burstedde

Universität Bonn, Germany

### 2:30-2:50 Free Boundary Solvers on a Forest of Octrees

Frederic Gibou and Arthur Guittet, University of California, Santa Barbara, USA; Mohammad Mirzadeh, Massachusetts Institute of Technology, USA; Carsten Burstedde, Universität Bonn, Germany

### 2:55-3:15 Simulation of Volcanic Ash Transport Using ForestClaw

Donna Calhoun, Boise State University, USA; Melody Shih, Columbia University, USA; Hans Schweiger, Larry Mastin, and Roger Denlinger, U.S. Geological Survey Cascades Volcano Observatory, USA

### 3:20-3:40 Applications of Space-Time Adaptive Mesh Refinement with a Block-Structured Higher-Order Cut Cell Framework

Hans Johansen, Amneet Pal S. Bhalla, Dan Graves, and Phillip Colella, Lawrence Berkeley National Laboratory, USA

### 3:45-4:05 Multigrid Methods for Complex Fracture Networks in Porous Media

Marco Favino, Università della Svizzera italiana, Switzerland; Jürg Hunziker and Klaus Holliger, University of Lausanne, Switzerland; Rolf Krause, Università della Svizzera italiana, Switzerland

Saturday, March 10

## MS118

### Highly Scalable Solvers for Computational PDEs - Part III of III

2:30 PM-4:10 PM

Room:52-201

### For Part 2 see MS107

Efficient numerical simulation of high-fidelity large-scale problems in science and engineering is still challenging. Complex applications include difficulties such as transient problems with widely varying time and spatial scales, strongly-coupled multiphysics, heterogeneous media, nonlinearities, etc. Developing efficient linear system solvers for these problems has many challenges, especially for large-scale simulations, which makes proper preconditioning critical. This minisymposium will focus on highly scalable preconditioners, e.g. multigrid or domain decomposition approaches, multiphysics solvers (e.g. based on block factorization techniques), nonlinear preconditioning, multiscale solvers for heterogeneous problems or space-time solvers. Contributions discussing algorithms that can exploit many-core processors and accelerators are also welcomed.

Organizer: Santiago Badia

Universitat Politècnica de Catalunya, Spain

Organizer: Hiroshi Kawai

Tokyo University, Japan

Organizer: Paul Lin

Sandia National Laboratories, USA

Organizer: Masao Ogino

Nagoya University, Japan

Organizer: Ryuji Shioya

Tokyo University, Japan

**2:30-2:50 Performance Comparison of GMG and BDD for Large Scale Numerical Simulation**

Ryuji Shioya, Hongjie Zheng, and Hiroshi Kawai, Tokyo University, Japan; Amane Takei, University of Miyazaki, Japan

**2:55-3:15 A Reduced Iterative Domain Decomposition Method for Mixed Variational Formulations Derived from Magnetic Field Problems**

Daisuke Tagami, Kyushu University, Japan; Shin-ichiro Sugimoto, Tokyo University of Science, Suwa, Japan

**3:20-3:40 A Nonlinearly Preconditioned Newton Method for a Hyperelasticity Problem**

Shihua Gong, Peking University, China; Xiao-Chuan Cai, University of Colorado Boulder, USA

**3:45-4:05 Scalable Newton-Krylov-Amg-Based Preconditioners for Implicit Continuum Plasma Simulations**

Paul Lin, John N. Shadid, Edward G. Phillips, Eric C. Cyr, and Roger Pawlowski, Sandia National Laboratories, USA

Saturday, March 10

**MS119**

**Research Compiler Infrastructure for High Performance Computing**

2:30 PM-4:10 PM

Room:52-202

Current supercomputers accommodate multiple levels of parallelism, such as core, instruction, and SIMD-level parallelism to achieve high operation ratio per watt. The role of compiler and programming model is becoming critical to exploit potential performance of the system. The minisymposium, Research Compiler Infrastructure for High Performance Computing, will introduce research activities and their compiler infrastructures to provide productive and high-performance programming environment for exascale computing. It consists of 4 talks provided from research groups in Japan, USA, and UK. RIKEN AICS in Japan will introduce the Omni compiler infrastructure which is a source-to-source translator. Research activities on parallel programming models including the PGAS model and OpenMP will be presented. Stony Brook University in USA will introduce research activities based on the open-source LLVM compiler. The talk includes SIMD vectorization, memory optimization for deep memory and NUMA architecture. Oak Ridge National Laboratory in USA will present the OpenARC compiler infrastructure, their source-to-source translator. The talk includes research activities on OpenACC compilation and optimization technique for FPGA. ARM in UK will present their compiler research for the ARM architecture, which is based on LLVM. Compiler development for SIMD vectorization for their new SIMD instruction set, Scalable Vector Extension (SVE) will be presented.

Organizer: Jinpil Lee

RIKEN Advanced Institute for Computational Science, Japan

**2:30-2:50 Omni Compiler: Compiler Infrastructure for Source-to-Source Translation**

Jinpil Lee and Mitsuhsa Sato, RIKEN Advanced Institute for Computational Science, Japan

**2:55-3:15 Directive-based Vectorization Strategies in LLVM for Current and Upcoming Many-Core Processors**

Martin Kong and Lingda Li, Brookhaven National Laboratory, USA; Barbara Chapman, Brookhaven National Laboratory and Stony Brook University, USA

**3:20-3:40 OpenARC: A Compiler Framework for Directive-based High-Performance Reconfigurable Computing**

Seyong Lee, Oak Ridge National Laboratory, USA

**3:45-4:05 Arm in HPC: Compilers and Tools**

Francesco Petrogalli, Arm Holdings, United Kingdom

Saturday, March 10

## MS120

### Techniques for Developing Massively-Parallel Linear Solvers - Part II of II

2:30 PM-4:10 PM

Room:52-204

#### For Part 1 see MS109

This minisymposium highlights four techniques to address the challenges of developing linear solvers at extreme scale. These techniques are (1) asynchronous-communication based algorithms (2) multi-level algorithms (e.g., domain decomposition and multi-grid methods) (3) hierarchical low-rank algorithms (e.g., the fast multipole method) and (4) task-based parallel algorithms. Both scalable algorithms and approaches that can exploit many-core and massively parallel architectures will be discussed.

Organizer: Yingzhou Li

Duke University, USA

Organizer: Yingzhou Li

Duke University, USA

Organizer: Chao Chen

Stanford University, USA

#### 2:30-2:50 HiCMA: Hierarchical Computations on Manycore Architectures

David E. Keyes and Hatem Ltaief, King Abdul Aziz University, Saudi Arabia; George Turkiyyah, American University of Beirut, Lebanon; Rio Yokota, Tokyo Institute of Technology, Japan

#### 2:55-3:15 Parallel Preconditioning Techniques for Electromagnetic Problems

Tao Cui, Chinese Academy of Sciences, China

#### 3:20-3:40 Efficient Low-rank Solver for Integral Equations on Distributed Memory Systems

Akihiro Ida, University of Tokyo, Japan

#### 3:45-4:05 AutoMG: An Automatic Tuning Multigrid Procedure for the Sequences of Linear Systems

Xiaowen Xu, Ran Xu, and Hengbin An, Institute of Applied Physics and Computational Mathematics, China; Zeyao Mo, CAEP Software Center for High Performance Numerical Simulations, China

Saturday, March 10

## MS121

### FPGAs in High Performance Computing (Part II of II)

2:30 PM-4:10 PM

Room:52-301

#### For Part 1 see MS110

Field-Programmable Gate-Array (FPGAs) belong to the class of fine-grained reconfigurable architecture. Historically incepted to simulate ASIC designs, FPGAs have today transcended their initial purpose and are being considered to accelerate general-purpose computations. Modern FPGAs are equipped with millions of reconfigurable cells, thousands of hard compute-blocks and large amount of on-chip memory; characteristics that allows FPGAs to compete with general purpose processors (CPUs) and graphics processors (GPUs). With Dennard scaling failing and with the threats of dark-silicon standing around the corner, FPGAs offer a unique way of tuning the hardware resources to the application, leaving no silicon unused. This property allows FPGAs to potentially be more power-efficient and still perform on par with CPUs and GPUs. This minisymposium is divided into two sessions, covering important parts of current and future FPGAs and their role in High-Performance Computing. The first session focuses on FPGA systems and includes concepts such as in-switch computing and multi-FPGA architectures. The second session focus on how to leverage FPGAs to accelerate HPC applications.

Organizer: Artur M. Podobas

Tokyo Institute of Technology, Japan

Organizer: Toshihiro Hanawa

University of Tokyo, Japan

Organizer: Taisuke Boku

University of Tsukuba, Japan

Organizer: Kentaro Sano

Tohoku University, Japan

#### 2:30-2:50 Accelerating Stencil Computations Using Opencl on FPGAs

Hamid Reza Zohouri and Artur M. Podobas, Tokyo Institute of Technology, Japan

#### 2:55-3:15 Using OmpSs to Run Applications in FPGAs

Xavier Martorell, Barcelona Supercomputing Center, Spain

#### 3:20-3:40 System Software for Future Heterogeneous, Reconfigurable Architecture

Kazutomo Yoshii and Franck Cappello, Argonne National Laboratory, USA

#### 3:45-4:05 Design Examples of Accelerator-in-Switch

Hideharu Amano, Keio University, Japan

continued in next column

Saturday, March 10

## MS122

### Modeling and Simulation of HPC Architectures and Applications - Part III of III

2:30 PM-4:10 PM

Room:52-302

#### For Part 2 see MS111

High performance computing architectures and applications are significantly complex that pen and paper models do not suffice for system design and performance prediction. A wide variety of modeling and simulation tools have been developed in the high performance computing community to study architectural designs (CPU, memory, GPGPUs and networks among other things), and to predict the performance of parallel codes on future architectures. In this multi-part minisymposium, we bring together researchers from the on-node and off-node modeling/simulation communities, discuss recent advances to simulation tools of varying levels of fidelity, and present recent validation and simulation studies performed using these tools.

Organizer: Abhinav Bhatele  
*Lawrence Livermore National Laboratory, USA*

Organizer: Rob Ross  
*Argonne National Laboratory, USA*

Organizer: Jeremiah Wilke  
*Sandia National Laboratories, USA*

#### 2:30-2:50 Network Design and Comparison Studies Using Tracer Simulations

Abhinav Bhatele, Lawrence Livermore National Laboratory, USA; Nikhil Jain, Lawrence Berkeley National Laboratory, USA; Todd Gamblin, Lawrence Livermore National Laboratory, USA; Laxmikant Kale, University of Illinois at Urbana-Champaign, USA

#### 2:55-3:15 Exploring HPC Storage and Interconnect Design Space Using CODES

Misbah Mubarak and Rob Ross, Argonne National Laboratory, USA; Christopher Carothers, Rensselaer Polytechnic Institute, USA; Philip Carns and Matthieu Dorier, Argonne National Laboratory, USA; Nikhil Jain, Lawrence Berkeley National Laboratory, USA; Abhinav Bhatele, Lawrence Livermore National Laboratory, USA; Noah Wolfe, Rensselaer Polytechnic Institute, USA

#### 3:20-3:40 Extending A Network Simulator for Power/Performance Prediction of Large Scale Interconnection Networks

Takatsugu Ono, Kyushu Sangyo University, Japan; Yuta Kakibuka, Kyushu University, Japan; Nikhil Jain, Lawrence Berkeley National Laboratory, USA; Abhinav Bhatele, Lawrence Livermore National Laboratory, USA; Shinobu Miwa, University of Electro-Communications, Japan; Koji Inoue, Kyushu University, Japan

#### 3:45-4:05 Software Defined Networking Applied to High Performance Computing

Michael Lang, Los Alamos National Laboratory, USA; Xin Yuan, Florida State University, USA

Saturday, March 10

## MS123

### Advances in Parallel Multiphysics and Multiscale Computing

2:30 PM-4:10 PM

Room:52-303

The increase in computational power and the diversity of software have advanced capability of solving large and complex multiphysics and multiscale applications, such as interdependent networks, gas, electric, water distribution and other unstructured networks. Here we present the on-going research, development and applications. We focus on new algorithmic developments, coupling techniques, as well as software implementations and their deployment in emerging computer systems.

Organizer: Hong Zhang  
*Argonne National Laboratory, USA*

Organizer: Shirang G. Abhyankar  
*Argonne National Laboratory, USA*

#### 2:30-2:50 Simulation of Coupled Power-Grid and Water Networks on Extreme Scale Computers

Hong Zhang and Shirang G. Abhyankar, Argonne National Laboratory, USA

#### 2:55-3:15 Development of An Architecture-Aware Highly Scalable Software Package for Solving PDEs

Yulong Ao and Chao Yang, Chinese Academy of Sciences, China

#### 3:20-3:40 Solving Large Graphs in Parallel with Multigrid

Tristan Konolige and Jed Brown, University of Colorado Boulder, USA

#### 3:45-4:05 High-Performance Multiphysics Simulations for Ion Channels

Yu-Hsiang Lan, National Taiwan University, Taiwan; MiSun Min, Argonne National Laboratory, USA; Paul Fischer, University of Illinois at Urbana-Champaign, USA; Tai-Chia Lin, National Taiwan University, Taiwan, ROC



Saturday, March 10

## MS124

### Performance in Particle Simulations: Parallelism, Prediction, Auto-Tuning

2:30 PM-4:10 PM

Room:52-304

Particle simulations are used in a wide range of applications including molecular dynamics, astrophysics and fluid dynamics. This minisymposium addresses current research on computational performance in particle simulations with particular regard to parallelization and auto-tuning on recent HPC hardware. Load balancing is a major research topic and subject of this minisymposium. Due to heterogeneity of hardware and due to potentially great variations in particle numbers per thread/process (heterogeneous particle systems), appropriate load balancing algorithms such as tree-based domain partitioning and implementations thereof are required. Besides, with various (and quite different) hardware systems being driven towards the exascale, the automatic exploration and exploitation of the performance of these systems with unified as well as with hardware-aware particle simulation code is highly relevant to mitigate manual tedious code optimization. Therefore, the minisymposium will further address performance prediction for and automatic tuning of particle simulations. The speakers will give insight into how load balancing, performance prediction or auto-tuning are used and further developed in their implementations which are given amongst others by the simulation codes *ls1 mardyn*, which is a molecular dynamics package used in process engineering, and *FDPS*, a framework for hybrid large-scale parallelization of particle-based simulations.

Organizer: Philipp Neumann  
*University of Hamburg, Germany*

Organizer: Felix Wolf  
*Technische Universitaet Darmstadt, Germany*

**2:30-2:50 Parallelism and Performance Prediction in Molecular Dynamics and Molecular-Continuum Simulations**

*Philipp Neumann, University of Hamburg, Germany*

**2:55-3:15 How Particle Methods Can Help Simulate the Structural Plasticity of the Brain at Scale**

Sebastian Rinke, Technische Universitaet Darmstadt, Germany; Markus Butz-Ostendorf, Biomax Informatics AG, Germany; Marc-Andre Hermanns and Mikael Naveau, Juelich Aachen Research Alliance, Forschungszentrum Juelich, Germany; *Felix Wolf*, Technische Universitaet Darmstadt, Germany

**3:20-3:40 Performance of FDPs on Sunway Taihulight and Other Modern HPC Systems**

*Jun Makino*, Kobe University, Japan; Masaki Iwasawa, Daisuke Namekata, Keigo Nitadori, Miyuki Tsubouchi, and Long Wang, RIKEN, Japan

**3:45-4:05 Fully Heterogeneous Load Balancing in *ls1 mardyn***

*Steffen Seckler*, Simon Griebel, and Nikola P. Tchipev, Technical University of Munich, Germany; Philipp Neumann, University of Hamburg, Germany; Hans-Joachim Bungartz, Technical University of Munich, Germany

Saturday, March 10

## MS125

### High Performance Computing in Optimization - Part III of III

2:30 PM-4:10 PM

Room:63-201/202

**For Part 2 see MS114**

High performance computing (HPC) becomes increasingly relevant in solving large-scale optimization problems in many applications areas such as operations research, energy systems, industrial engineering, advanced manufacturing, and others. Despite being in its infancy, HPC optimization was successfully used to solve problems of unprecedented sizes. The mainstream computational approaches are parallelization of linear algebra, optimization decomposition algorithms and branch-and-bound searches. This minisymposium on HPC optimization will host a series of presentations on parallel methods and software implementations and is aimed at fostering interdisciplinary collaborations between the optimization and parallel processing communities.

Organizer: Katsuki Fujisawa  
*Kyushu University, Japan*

Organizer: Yuji Shinano  
*Zuse Institute Berlin, Germany*

Organizer: Olaf Schenk  
*Università della Svizzera italiana, Switzerland*

Organizer: Kibaek Kim  
*Argonne National Laboratory, USA*

Organizer: Cosmin G. Petra  
*Lawrence Livermore National Laboratory, USA*

**2:30-2:50 Parallel Quasi-Newton Methods for the Optimization of Complex Systems Using High-Performance Computing**

*Cosmin G. Petra*, Lawrence Livermore National Laboratory, USA

**2:55-3:15 Asynchronous Dual Decomposition for Stochastic Mixed-Integer Programs**

*Kibaek Kim*, Argonne National Laboratory, USA; *Cosmin G. Petra*, Lawrence Livermore National Laboratory, USA; *Victor M. Zavala*, University of Wisconsin, Madison, USA

**3:20-3:40 Assessing Performance of Parallel Branch-and-Bound Algorithms**

*Ted Ralphs*, Lehigh University, USA; *Stephen Maher*, Lancaster University, United Kingdom; *Yuji Shinanno*, Zuse Institute Berlin, Germany

**3:45-4:05 Scalable and Provably Convergent Asynchronous Progressive Hedging for Scenario-Based Decomposition of Stochastic Convex Programs**

*Jean-Paul Watson*, Sandia National Laboratories, USA

Saturday, March 10

## MS126

**Parallel-in-Time Integration Methods for High-Performance Computing - Part III of III**

2:30 PM-4:10 PM

Room: 57-202

**For Part 2 see MS115**

With million-way concurrency at hand, the efficient use of modern HPC systems has become one of the key challenges in computational science and engineering. New mathematical concepts are needed to fully exploit massively parallel architectures. For the numerical solution of time-dependent PDEs, recent developments in the field of time-parallel methods have opened new ways of overcoming both strong and weak scaling limits. This minisymposium brings together scientists from the fields of parallel-in-time integration, multigrid methods, applications and more. Talks will cover a broad range of topics, from implementation and performance analysis to convergence acceleration and novel algorithms.

Organizer: *Mikio Iizuka*

*Kyushu University, Japan*

Organizer: *Kenji Ono*

*Kyushu University, Japan*

Organizer: *Daniel Ruprecht*

*University of Leeds, United Kingdom*

Organizer: *Robert Speck*

*Jülich Supercomputing Centre, Germany*

**2:30-2:50 PinT Computation of Swarm Behavior**

*Toshiya Takami*, Oita University, Japan

**2:55-3:15 A Spectral Coarse Solver for the Parallel-in-time Solution of the Time-periodic Eddy Current Problem**

*Iryna Kulchytska-Ruchka*, *Herbert De Gersem*, and *Sebastian Schoeps*, Technische Universität Darmstadt, Germany

**3:20-3:40 Parallel Multigrid Reduction in Time (MGRIT) with Application to Power Grid Simulations**

*Robert D. Falgout*, Lawrence Livermore National Laboratory, USA; *Matthieu Lecouvez*, CEA, France; *Jacob B. Schroder* and *Carol S. Woodward*, Lawrence Livermore National Laboratory, USA

**3:45-4:05 A Space-Time Multigrid Method for Electrophysiology**

*Patrick Zulian*, *Pietro Benedusi*, *Alessio Quaglino*, and *Rolf Krause*, Università della Svizzera italiana, Switzerland

Saturday, March 10

## CP15

### Advanced Simulation - Part II of II

2:30 PM-4:10 PM

Room:52-102

Chair: Marco Favino, Università della Svizzera italiana, Switzerland

#### 2:30-2:50 Recent Advances on Finite Element Method and Simulation of Invisibility Cloaks with Metamaterials

Jichun Li, University of Nevada, Las Vegas, USA

#### 2:55-3:15 Lessons and Accomplishments of the Computational Science Undergraduate Research Experiences Program

Kwai L. Wong, University of Tennessee  
and Oak Ridge National Laboratory,  
USA

### Intermission

4:10 PM-4:20 PM

Saturday, March 10

## IP7

### Communication Hiding Through Pipelining in Krylov Method

4:20 PM-5:05 PM

Room:63-201/202

Chair: Olaf Schenk, Università della Svizzera italiana, Switzerland

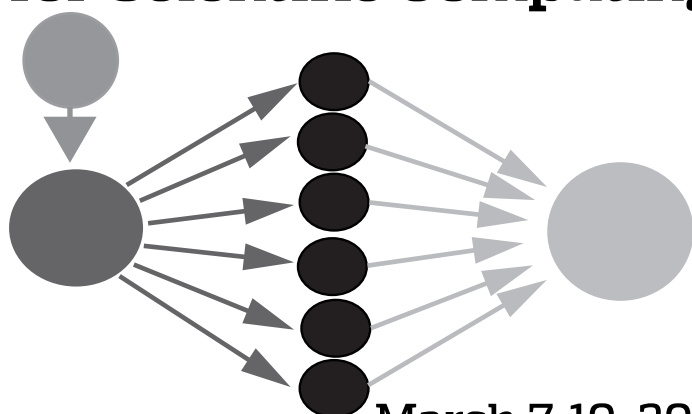
The performance of Krylov method suffers in the strong scaling limit due to the synchronization latencies mainly associated with the dot-products. We review the development of pipelined Krylov methods that exploit asynchronous communication to overlap communication and computation. For example, by reordering the operations in the conjugate gradients it is possible to execute the dot-product simultaneously with the sparse matrix vector products. This gives the dot-product more time to complete while doing useful work. We also introduce deep pipelines, where the dot-products are overlapping with multiple sparse matrix-vector products. This gives the dot-product even more time to complete and these deep pipelines solve most of the scaling problems. However, due to this reorganization of the algorithm the propagation of rounding errors is altered resulting in different final accuracy, requiring a detailed numerical analysis. We conclude with experiments that illustrate the improved scalability.

Wim I. Vanroose

University of Antwerp, Belgium

**PP18 Abstracts**

SIAM Conference on  
**Parallel Processing  
for Scientific Computing**



March 7-10, 2018  
Waseda University  
Tokyo, Japan

Abstracts are printed as submitted by the authors.



**IP1****Verified Numerical Computation and Large Scale Computing**

In this talk, we overview the states of art of the verified numerical computation. One of the stresses is on large scale computation. We first treat computer assisted rigorous inclusion of eigenvalues of the self-adjoint operator on bounded domain. Especially, we report our recent work on how to calculate rigorous lower bounds of eigenvalues. Especially, we summarize how to get a rigorous lower bound of minimum eigenvalue. Then, we briefly review its application to computer assisted proof of nonlinear elliptic boundary value problems through the Newton-Kantorovich type arguments. Namely, inclusion of minimum eigenvalue is useful for bounding the operator norm of inverses of linear operators. In the process of inclusion of eigenvalues, usually large-scale computation is needed. We will show an example in which, using the Kei-computer, several eigenvalues are rigorously included for a square matrix with more than 1 million-dimension.

Shin'ichi Oishi

Waseda University  
oishi@waseda.jp

**IP2****Computing Beyond Moores Law**

Moores Law is a techno-economic model that has enabled the Information Technology (IT) industry to nearly double the performance and functionality of digital electronics roughly every two years within a fixed cost, power and area. Within a decade, the technological underpinnings for the process Gordon Moore described will come to an end as lithography gets down to atomic scale. At that point, it will be feasible to create lithographically produced devices with characteristic dimensions in the 3nm5nm range. This range corresponds to a dozen or fewer Si atoms across critical device features and will therefore be a practical limit for controlling charge in a classical sense. The classical technological driver that has underpinned Moores law for the past 50 years is already failing and is anticipated to flatten by 2025. This talk provides an updated view of what a 2021-2023 system might look like and the challenges ahead, based on our most recent understanding of technology roadmaps. It also will discuss the tapering of historical improvements in lithography, and how it affects options available to continue scaling of successors to the first exascale machine.

John Shalf

Lawrence Berkeley National Laboratory  
jshalf@lbl.gov

**IP3****Toward Community Software Ecosystems for High-Performance Computational Science**

Software—crosscutting technology that connects advances in mathematics, computer science, and domain-specific science and engineering—is a cornerstone of long-term collaboration and scientific progress. As we leverage unprecedented high-performance computing resources to work toward predictive science, software complexity is increasing due to multiphysics and multiscale modeling, the coupling of simulations and data analytics, and the demand for greater reproducibility in the midst of disruptive architectural changes. Applications increasingly require the com-

bined use of independent software packages, which have diverse sponsors, priorities, and processes for development and release. These challenges create the unique opportunity to improve how scientific software is designed, developed, and sustained—with explicit work toward scientific software ecosystems. This presentation will introduce the xSDK, or Extreme-scale Scientific Software Development Kit, where community-defined policies are increasing the quality and interoperability across numerical libraries as needed by the DOE Exascale Computing Project. We will also discuss complementary efforts to increase scientific software productivity and sustainability.

Lois Curfman McInnes

Mathematics and Computer Science Division  
Argonne National Laboratory  
curfman@mcs.anl.gov

**IP4****Enabling Multi-Peta-Scale Fully Implicit Simulations of Atmospheric Dynamics on Sunway TaihuLight**

This joint talk is comprised of two parts. In the first part, we will introduce the technical details and the underlying design philosophy of Sunway TaihuLight, which is the world's first system with a peak performance greater than 100 PFlops, and a parallel scale of over 10 million cores. In contrast with other existing heterogeneous supercomputers, which include both CPU processors and PCIe-connected many-core accelerators (NVIDIA GPU or Intel MIC), the computing power of TaihuLight is provided by a homegrown many-core SW26010 CPU that includes two types of processing elements in one chip. With a comparison between other leadership systems, we would highlight the major programming challenges for utilizing the system for extreme-scale scientific simulations. In the second part of the talk, we will present some technical details of our recent work that won the ACM Gordon Bell Prize in 2016. In the work, we have designed an ultra-scalable fully implicit solver for solving the atmospheric dynamics at the nonhydrostatic scale. The fully implicit solver can scale to the entire system of Sunway TaihuLight and sustain an aggregate performance of 8 PFlops. Remarkably, the simulation capability of the fully implicit solver at the full system scale is nearly two orders of magnitude better than a highly optimized explicit solver with an aggregate performance of 23.66 PFlops.

Haohuan Fu

Tsinghua University  
National Supercomputing Center in Wuxi  
haohuan@tsinghua.edu.cn

Chao Yang

Software Institute, Chinese Academy of Sciences  
P. R. China  
yangchao@iscas.ac.cn

**IP5****Fast Scalable Implicit Solver for Low-Ordered Unstructured Finite Element Analysis and Its Application**

We developed a fast scalable implicit solver that uses low-order unstructured finite element (FE) analysis to estimate static/dynamic and nonlinear/linear responses of a solid body. As this type of FE analysis uses sparse matrix-vector multiplication with random data access, it does not

generally perform well on modern computer architectures. We overcame this bottleneck by developing special algorithms that are better suited to current computer architectures. These algorithms have many characteristics, including multi-grid methods, adaptive conjugate gradient methods, multi-precision arithmetic, element-by-element methods, predictors based on Adams-Bashforth methods and grid-structuring methods. Our solver performs well in terms of scalability, time-to-solution and peak performance. The superiority of this solver on the K computer led to it being selected as a finalist for the SC14 and SC15 Gordon Bell Prize, and for the SC16 Best Poster Award. In this talk, we describe the solver in detail with reference to its applications to earthquake problems, which we were able to analyze using nonlinear wave FE analysis with a trillion degrees of freedom.

Tsuyoshi Ichimura  
Earthquake Research Institute  
The University of Tokyo  
ichimura@eri.u-tokyo.ac.jp

## IP6

### ChainerMN: Scalable Distributed Deep Learning with Chainer

We'll present ChainerMN, a multi-node distributed deep learning framework, together with the basics of distributed deep learning. Even though GPUs are continuously gaining more computation throughput, it is still very time-consuming to train state-of-the-art deep neural network models. For better scalability and productivity, it is paramount to accelerate the training process by using multiple GPUs. To enable high-performance and flexible distributed training, we developed ChainerMN, built on top of Chainer. We'll first introduce the basic approaches to distributed deep learning. Then, we'll explain the design choice, basic usage, and implementation details of Chainer and ChainerMN. We'll report benchmark results using hundreds of GPUs and discuss the future directions of distributed deep learning.

Takuya Akiba  
Preferred Networks, Inc.  
akiba@preferred.jp

## IP7

### Communication Hiding Through Pipelining in Krylov Method

The performance of Krylov method suffers in the strong scaling limit due to the synchronization latencies mainly associated with the dot-products. We review the development of pipelined Krylov methods that exploit asynchronous communication to overlap communication and computation. For example, by reordering the operations in the conjugate gradients it is possible to execute the dot-product simultaneously with the sparse matrix vector products. This gives the dot-product more time to complete while doing useful work. We also introduce deep pipelines, where the dot-products are overlapping with multiple sparse matrix-vector products. This gives the dot-product even more time to complete and these deep pipelines solve most of the scaling problems. However, due to this reorganization of the algorithm the propagation of rounding errors is altered resulting in different final accuracy, requiring a detailed numerical analysis. We conclude

with experiments that illustrate the improved scalability.

Wim I. Vanroose  
Universiteit Antwerpen  
Belgium  
wim.vanroose@ua.ac.be

## CP1

### Fast Scalable Solvers for Charge Models in Molecular Dynamics Applications

Polarizable and reactive force fields which incorporate dynamic charges into their molecular models have been shown to address the shortcomings of classical force fields with static charges. A crucial component of polarizable and reactive force fields is a charge distribution model. Accurate solution of the charge distribution problem, as formulated by several different models proposed to date, manifests itself as large-scale sparse linear systems. Krylov subspace methods employed to solve these systems represent a significant limitation in terms of solution time and scalability in large simulations. In this talk, we will discuss our recent efforts to accelerate these parallel iterative solvers. These discussions will span the Electronegativity Equilibration Method (EEM) and Charge Equilibration (QEq) model which are relatively old approaches, as well as newer models such as ACKS2 and PQEq that have been proposed to address the shortcomings of EEM and QEq. In particular, we will present results from using various methods such as incomplete LU (ILU) based preconditioning, sparse approximate inverse (SAI) preconditioning and Chebyshev filtering. We will discuss the trade-offs involved in choosing a good preconditioner and its efficient implementation on shared and distributed memory architectures.

H. Metin Aktulga  
Michigan State University  
hma@cse.msu.edu

## CP1

### A Scalable Block-Iterative (Non)-Linear Solver for Adaptive 3D Organic Semiconductor Simulation

We present a 3D adaptive code for large scale simulation of complex organic semiconductor devices in stationary, transient and frequency domain regimes on Octree meshes. We address Issues related to error estimation and mesh adaptation and present an application-tailored block preconditioning strategy for solving the (non)-linear coupled systems of equations derived from the model discretization.

Carlo De Falco  
MOX Modeling and Scientific Computing  
Politecnico di Milano  
carlo.defalco@polimi.it

## CP1

### Leveraging Kokkos/Tpetra for Performance Portability in the Thyra Abstraction Layer

The Thyra package in Trilinos provides a set of interfaces relating to vectors, vector spaces, and linear operators to support interoperability between different types of numerical software. In order to support specialized vector operations which may be required by an abstract numerical algorithm, Thyra has previously performed all element-wise vector transformations and reductions using the RTOp Trilinos package. While the RTOp approach

provides for flexibility and extensibility, the current implementation lacks of performance portability and disregards any threading support provided by the underlying linear algebra library. In this presentation, we will describe changes which have been made to Thyra to enable it to benefit from the performance portability provided by the Kokkos and Tpetra packages from Trilinos. We will then demonstrate the effect of these changes for several nonlinear solver test problems and consider some issues related to nonlinear solver performance.

Alexander Toth

Sandia National Laboratories  
artoth@sandia.gov

Roger Pawlowski

Multiphysics Simulation Technologies Dept.  
Sandia National Laboratories  
rppawlo@sandia.gov

Ross Bartlett

Sandia National Laboratories  
rabartl@sandia.gov

## CP1

### BDDC and FETI-DP Methods in PETSC

In the last decade, BDDC and FETI-DP methods have proven to be powerful solvers for large and sparse linear systems arising from the finite element discretization of elliptic PDEs, having convergence rates independent of the heterogeneity of the PDE coefficients. I will present the current implementation in the PETSc library, together with numerical results for a variety of discretizations and partial differential equations. Robustness and prospects of these methods will be also discussed.

Stefano Zampini

King Abdullah University of Science and Technology  
stefano.zampini@kaust.edu.sa

## CP2

### Performance Assessment of Hybrid Parallelism for Large-Scale Reservoir Simulation on Petascale Architecture

Scalable solvers are of utmost importance for large-scale reservoir simulation as number of cores increases and memory hierarchy becomes more complex in today's high performance computing platforms. This trend has reshaped the simulation landscape, where the bulk synchronous parallel model must make way for additional means to exploit the underlying architecture. We analyze the performance of a hybrid parallel solver of large-scale reservoir simulation on multi- and many-core architectures.

Amani Alonazi

KAUST  
amani.alonazi@kaust.edu.sa

Marcin Rogowski, Ahmed Al-Zawawi

SAUDIA ARAMCO  
marcin.rogowski@aramco.com,  
ahmed.zawawi.1@aramco.com

David E. Keyes

KAUST

david.keyes@kaust.edu.sa

## CP2

### Radial-Basisfunction Interpolation for Black-Box Multi-Physics Simulation

RBFs are an accurate interpolation method for black-box surface coupled multi-physics simulations with non-matching meshes where we have for non-meshed scattered surface data. Non-homogeneous data distributions (in high-order or adaptive discretizations) and the global support of radial basis functions are challenges in terms of accuracy, efficiency, and scalability. We present tailored RBF methods for large, distributed multi-physics scenarios. The interpolation is augmented by a global low-order polynomial and presents a way to separately compute the polynomial. This, along with cut-off basis functions, which maintains a sparse structure of the interpolation system and a tolerable condition. We further improve accuracy and condition with a rescaling approach for the interpolant and the basisfunctions. High parallel efficiency is achieved with a sophisticated filtering approach for both meshes in the mapping, allowing fully local computation of the system matrix. For preliminary results, see [Lindner et.al., Radial basis function interpolation for black-box multi-physics simulations, 2017]. All methods are implemented in the open-source library preCICE [Uekermann, Partitioned Fluid-Structure Interaction on Massively Parallel Systems, 2016]. It provides data-mapping, accelerated fixed-point iteration schemes for the overall coupled system, and inter-code communication. Its API is designed for quick integration into existing codes.

Florian Lindner, Miriam Mehl

University of Stuttgart  
florian.lindner@ipvs.uni-stuttgart.de,  
miriam.mehl@uni-stuttgart.de

Benjamin Uekermann

Department of Informatics  
Technical University of Munich  
uekerman@in.tum.de

## CP2

### Persistent Homology of Geospatial Signals

At Descartes Labs, we are building a data refinery of remotely sensed imagery in order to enable robust, cloud-based geospatial analytics at scale. Geometrically this translates into constructing a model of the Earth across space and time by merging many image bands from a plethora of sensors at heterogeneous resolutions. Complete coverage of any one point in space and time cannot be guaranteed. We present how we use persistent homology to track where there are gaps in coverage and what impacts they have on signal production.

David Nicholaeff

XCP-2 Eulerian Applications  
Los Alamos National Lab  
dnic@descarteslabs.com

## CP2

### A Parallel Library of P-Adaptive High Order Finite Elements for Geophysical Applications

High order DG methods pose stringent stability restrictions on explicit time discretization methods. In recent work

by the authors, a strategy for the reduction of the computational cost of DG methods has been proposed. A novel semi-implicit, semi-Lagrangian time discretization technique is employed, that allows the use of much longer time steps than explicit schemes. Furthermore, a dynamically adaptive choice of the polynomial degree is employed in each element. This adaptive strategy allows to reduce the number of degrees of freedom by a factor of up to 50%. In this work, we will outline the ongoing development of a parallel library that will allow to implement these adaptive techniques by taking full advantage of degree adaptivity in each coordinate direction of generic hexahedral elements. The key ingredients of the parallelization strategy will be presented along with preliminary numerical results in three dimensions. The preliminary results achieved are promising and hint that the proposed approach is useful to achieve the maximum possible flexibility in the application of high order discontinuous finite element methods to environmental flows.

Giovanni Tumolo  
ICTP, Trieste, Italy  
grumolo@ictp.it

Luca Bonaventura  
Politecnico di Milano  
MOX, Dipartimento di Matematica F. Brioschi  
luca.bonaventura@polimi.it

### CP3

#### A Programming Framework for Performance Tuning in Julia

We present a programming framework which optimizes Julia codes and translates them into C++ automatically. Our framework transforms Julia codes by user-specified rules. The rule files can express various code optimization techniques used in the high-performance computing field. These codes are translated into C++ codes using Eigen library for matrix operations. The generated codes can be parallelized by OpenMP. We also report the performance results of programs using our framework.

Ayumu Gomi  
Graduate School of Systems and Information Engineering  
University of Tsukuba  
agomi@hpcs.cs.tsukuba.ac.jp

Daisuke Takahashi  
Center for Computational Sciences  
University of Tsukuba  
daisuke@cs.tsukuba.ac.jp

### CP3

#### PerfMemPlus: A Memory Performance Analysis Tool Based on Linux Perf

In high performance computing applications many performance bottlenecks are caused by the memory system. We present a new tool for memory performance analysis called PerfMemPlus to identify such bottlenecks. It uses Linux perf as backend which already comes with support for many hardware architectures and continued maintenance. In addition our tool adds support for tracking objects and specialized configuration and analysis features for more effective analysis of memory performance problems.

Christian R. Helm  
The University of Tokyo

christian@eidos.ic.i.u-tokyo.ac.jp

Kenjiro Taura  
University of Tokyo  
tau@eidos.ic.i.u-tokyo.ac.jp

### CP3

#### Acceleration of MuPAT on MATLAB/scilab Using Parallel Processing

We developed MuPAT (Multiple Precision Arithmetic Toolbox) on MATLAB/Scilab for "Ease of Use" of multiple precision arithmetic using Double-double arithmetic and Quad-double arithmetic. They require from 10 to 1,000 double precision operations for each Double-double and Quad-double arithmetic operation. To accelerate these MuPAT operations in the conventional PC environment, we applied parallel processing features such as FMA, SIMD and Multi threadings.

Hotaka Yagi  
Tokyo University of Science  
1415095@ed.tus.ac.jp

Emiko Ishiwata  
Department of Mathematical Information Science,  
Tokyo University of Science, Japan  
ishiwata@rs.kagu.tus.ac.jp

Hidehiko Hasegawa  
University of Tsukuba, Japan  
hasegawa@slis.tsukuba.ac.jp

### CP3

#### Performance Optimization of the SCB Kernel on the X86 Multi-Core Platform

The subcell balance (SCB) method is one of the widely used methods for radiative transfer problems and not straightforward to achieve good performance on modern multi-core CPUs because of its data dependency. In this presentation, we will introduce our recent work of optimizing the SCB kernel which comes from a real-world application on the X86 multi-core platform. Firstly, we will give a performance analysis of SCB using the well known Roofline model. Then, performance optimizations considering both data locality and parallelism for boosting the performance of SCB are to be described in detail. We will also explain the confusing performance difference between MPI and OpenMP which is observed during the multicore parallelism. Finally, the experiment result shows that our fully tuned code has achieved a significantly speedup compared with original code and notably reduce the running time of the real-world application by 22%.

Peng Zhang, Zhang Yang  
Institute of Applied Physics and Computational Mathematics  
zhang\_peng@iapcm.ac.cn, yang\_zhang@iapcm.ac.cn

### CP4

#### Recent Development in Distributed-Memory Hierarchical Interpolative Factorization

Distributed-memory hierarchical interpolative factorization is a distributed-memory algorithm for the hierarchical interpolative factorization, which is a novel framework for efficient solution of partial differential equations and in-



tegral equations. This algorithm can be applied to most elliptic problems as well as low-to-medium frequency hyperbolic problems. Recently, we have tested the parallel algorithms on massive parallel machine. Numerical experiments demonstrate that the problems with almost billions of unknowns can be efficiently solved with thousands of processes in minutes. We also implement functionality for integral equations. This talk focus on the recent developments in distributed-memory hierarchical interpolative factorization.

Yingzhou Li  
Stanford University  
yingzhou.li@duke.edu

Yingzhou Li  
Duke University  
yingzhou.li@duke.edu

Lexing Ying  
Stanford University  
Department of Mathematics  
lexing@stanford.edu

#### CP4

##### A Method to Calculate SVD of a Sparse T-S Matrix with Iterative Correction of Orthogonality

For a tall-skinny matrix  $A$  of large size, we calculate SVD  $A = Q\Sigma W^T$  by the use of the eigenvalue decomposition of  $A^T A$ , and the error of the column orthogonality of  $Q$  is corrected iteratively. The matrix  $Q$  will be dense even the original matrix  $A$  is sparse, however the dense store of  $Q$  can be avoided if the original matrix  $A$  is kept unchanged and  $Q$  is represented implicitly by the use of the relation  $Q = AW\Sigma^\dagger$ .

Hiroshi Murakami  
Department of Mathematics and Information Sciences,  
Tokyo Metropolitan University  
mrkmhrsh@tmu.ac.jp

#### CP4

##### OpenMP Tasks for QR Factorization in PLASMA

We describe recent developments of the QR factorization functions in PLASMA, a numerical linear algebra library for dense matrices on multicore systems. Parallel execution relies on the data-dependent tasks of the OpenMP runtime. An emphasis is given on rectangular matrices, for which a new version of the tree-based hierarchical QR has been implemented.

Jakub Sistek  
The University of Manchester  
School of Mathematics  
jakub.sistek@manchester.ac.uk

Jakub Kurzak  
University of Tennessee Knoxville  
kurzak@eecs.utk.edu

Julien Langou  
University of Colorado Denver  
julien.langou@ucdenver.edu

#### CP4

##### A Hybrid Parallel Block Low-Rank Solver with

##### Flexible Factorization Algorithms and Communication Schemes

We present a hybrid parallel dense BLR linear solver that accommodates arbitrary  $LU/LL^T$  factorization algorithms and communication schemes. A preprocessing step generates a DAG describing distributed-memory parallelism; a simple execution engine traverses the DAG to complete the factorization. We analyze various factorizations algorithms (eg., right-looking) and communication schemes (global synchronous, point-to-point asynchronous), and we study their performance in the context of BLR approximations. We experiment on electromagnetic problems, obtained with the multiphysics code LS-Dyna.

Clement Weisbecker  
Livermore Software Technology Corporation  
clement@lstc.com

Cleve Ashcraft  
Livermore Software Technology  
Corporation  
cleve@lstc.com

Francois-Henry Rouet  
Livermore Software Technology Corp.  
frouet@lstc.com

#### CP5

##### Application of Transactional Memory to Scalable Linear Solvers

As threading has become a requirement for modern multi- and many-core platforms, Transactional Memory (TM) has also received renewed attention in recent years. Our previous work indicated that both mesh [B.L. Bihari, M. Wong, B.R. de Supinski and Lori Diachin, "On the Algorithmic Aspects of Using OpenMP Synchronization Mechanisms: The Effects of Transactional Memory", *IWOMP 2014: Using and Improving OpenMP for Devices, Tasks, and More*, LNCS 8766, pp. 115-129, 2014.] and Algebraic Multigrid (AMG) smoothers [B.L. Bihari, U.M. Yang, M. Wong and B.R. de Supinski, "Transactional Memory for Algebraic Multigrid Smoothers", *IWOMP 2016: OpenMP: Memory, Devices, and Tasks*, LNCS 9903, pp. 320-335, 2016.] can greatly benefit from the freshly updated values automatically synchronized by TM. In this study we extend and generalize our earlier work on TM for Gauss-Seidel (GS) type AMG smoothers to other AMG smoothers — such as  $l_1$  — from the GS family. We implement this work in the well-known *hypr* linear solver library using the "early access" IBM Power8 systems, and time- and resources-permitting, the new Power9 architecture as well. We show that when the recently introduced "time-to-convergence" criterion is used these TM-based iterative solvers outperform their traditional counterparts.

Barna Bihari  
Lawrence Livermore National Lab  
bihari1@llnl.gov

Ulrike Yang  
Lawrence Livermore National Laboratory  
yang11@llnl.gov

#### CP5

##### Performance Evaluation of a Modified Communication-Avoiding Generalized Minimal

## Residual Method on Many Core Platforms

Communication-avoiding (CA) Krylov methods are promising solutions for severe communication bottlenecks on future exascale machines. However, most CA Krylov methods require additional computation instead of reduced communication. In order to maximize the impact of CA, one needs to reduce the cost of communication either via implementation or optimization. In this work, we address this issue on a CA generalized minimal residual (GMRES) method. We propose a modified CA-GMRES method, which reduces both computation and memory access by 30% with keeping the same CA property as the original CA-GMRES method. Compared with a GMRES method based on the modified Gram-Schmidt method, the modified CA-GMRES method based on the Cholesky QR factorization reduces the amount of computation by 30% and improves the arithmetic intensity Flop/Byte from 0.18 to 1.38. These numerical properties, less communication and computation with higher arithmetic intensity, are promising features for future exascale machines with limited memory and network bandwidths. The modified CA-GMRES method is applied to a large scale non-symmetric matrix in an implicit solver of the gyrokinetic toroidal five dimensional Eulerian code GT5D, and its performance is estimated on the Oakforest-PACS (KNL). The numerical experiment shows that compared with the GCR method, computing kernels are accelerated by 1.5x, and the cost of data reduction communication is reduced from 12.5% to 1% of the total cost at 1,280 nodes.

Yasuhiro Idomura, Takuya Ina, Akie Mayumi, Susumu Yamada  
Japan Atomic Energy Agency  
idomura.yasuhiro@jaea.go.jp, ina.takuya@jaea.go.jp,  
mayumi.akie@jaea.go.jp, yamada.susumu@jaea.go.jp

Kazuya Matsumoto  
University of Aizu, Japan  
kazuya-m@u-aizu.ac.jp

Yuuichi Asahi  
CEA, Cadarache, IRFM  
yuuichi.asahi@cea.fr

Toshiyuki Imamura  
RIKEN Advance Institute for Computational Science  
imamura.toshiyuki@riken.jp

## CP5

### A Novel Adaptive Algebraic Multigrid Preconditioner for Large-scale Problems

Algebraic multigrid methods (AMG) are highly used as preconditioners to Krylov space solvers for the solution of linear systems of equations. Several setup techniques for these preconditioners have been developed since their introduction ranging from the classical multigrid up to the most recent adaptive approaches. These share an objective in common which is to alleviate the hypothesis on the behavior of the algebraic smooth errors. In this work, we present a novel AMG technique characterized by further adaptivity into its setup process with the objective to produce a better and cheaper multigrid preconditioner in comparison to other adaptive versions. The default relaxation method used here is the adaptive factorized sparse approximate inverse (aFSAI), which is strongly scalable and whose accuracy can be controlled by tuning the approximate in-

verse density. A set of test vectors representing the algebraic smooth error components is uncovered by applying the Lanczos method to the smoothing operator. These vectors are then used in the calculation of affinity measures between grid points in the coarsening process as well as in the construction of the interpolation operator through a least squares fitting process. We show the performance of the proposed preconditioner in the solution of large-scale linear problems arising from different engineering applications and preliminary results on the parallelization of this method for HPC systems.

Victor A. Paludetto Magri, Massimiliano Ferronato,  
Andrea Franceschini, Carlo Janna  
University of Padova  
victor.magri@dicea.unipd.it, ferronat@dmsa.unipd.it,  
franc90@dmsa.unipd.it, janna@dmsa.unipd.it

## CP5

### Autotuning Qr Factorizations

The increasing complexity and heterogeneity of computer architectures makes it challenging to design both efficient and portable codes. Indeed, designing generic linear algebra kernels that attempt to fit all architectures would not be efficient on any given architecture. Moreover, the careful and customized design of a kernel for a specific architecture will be hardly efficient on the next generation of architectures. Furthermore, writing tailored kernels for every architecture is a daunting task that would require too much time and effort. We present our work on applying the auto-tuning paradigm to target this issue for QR factorization kernels by generating automatically codes specific to a given architecture. The auto-tuning consists in writing an all-purpose highly parametrized kernel template. Every combination of parameters will induce one instance of the kernel. The auto-tuner would then traverse the parameters search space in order to find the best set of parameters for a given architecture. Following this meta-programming methodology, the power of computers is taken advantage of to generate the best kernels for the architecture at hand.

Wissam M. Sid-Lakhdar  
Texas A&M University  
Computer Science and Engineering  
wissam@lbl.gov

## CP6

### Optimizations of Implicit Unstructured Aerodynamics Computations for Emerging Many and Multi-Core HPC Architectures

Scalability of scientific workloads on next-generation computing systems is critically dependent upon the performance of shared-memory multi- and many-core compute nodes, which we address for the important problem class of unstructured meshes. We investigate several state-of-the-practice shared-memory optimization techniques applied to key routines of an unstructured computational aerodynamics application with irregular memory accesses. We target the Intel Knights Landing (KNL) processor, identifying and addressing performance challenges without compromising the floating point numerics. We employ low- and high-level architecture-specific code optimizations involving thread and data-level parallelism. Our approach is based upon a multi-level hierarchical distribution of work and data across both the threads and the SIMD units within each thread. On a 64-core KNL chip, these trans-

formations achieve nearly 2.9x speedup in dominant routines relative to the baseline. They exhibit almost linear strong scalability to the full core count of KNL (64), and thereafter some improvement with hyperthreading. At substantially fewer Watts in both cases, we achieve up to about 1.4x speedup relative to the performance of 72 threads of a 36-core Haswell CPU and roughly equivalent performance to 112 threads of a 56-core Skylake processor. These optimizations are expected to be of value for many other unstructured mesh PDE-based scientific applications as multi- and many-core architecture evolves

Mohammed A. Al Farhan, David E. Keyes  
KAUST  
mohammed.farhan@kaust.edu.sa,  
david.keyes@kaust.edu.sa

## CP6

### Parallel Efficient Mesh Deformation Method Based on Support Vector Regression

We present a parallel mesh deformation method based on the support vector regression for unstructured meshes. The proposed method first trains three SVMs by the coordinates of the boundary points and their known displacements in each direction, and then predicts the displacements of the internal points. The training part of the method is executed with all global boundary points in each decomposed domain. The internal evaluation part is predicted separately without any data dependency.

Xiang Gao, Chuanfu Xu, Dali Li  
National University of Defense Technology  
gaoxiang12@nudt.edu.cn, xuchuanfu@nudt.edu.cn,  
achhx@nudt.edu.cn

## CP6

### Multithreaded Mesh Generation

In this talk, we present a new scalable fine scale parallelization scheme to generate tetrahedral meshes. Our first contribution is a very efficient serial implementation. From this base we developed a multi-threaded version of the Delaunay kernel that concurrently insert points into the tetrahedrization. We use Hilbert curve coordinates to distribute the work between threads. Our strategy is free from heavy synchronization overhead. The key idea to manage conflicts is to modify the partitions by enlarging the space filled by the Hilbert curve. Our implementation is very simple and outperforms previous existing methods.

Jean-François Remacle  
Univ. catholique de Louvain  
jean-francois.remacle@uclouvain.be

## CP7

### Comparison of Different Inexact BDDC Variants for Systems of PDEs

Traditionally, domain decomposition methods as BDDC (Balancing Domain Decomposition by Constraints) or FETI-DP (Finite Element Tearing and Interconnecting - Dual Primal) use sparse direct solvers as building blocks, i.e., to solve local subdomain problems and/or a coarse problem. Especially the direct solution of a coarse problem can lead to an insufficient weak scaling behavior, since the coarse space grows proportionally with the number of processors and subdomains. Therefore, the sparse direct

solvers are often replaced by spectrally equivalent preconditioners, e.g. AMG methods, or multilevel domain decomposition approaches are considered. In this talk, highly scalable implementations of different inexact BDDC variants as well as multilevel BDDC are presented and compared. Also scalability results in two and three dimensions for up to more than 100K computational cores are shown. In the inexact BDDC methods, the inverse action of the partially coupled stiffness matrix is replaced by V-cycles of an AMG method. But classical AMG approaches can be insufficient for a fast convergence of the different inexact BDDC methods applied to realistic problems. The benefit of using an explicit interpolation of rigid body motions in the AMG interpolation (GM approach) is therefore discussed. It is illustrated that an appropriate AMG interpolation is indeed mandatory for fast convergence.

Martin Lanser  
Universitaet zu Koeln  
Mathematisches Institut  
martin.lanser@uni-koeln.de

Axel Klawonn  
Universitaet zu Koeln  
axel.klawonn@uni-koeln.de

Oliver Rheinbach  
Technische Universität Bergakademie Freiberg  
oliver.rheinbach@math.tu-freiberg.de

## CP7

### Robust Domain Decomposition Methods for High Performance Computing in Aerospace Industrial Context

Multi-pre-conditionned algorithms (as AMPFETI) have shown to overcome classical weaknesses of historical domain decomposition strategies (FETI and BDD) and are promising for use in an industrial context. This talk will present recent developments around AMPFETI to improve scalability and enable the use of recycling techniques. We will show applications on practical use cases coming from the aeronautic industry, some of which requiring to solve huge sequences of increasingly large linear systems.

Augustin Parret-Fréaud  
Safran Tech, Rue des Jeunes Bois, Châteaufort  
CS 80112 78772 Magny-Les-Hameaux  
augustin.parret-freaud@safrangroup.com

Christophe Bovet  
Onera - The French Aerospace Lab  
92322 Châtillon, France  
christophe.bovet@onera.fr

Pierre Gosselet  
LMT-Cachan (ENS Cachan, CNRS, Université Paris-Saclay)  
61 avenue du Pdt Wilson, 94235 Cachan, France  
pierre.gosselet@lmt.ens-cachan.fr

Nicole Spillane  
CMAP (UMR 7641, Ecole Polytechnique, CNRS)  
Route de Saclay, 91128 Palaiseau, France  
nicole.spillane@polytechnique.edu

## CP7

### Multi-Level Domain Decomposition Solvers for

### **Intrusive Spectral Stochastic Fem Versus Non-Intrusive Sparse Grid Based Solvers for Large Random Dimensions**

Using the Galerkin projection based intrusive spectral finite element method, uncertainty quantification of practical engineering systems, modeled by stochastic PDE having a large number of random parameters, may involve solving linear systems with billions of unknown, even when the spatial mesh resolution is coarse. To this end, three-level non-overlapping domain decomposition (DD) solvers are developed which exploit efficient preconditioners for preconditioned conjugate gradient method (PCGM) at each level. The numerical and parallel scalabilities (up to 5000 cores) of these solvers are first studied for a stochastic PDE with non-Gaussian parameters having up to 25 random variables. The computational efficiency of the intrusive DD solvers are then demonstrated over sparse grid based non-intrusive solvers.

Abhijit Sarkar  
Associate Professor  
carleton University, Ottawa, Canada  
abhijit\_sarkar@carleton.ca

Ajit Desai  
Carleton University, Canada  
ajit.desai@carleton.ca

Mohammad Khalil  
Sandia National Laboratories  
mkhalil@sandia.gov

Chris Pettit  
United States Naval Academy, USA  
pettitcl@usna.edu

Dominique Poirel  
Royal Military College of Canada, Canada  
poirel-d@rmc.ca

### **CP7**

#### **Simulation of Wind Flows in Large Urban Areas Using a Scalable Domain Decomposition Method**

Many wind environmental problems, such as natural ventilation design and pedestrian comfort rely on accurate knowledge of the wind flow field. In this talk, we study a scalable domain decomposition based incompressible Navier-Stokes solver for solving wind flows in the urban areas. Two real-world scale urban areas in the city of Shenzhen and Guangzhou are applied to test our algorithm. We will report the parallel performance of the method on Tianhe-2A supercomputer.

Zhengzheng Yan  
Shenzhen Institutes of Advanced Technology  
zhengzh.yan@gmail.com

Tangzheng Du  
Chinese Academy of Sciences  
Shenzhen Institutes of Advanced Technology  
tz.du@siat.ac.cn

Rongliang Chen  
Chinese Academy of Sciences  
rl.chen@siat.ac.cn

Xiao-Chuan Cai

University of Colorado, Boulder  
Dept. of Computer Science  
cai@cs.colorado.edu

### **CP8**

#### **Parallelization of Recursive Inverse Factorization using the Chunks and Tasks Programming Model**

In this talk we will present a parallel algorithm for inverse factorization of symmetric positive definite matrices. This algorithm utilizes a hierarchical representation of the matrix and computes an inverse factor by combining an iterative refinement procedure with a recursive binary principal submatrix decomposition. Although our algorithm should be generally applicable, this work was mainly motivated by the need for efficient and scalable inverse factorization of the basis set overlap matrix in large scale electronic structure calculations. We show that for such matrices the computational cost increases only linearly with system size. We will also describe parallelization within the Chunks and Tasks programming model, demonstrate the weak scaling behaviour of the algorithm and give theoretical justification of the scaling behaviour based on critical path length estimation.

Anton Artemov  
Uppsala University  
anton.artemov@it.uu.se

Elias Rudberg  
Division of Scientific Computing, Department of I.T.  
Uppsala University  
elias.rudberg@it.uu.se

Emanuel H. Rubensson  
Uppsala University  
emanuel.rubensson@it.uu.se

### **CP8**

#### **A Distributed Shared Memory Library with Global-View Tasks on High-Performance Interconnects**

Recent high-performance interconnects encourage system designers to revisit Distributed Shared Memory (DSM) models, which provide shared memory on distributed memory machines, as an alternative to message-passing. We propose a DSM library with a global-view task scheduler to achieve a unified programming model for both high productivity and efficiency. To address the scalability issues of conventional DSM systems, we are focusing on reducing software overhead arising from the DSM and its underlying communication layer.

Wataru Endo, Kenjiro Taura  
University of Tokyo  
wendo@eidoss.ic.i.u-tokyo.ac.jp,      tau@eidoss.ic.i.u-tokyo.ac.jp

### **CP8**

#### **Efficient Multiple Precision Dense Matrix Multiplication Library with Automatic Turning Tool**

Reliable software-based long precision floating-point arithmetic libraries, such as QD and MPFR/GMP, are effective for ill-conditioned problems. We have developed "BNC-matmul" library for multiple precision dense matrix multiplication using an automatic tuning method in select-



ing the most effective algorithm among simple triple loop, blocking, and Strassen's algorithms. The tuning time was decreased by predicting the computational times for the blocking algorithm in several block sizes. We will show these techniques and the performance of BNCmatmul.

Tomonori Kouya  
Shizuoka Institute of Science and Technology  
kouya.tomonori@sist.ac.jp

## CP8

### Block Red-Black Milu(0) Preconditioner with Relaxation on GPU

The combination of the block red-black ordering and the relaxed modified incomplete LU factorization without fill-ins (MILU(0)) provides stable and highly parallelizable preconditioners for Krylov solvers. In this strategy, the number of iterations decreases as the block size becomes larger. To take advantage of the preconditioner and GPU, we consider the implementation in which large block sizes and performant data access patterns are compatible. The effectiveness are investigated by solving Poisson's equation for PIC simulations.

Akemi Shioya  
University of Electro-Communications  
akemi.shioya@uec.ac.jp

Yusaku Yamamoto  
The University of Electro-Communications, Japan  
yusaku.yamamoto@uec.ac.jp

## CP9

### A Waveform-Relaxation Method for the Simulation of Rate-Based Models in a Distributed Spiking Neural Network Simulator

Contemporary modeling approaches to the dynamics of neural networks include two important classes of models: biologically grounded spiking neuron models and functionally inspired rate-based units. Although it is a major challenge in neuroscience to form a bridge between spike- and rate-based models and several efficient simulators enabling the distributed simulation of large scale networks exist for both approaches these tools are usually restricted to either rate-based or spike-based models only. We here present a unified simulation framework implemented in the spiking neural network simulator NEST that supports the combination of the two approaches for multi-scale modeling and enables the quantitative validation of mean-field approaches by spiking network simulations. In addition to the standard implementation that relies on communication in short intervals we present an iterative approach based on waveform-relaxation techniques which reduces communication and increases the performance for large-scale simulations of rate-based models with instantaneous interactions in the setup of a spiking neural network simulator.

Jan Hahne  
University of Wuppertal  
hahne@math.uni-wuppertal.de

Matthias Bolten  
Bergische Universität Wuppertal  
Fakultät für Mathematik und Naturwissenschaften

bolten@math.uni-wuppertal.de

## CP9

### Feature Scaling Method for Supervised Spectral Clustering

We propose a supervised spectral clustering method for clustering high dimensional data. To exploit the prior knowledge of partial data and achieve flexible and accurate classification by scaling the features, we formulate a rectangular eigenproblem whose eigenvector has scaling factors. The solution is to use a contour integral method, in which least squares problems are solved at quadrature points in parallel. Numerical experiments show the proposed method outperforms previous methods in terms of accuracy.

Momo Matsuda  
University of Tsukuba  
matsuda@mama.cs.tsukuba.ac.jp

Keiichi Morikuni, Tetsuya Sakurai  
Department of Computer Science  
University of Tsukuba  
morikuni@cs.tsukuba.ac.jp, sakurai@cs.tsukuba.ac.jp

## CP9

### Neural Network Architectures for Kalman Filters

The goal of this talk is to draw the parallel between the prediction-correction nature of the classic and extended Kalman Filter and types of neural networks architectures, like the classical feed-forward layered and recurrent multi-layered perceptron. We want to introduce how the idea of backpropagation, the cornerstone of neural networks plays into Kalman Filters, and how evolving weights in a multi-stream fashion can be orders of magnitude more efficient in terms of the number of training epochs than standard backpropagation. For this presentation is assumed previous knowledge on the derivation of Kalman Filter formulation and stochastic process theory.

Elizabeth C. Ramirez  
Columbia University  
er2732@columbia.edu

## CP9

### Predicting Biological Cell Aggregation Using Scalable Random Forest Decision Trees

Random forest algorithms can be used to find the best combination of variables for predicting events in experimental or simulation data. In this work, we explore the potential of random forest algorithms for predicting the onset of cell aggregation in the amoeba *Dictyostelium Discoideum*, and for the identification of "pacemaker" cells which act as leaders in aggregation. The random forest algorithm is applied to image data of the cells in time, which are labeled by the time and location of cell aggregation. Velocity fields are extracted using particle image velocimetry and using particle tracking methods, which provides a dynamical system in both the Eulerian (lab) and Lagrangian (individual amoeba) reference frames. Then, a list of variables associated with the velocity fields such as magnitude, angle, spatial correlation, and vorticity is created. A large number of random decision forests are created using the variables, and are traversed in order to learn the location and time of cell aggregation from the labeled data. Then,

the trained algorithm is used to predict the onset of aggregation in unlabeled data. The computational costs are also analyzed using both CPU and GPU parallelism. The application of random forests to find predictive variables in physical systems has a broad range of experimental and theoretical applications.

Sulimon Sattari

University of California, Merced  
School of Natural Sciences  
sulimon@es.hokudai.ac.jp

Tamiki Komatsuzaki

Hokkaido University  
tamiki@es.hokudai.ac.jp

## CP10

### Scalable Krylov-FFT-Type Algorithms for Subsurface Scattering Problems

In this paper, we consider parallel Krylov-FFT type high-resolution algorithms for the solution of a subsurface electromagnetic scattering problem. The 3D Helmholtz equation is discretized by high-order compact finite-difference schemes. The matching order compact non-reflecting boundary conditions are applied to preserve the high accuracy of the numerical solution. The resulting systems of finite-difference equations are solved by a preconditioned GMRES method. The FFT-based preconditioner is used for efficient implementation of the developed iterative approach. The main focus of the paper is on the efficient parallel implementation of the proposed algorithms in the shared (OpenMP) and distributed (MPI) memory environment. We also consider the efficiency of the proposed implementation on GPUs using CUDA C. The complexity and scalability of the parallel preconditioned iterative method are analyzed on forward scattering problems with realistic ranges of parameters in soil and mine-like targets. This research establishes the fact that sequential numerical algorithms are drastically inferior to their parallel counterparts. This will be demonstrated by comparing the sequential and parallelized run times of the 2nd, 4th, and 6th order compact algorithms for approximating solutions to the three dimensional Helmholtz equation. The comparison will use several different computational grid sizes and use both CPUs and GPUs via the parallel technologies OpenMP, MPI and CUDA.

Yury A. Gryazin, Ron Gonzales, Yun Teck Lee  
Idaho State University  
gryazin@isu.edu, gonzrona@isu.edu, leeyunt@isu.edu

## CP10

### A Randomized, Communication-Efficient Algorithm for Computing a Rank-Revealing UTV Matrix Decomposition

The talk describes techniques based on randomized projections for efficiently building near-optimal rotation matrices  $U$  and  $V$  for a rank-revealing UTV decomposition. The result is a factorization method that in tested cases yields low-rank approximations approaching those of the SVD in quality, but whose computational costs are competitive with a column-pivoted QR factorization. Furthermore, the algorithm is communication-efficient, enabling effective implementation in parallel environments such as multi-core CPUs, GPUs, distributed memory, etc.

Nathan Heavner

University of Colorado Boulder  
nathan.heavner@colorado.edu

Per-Gunnar Martinsson

University of Oxford  
martinsson@maths.ox.ac.uk

Gregorio Quintana-Orti

Departamento de Ingeniería y Ciencia de Computadores  
Universidad Jaume I  
gquintan@icc.uji.es

## CP10

### An Effective GPU Implementation of the Adaptive Pattern FSAI Preconditioner for SPD Problems

GPUs are, by now a well established platform exhibiting significantly better performance than conventional CPUs. However, in general, only algorithms having a regular and fixed memory access pattern can exploit GPUs at their best whereas most of existing kernels need to be redesigned. FSAI, an effective preconditioner, has been already successfully ported on GPU in its static pattern version. In this communication we show how the more complex but much more effective adaptive version of FSAI can also take advantage of GPU massive parallelism. A few examples on real world problems show the effectiveness of the proposed approach with a significant enhancement with respect to the CPU implementation.

Massimo Bernaschi, Mauro Carrozzo  
Istituto Applicazioni del Calcolo, CNR, Italy  
massimo.bernaschi@cnr.it, m.carrozzo@iac.cnr.it

Ferronato Massimiliano, Carlo Janna

Dept. ICEA - University of Padova  
massimiliano.ferronato@unipd.it, carlo.janna@unipd.it

## CP11

### Scalable Spectral Clustering

Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG) [https://doi.org/10.1137/S1064827500366124] has practically demonstrated [https://arxiv.org/abs/1708.07481] to efficiently solve eigenvalue problems for graph Laplacians that appear in spectral clustering, e.g., determining 100% correct 160 clusters in Stochastic Block Partition Streaming Graph Challenge (SBPSGC) static 2M/41M vertices/edges graph in <2,000 sec and with 50GB memory using LOBPCG Python code with threaded ATLAS. We now test scalability partitioning SBPSGC static graphs using PETSc/SLEPc/LOBPCG code with MPI in shared and distributed memory environments.

Andrew Knyazev

Mitsubishi Electric Research Laboratories (MERL)  
knyazev@merl.com

## CP11

### Low Communication Neighbor Discovery for Matrix Migration

Migration of arbitrary sparse matrices from one set of processors to another is a necessary part of matrix transposition, matrix repartitioning and sparse-matrix matrix multiplication. Far too often, neighbor information from the

original pre-migration matrix is discarded and new neighbors are discovered *ex nihilo* — a process which is at least logarithmic in the number of processors. For these three cases, we demonstrate an algorithm to communication additional information during the migration in order to construct the neighbor information of the post-migration matrix with  $O(1)$  communication. The algorithm works for general matrices and does not presume any structure barring a bounded number of nonzeros per matrix row.

Christopher Siefert  
Sandia National Laboratories  
csiefer@sandia.gov

Christopher Luchini  
Sci Tac LLC  
luchini@sci-tac.com

## CP11

### A Molecular Fingerprint Algorithm for Learning Force Fields from Ab Initio Calculations

Molecular fingerprints, i.e. feature vectors describing atomistic neighborhood configurations, is an important abstraction and a key ingredient for data-driven modeling of molecular energy and interatomic force. In this talk, I will present a molecular fingerprint algorithm, which is robust and efficient, for fitting per-atom scalar and vector quantities. The fingerprint is essentially a continuous density field formed through the superimposition of smoothing kernels centered on the atoms. We suggest that the distance between the fingerprints be measured as a weighted L2 norm of their difference. Rotational invariance of the fingerprint is achieved by aligning, for each fingerprint instance, the neighboring atoms onto a set of local canonical coordinate frame computed from a kernel minimization optimization procedure. We show that this approach is superior over PCA-based methods especially when the atomistic neighborhood is sparse and/or contains symmetry. To minimize the computational cost, we derive numerical recipes for discrete sampling and evaluation of the fingerprint using optimal quadrature rules. We also experiment on the choice of weight functions for constructing the density fields, and characterize their performance for fitting interatomic potentials. The applicability of the fingerprint is demonstrated through a set of realistic benchmark problems.

Yu-Hang Tang  
Brown University  
Tang.Maxin@gmail.com

## CP12

### FMM-Accelerated Wave Scattering Solver on Emerging Many and Multi-Core High-Performance Computing Architectures

Algorithmic and architecture-oriented optimizations are essential for achieving performance worthy of anticipated energy-austere exascale systems comprised of manycore accelerators. We present an acoustics wave scattering application that employs FMM as a matrix-free method for solving the linear system generated by the discretization of the governing Helmholtz equations. The FMM-Helmholtz kernel is capable of treating non-trivial singular and near-field integration points within FMM as opposed to external singularity correction. This cuts down redundant computations and data movements. We use a tree-based FMM implementation to augment the Directed Acyclic Graph

(DAG) of Intel's TBB to traverse the tree on emerging Intel many- and multi-core HPC architectures. We extract the potential SIMD and thread-level parallelisms of different computationally intensive kernels of FMM. We employ hand-written vectorization via utilizing the AVX-512 intrinsics to optimize the low-level kernels. We apply our hierarchical sparse data exchange protocol for exchanging far-field cells to scale on challenging boundary distributions such as airfoils and fuselages. We build a performance model to analyze the behavior of hybrid programming paradigms (MPI+TBB). Multiple strategies are developed to extract the best combinations of different programming models within a chip. These include thread/task pinning to a thread, core, tile, quadrant, and node, through low-level programming interfaces.

Mustafa Abduljabbar  
Extreme Computing Research Center  
King Abdullah University of Science and Technology  
mustafa.abduljabbar@kaust.edu.sa

Noha Al-Harhi  
King Abdullah University of Science and Technology  
noha.harhi@kaust.edu.sa

Mohammed A. Al Farhan  
KAUST  
mohammed.farhan@kaust.edu.sa

Rui Chen, Huda Ibeid  
King Abdullah University of Science and Technology  
rui.chen@kaust.edu.sa, huda.ibeid@kaust.edu.sa

Rio Yokota  
Tokyo Institute of Technology  
rioyokota@gsic.titech.ac.jp

Hakan Bagci  
King Abdullah University of Science and Technology  
hakan.bagci@kaust.edu.sa

David E. Keyes  
KAUST  
david.keyes@kaust.edu.sa

## CP12

### Parallel 3d Ffts for Scientific Application Codes on Exascale Architectures

3d FFTs are central to many scientific application codes ranging from Materials Science and Chemistry to Accelerator design and Spectral Methods in general. 3d FFTs are amongst the most demanding algorithms, in terms of communication latency and bandwidth on large parallel machines, to scaling to large core counts. This is due to the required transpose of the complete grid of data points and the fact that the calculation cost of FFTs only scales as  $N \log N$ , where  $N$  is the grid size. This gives a ratio of calculations to communications that scales as  $\log N$  compared to say  $N$  for a parallel matrix multiply. While this limits parallel scaling most scientific codes do not perform parallel FFTs across all the compute nodes they are running on as they typically have multilevels of parallelism in the code. Even so the FFTs still significantly limit performance and parallel scaling on many-core Exascale type architectures. In order to avoid the latency bottle neck as well as reduce communications as much as possible it is necessary to have well designed hybrid OpenMP/MPI FFTs on many-core architectures that can also overlap communi-

cations and calculations. We present research into the best methods of performing these types of parallel 3d FFTs for different scientific applications on the new many-core massively parallel machines such as the Cray XC40 Intel Phi KNL based machine Cori at NERSC, LBNL focusing on applications in materials science and chemistry.

Andrew M. Canning  
Lawrence Berkeley National Laboratory  
acanning@lbl.gov

## CP12

### Application of Parallel Hierarchical Matrices in Spatial Statistics and Parameter Identification

The main goal of this work is to introduce the parallel hierarchical (H-) matrix library HLIBpro [Kriemann'05] to the statistical community. We implement the HLIBCov package, which is an extension of the HLIBpro [Ronald Kriemann. Parallel H-matrix arithmetics on shared memory systems. *Computing*, 74(3):273–297, 2005] for typical linear algebra operations in spatial statistics and data assimilation. Intel's Threading Building Blocks are used for parallelization on a shared memory systems. We approximate large covariance matrices, maximize likelihood functions and estimate unknown/uncertain parameters of covariance matrices [A. Litvinenko, Y. Sun, M. G. Genton, and D. Keyes. Likelihood approximation with hierarchical matrices for large spatial datasets, arXiv:1709.04419, 2017]. We show that an approximate hierarchical Cholesky factorization of a dense covariance matrix of size  $2M \times 2M$  can be computed on a modern multi-core desktop in two minutes. The H-matrix format has computational cost  $O(k^2 n \log^2 n/p)$  and storage  $O(kn \log n)$ , where the rank  $k$  is a small integer (typically  $k < 25$ ),  $p$  the number of cores and  $n$  the number of locations on a fairly general mesh. For reproducibility we provide the C++ code, the documentation, and the synthetic data. HLIBpro is freely available on [www.hlibpro.com] in binary form for academic purposes.

Alexander Litvinenko  
Stochastic Numerics and Bayesian Computational groups, KAUST  
alexander.litvinenko@kaust.edu.sa

## CP12

### Large-Scale Parallel Finite-Element Analysis of Seismic Response Considering Soil-Structure Interaction

A large-scale parallel FE analysis of a 12 million element model is performed to investigate the safety of a structure under a huge earthquake in consideration of the interaction between the structure and the surrounding soil. A dynamic system equation is integrated in time up to 20,000 steps to represent the five-second long motion of an observed earthquake. 76,800 cores of a supercomputer at The University of Tokyo are used to execute this computation.

Yoshiyuki Takahashi  
Kajima Corporation  
takayoshi@kajima.com

Naoki Morita  
Graduate School of Frontier Sciences  
The University of Tokyo  
morita@multi.k.u-tokyo.ac.jp

Toshihide Saka  
Kajima Technical Research Institute  
sakat@kajima.com

Gaku Hashimoto  
Graduate School of Frontier Sciences  
The University of Tokyo  
ghashimoto@k.u-tokyo.ac.jp

Hiroshi Okuda  
The University of Tokyo  
okuda@edu.k.u-tokyo.ac.jp

Toshihiro Koiso  
Kajima Technical Research Institute  
koisot@kajima.com

Kazuhiko Yamada  
Kajima Corporation  
yamada-kazuhiko@kajima.com

## CP13

### Event Detection in Multi-Variate Scientific Simulations Using Feature Anomaly Metrics

In multi-variate multi-physics scientific simulations, anomalous events occur at locations in space-time domain that are hard to predict, for example ignition fronts in combustion. It is often required to identify these events promptly and precisely such that necessary actions may be taken (e.g., triggering in-situ analysis, data checkpoint, mesh refinement), which is challenging in a distributed setting since these events are local in space and/or time. We propose the use of feature anomaly metrics (FAMs) to trigger the detection of such events. Due to tightly coupled physics, anomalies do not always manifest as outliers in individual variables, but as clusters away from the axes in the joint variable space. The FAM quantifies the contribution of each variable to anomalies in the joint variable space based on its alignment with vectors that point towards anomalous clusters. To construct such vectors, we seek a change of basis in a manner analogous to PCA. While PCA yields a change of basis guided by the co-variance matrix, a measure of data spread, we desire a change of basis guided by a measure of data outlieriness, co-Kurtosis, which is a symmetric fourth-order tensor. We employ symmetric CP decomposition of the co-Kurtosis tensor to perform a change of basis and construct FAMs. We examine the efficacy of FAMs in identifying anomalous events in synthetic data as well as canonical 1D combustion simulation data.

Aditya Konduri, Hemanth Kolla, Julia Ling, Philip Kegelmeyer, Timothy Shead  
Sandia National Laboratories  
akondur@sandia.gov, hnkolla@sandia.gov,  
jling@sandia.gov, wpk@sandia.gov, tshead@sandia.gov

Daniel M. Dunlavy  
Sandia National Laboratories  
Computer Science and Informatics  
dmdunla@sandia.gov

Warren Davis  
Sandia National Laboratories  
wldavis@sandia.gov

## CP13

### The Journey from On-Premises to Serverless in Sci-

## entific Computing

The evolution of scientific computing from on-premises infrastructures to Clouds has resulted in unprecedented changes in the last decades. The transition from physical clusters of PCs to cost-aware virtual elastic clusters introduced significant advantages for scientific computing. Then, the advancements of hypervisors and container-based technologies paved the way for serverless computing to surge in the field of scientific computing. Functions coded in different programming languages are executed in response to events on the infrastructure of a public Cloud provider, as is the case of AWS Lambda. This has introduced significant elasticity improvements, with respect to the use of Virtual Machines in an Infrastructure as a Service Cloud. Thus, it is now possible to execute in parallel thousands of invocations of the same function to perform complex distributed computations under a time-constrained execution limit. In this talk we describe the challenges of this journey and the open-source solutions developed in the context of large-scale projects, such as INDIGO-DataCloud, adopted to address these issues. We cover from more mature developments being already used in the EGI Federated Cloud to support parallel computing in scientific communities, to innovative technologies such as the execution of Docker containers on a serverless computing platform (AWS Lambda) to perform parallel Deep Learning analysis of datasets.

Germán Moltó

Institute for Molecular Imaging Technologies  
Universitat Politècnica de València  
gmolto@dsic.upv.es

Miguel Caballer, Carlos de Alfonso, Alfonso Pérez,  
Amanda Calatrava, Ignacio Blanquer  
Institute for Molecular Imaging Technologies  
Universitat Politècnica de Valencia  
micafer1@upv.es, caralla@upv.es, alpegon3@upv.es, am-  
caar@i3m.upv.es, iblanque@dsic.upv.es

## CP13

### BIND: Platform for Integrating Tasks of Bigdata Application in Cloud

For one application using big data, multiple and heterogeneous works are necessary. Thus, it is necessary to integrate these works. Also, an application for big data handles massive data, so the parallelism improves the performance of the application. BIND is implemented to satisfy those needs with python. BIND is a data-analytic cloud platform based on HDFS for big data processing applications, integrating whole tasks (e.g., crawling, script, big data processing, and visualization). In BIND, master receives heartbeat messages from slaves. These messages include serviceable resources of slaves with which master makes order of priority to execute tasks. With this order, master sends execute command to each slave. BIND also can be executed like stream processing through splitting and shifting whole processes like pipeline of computer architecture. It means that each machine in cloud is an element of a pipeline and these machines are executed in time-sliced fashion. This concept help users see approximate results quickly. Also, BIND adopts different data to same task in each slave. It needs data parallelism that is limited by the number of slaves. For syntax sugar, Bind adopts jQuery-based programming style. Each task is considered as independent Node, and it helps users do programing easily. For evaluating BIND, twitter data-analytic application and sentence classification application using CNN are used. BIND reduces the

execution time of two applications considerably.

Yoonsu Shin, songkuk kim

School of Integrated Technology

Yonsei University

sinowl@yonsei.ac.kr, songkuk@yonsei.ac.kr

## CP13

### Tuple Space Implementation in a Parallel Workflow Engine, Opendiel

A tuple server allows for processes to store their data to be sent so that each process does not have to block until the receiving process can receive the data. Using multiple, distributed tuple servers allows for this paradigm to become highly scalable, taking advantage of CPU cores and memory across multiple nodes on distributed memory systems. In this presentation, we will discuss the implementation and interfaces of a parallel workflow engine.

Zachary Trzil

University of Tennessee

ztrzil@vols.utk.edu

Kwai L. Wong

Joint Institute for Computational Science

University of Tennessee/ORNL

kwong@utk.edu

## CP14

### Scaling Machine Learning Models to Optimize Parallel Applications Using Feature Space Partitioning

With the rapid growth of complexities in HPC infrastructures, there is an increased need to build scalable machine learning models to automatically tune scientific codes for better performance. The massive training time might hinder the ease of adoption of machine learning based approaches. With the massive and powerful computational resources becoming increasingly available, scaling traditional machine learning model training seems promising. However, simple trial and run or greedy methods are not optimal with a variety in the training data. In this work, we focus on two important aspects to improve the prediction accuracies of machine learning models that aim to tune the optimal number of threads for a parallel code, while minimizing the training time. These are (i) feature space partitioning, a method to create sub-spaces from the training data and (ii) ensemble models, where localized models are generated on the partitioned sub-spaces and collectively form a powerful model with improved prediction accuracies. This modular approach is flexible to add new local models as and when new training data is available without the need to retrain everytime.

Murali Emani

Lawrence Livermore National Laboratory

emani1@llnl.gov

## CP14

### Assessing and Improving the Numerical Solution of Atmospheric Physics in E3SM

The numerical errors arising from parameterized subgrid-scale atmospheric physics have been largely overlooked in Earth system models. As spatial resolution increases, the computational effort expended on atmospheric dynamics will only provide limited gains in overall accuracy due to



numerical errors from parameterizations. In this talk, we focus on assessing the time integration and process coupling methods utilized in the Energy Exascale Earth System Model (E3SM) and explore approaches to improve process convergence and accuracy, while considering computational costs. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-ABS-738441

David J. Gardner, Carol S. Woodward  
Lawrence Livermore National Laboratory  
gardner48@llnl.gov, woodward6@llnl.gov

Hui Wan  
Pacific Northwest National Laboratory  
Richland, WA 99354  
hui.wan@pnnl.gov

Chris Vogl  
Lawrence Livermore National Laboratory  
vogl2@llnl.gov

Philip Rasch  
Pacific Northwest National Laboratory  
philip.rasch@pnnl.gov

#### CP14

##### **Tensor Decomposition to Perform Change of Basis in Multi-Variate Hpc Data to Preserve Higher Order Statistical Moments**

In multi-variate scientific computing data, it is often of interest to perform a change of basis using methods such as principal component analysis (PCA) for dimensionality reduction, identifying underlying manifolds or discovering reduced order models (ROMs). The resulting transformation preserves the co-variance amongst the multiple variables and generalizes well only under certain conditions; the new basis vectors are linearly uncorrelated and can be deemed independent only if the underlying multi-variate distribution is joint-normal. For many applications (e.g. turbulent combustion) the data distribution (marginal and/or joint) is not normal and a basis transformation that preserves moments higher than co-variance-co-skewness, co-Kurtosis is desirable. We propose a general method for basis transformation aimed at preserving higher order joint statistical moments using symmetric Tensor decomposition. Such a transformation is useful, for instance, in identifying directions in the multi-variate space along which anomalies or outliers appear, which will not be captured by PCA. While co-variance is a matrix (order two Tensor) higher order joint moments constitute higher order Tensors, requiring appropriate Tensor decomposition methods. We propose to use symmetric CP decomposition to identify a changed basis that preserves/maximizes the corresponding higher order moment and demonstrate the method using synthetic as well as real data from a combustion simulation.

Hemanth Kolla, Aditya Konduri  
Sandia National Laboratories  
hnkolla@sandia.gov, akondur@sandia.gov

Prashant Rai  
Sandia National Laboratories  
Livermore, CA, USA  
pmrai@sandia.gov

Tammy Kolda, Warren Davis

Sandia National Laboratories  
tgkolda@sandia.gov, wldavis@sandia.gov

#### CP14

##### **Towards a General Purpose Multirate Time Integration Software Suite for Multiphysics Problems**

With the advance of computational power, simulations are becoming increasingly heterogeneous and multiscale. For such systems, advancing the entire computation at the step size of the fastest scale is highly inefficient. There is a growing need for time stepping methods that can instead advance each problem component separately at its natural time scale and couple the components efficiently. To this end, we are developing a prototype suite of general purpose multirate integrators that are suitable across a wide range of multiphysics problems. The suite is being developed for eventual inclusion into the SUNDIALS solver library. The main challenge with developing a multirate method is designing the coupling between components such that the method is accurate and stable while also computationally efficient. In this talk, we describe our experience in testing and selecting for the highest performing methods for inclusion into the software suite, and discuss the aspects of their design that make them efficient. We cover explicit-explicit, implicit-implicit, and implicit-explicit methods, some from the literature and others of our design. We also describe the implementation challenges and additional user requirements of developing software for multirate problems. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

John Loffeld  
Center for Applied Scientific Computing  
Lawrence Livermore National Laboratory  
loffeld1@llnl.gov

Slaven Peles  
Lawrence Livermore National Laboratory  
peles2@llnl.gov

Carol S. Woodward  
Lawrence Livermore National Laboratory  
woodward6@llnl.gov

#### CP15

##### **Recent Advances on Finite Element Method and Simulation of Invisibility Cloaks with Metamaterials**

In the June 23, 2006's issue of Science magazine, Pendry et al and Leonhardt independently published their papers on electromagnetic cloaking. In Nov.10, 2006's Science magazine, Pendry et al demonstrated the first practical realization of such a cloak with the use of artificially constructed metamaterials. Since then, there is a growing interest in using metamaterials to design invisibility cloaks. In this talk, I will first give a brief introduction to invisibility cloaks with metamaterials, then I will focus on some time-domain cloaking models. Well-posedness study and time-domain finite element method for these models will be presented. Finally, I will show some numerical simulations of time-domain cloaking and optical black holes. I will conclude the talk with some open issues. [1] J. Li and Y. Huang, Time-Domain Finite Element Methods for Maxwell's Equations in Metamaterials, Springer Series in Computational Mathematics, vol.43, Springer, 2013. [2]

J. Li, Y. Huang and W. Yang, Well-posedness study and finite element simulation of time-domain cylindrical and elliptical cloaks, *Math. Comput.* 84 (2015) 543–562. [3] J. Li, Y. Huang, W. Yang and A. Wood, Mathematical analysis and time-domain finite element simulation of carpet cloak, *SIAM J. Appl. Math.* 74(4) (2014) 1136–1151. [4] W. Yang, J. Li and Y. Huang, Mathematical analysis and finite element time domain simulation for electromagnetic cloaks of arbitrary shapes (in review).

Jichun Li

Department of Mathematical Sciences  
University of Nevada, Las Vegas, NV 89154-4020  
jichun.li@unlv.edu

## CP15

### Lessons and Accomplishments of the Computational Science Undergraduate Research Experiences Program

The Computational Science for Undergraduate Research Experiences (CSURE) and the Research Experiences for Computational Science, Engineering, and Mathematics (RECSEM) program are NSF funded Research Experiences for Undergraduates (REU) programs organized by the University of Tennessee's Joint Institute for Computational Sciences (JICS), [www.jics.utk.edu/recsem-reu](http://www.jics.utk.edu/recsem-reu). The main goal of these programs is to direct a group of undergraduate students to explore the emergent computational science models and techniques using the high performance computing resources at the National Institute for Computational Sciences (NICS) and XSEDE. In this talk we discuss the experiences and resolutions in managing and coordinating the program, delivering cohesive tutorial materials, mentoring of individual projects, lessons learned, and improvement over the course of the program.

Kwai L. Wong

Joint Institute for Computational Science  
University of Tennessee/ORNL  
kwong@utk.edu

## MS1

### Achieving Performance While Preserving Portability for NGS Application

DNA damage is a major source of mutation. Patient samples contain natural genetic variations created by mutations such as indels, or single nucleotide polymorphisms (SNPs). We need sophisticated sequence alignment tools to accelerate identification of mutations of various types such as substitution, insertion, deletion (indels) or frameshift. Alignment is computationally intensive and expensive. State-of-the-art alignment algorithms BWA, bowtie, BaraCUDA and nbowtie do not take advantage of the rich resources of novel hardware. The same concern holds good for assembly algorithms. Why? These algorithms are either closely tied to a particular programming language or hardware causing extensive intervention from the programmer to massively restructure the code. Such interventions can slow tools productivity and can be erroneous too. These limit tools usage by biologists who would rather spend quality time on science and not on programming and debugging. Our research plan aims to produce a fast, accurate, reliable, reproducible and portable tool. A tool that will help us be ready to tackle an unexpected event that anticipates a quick response and treatment. Our software tool will use high-level programming models and also use docker container, a major requirement to address repro-

ducible research and minimize the IT support requirement in any clinical lab setting thus easing the deployment and portability aspect of usability.

Sunita Chandrasekaran

University of Delaware  
schandra@udel.edu

## MS1

### Decluttering HPC Software Chaos with SPACK

HPC resources are available to scientists at national laboratories, universities, and industry. They enable solutions to problems that are intractable at smaller scale. However, technical barriers to entry in HPC are extremely high. Scientists must know how to build and run complex applications, as well as tens or hundreds of dependency libraries, just to run simple experiments. Spack is an open source package manager that allows scientists to easily install thousands of packages automatically. It is modeled after common system-level package managers, which have made commodity Linux systems extremely easy to use. Spack is also tailored for HPC. It allows users to experiment with different compilers, optimizations, build options, and dependency versions, so that scientists can easily build performant software on supercomputers. Spack has a rapidly growing community, with over 180 contributors at organizations worldwide. This talk will introduce Spack and show how it can make scientists more productive.

Todd Gamblin

Lawrence Livermore National Laboratory  
tgamblin@llnl.gov

## MS1

### Increasing User and Application Diversity Through Programmable Cyberinfrastructure with Jetstream

Jetstream is the first production cloud funded by the US National Science Foundation (NSF) for conducting general-purpose science and engineering research as well as an easy-to-use platform for education activities. Unlike many high-performance computing systems, Jetstream provides an interactive graphical user interface. This interface provides for a lower barrier of entry for use by educators, students, scientists, and engineers. Jetstream is built upon OpenStack and enables programmable cyberinfrastructure by leveraging APIs to support applications and science gateways that are not as well represented on other NSF resources. A key part of Jetstream's mission is to extend the reach of the NSF's eXtreme Digital (XD) program to a community of users who have not previously utilized those resources. The Pervasive Technology Institute at Indiana University also operates a research desktop environment focused on the idea of making a standard Linux desktop available remotely. The desktop contains all the normal HPC command line tools, but also provides access to interactive applications like Matlab, R-Studio and Jupyter by allowing users to run those tools on their remote desktop without having to submit batch jobs. The goal of this project is to lower the barrier of entry and broaden adoption of traditional HPC and high-throughput computing environments. We will discuss national and local implications and experiences gained from operating these environments.

David Hancock, Robert Henschel  
Indiana University

dyhancoc@iu.edu, henschel@iu.edu

## MS1

### HPC As a Tool for Better Science The Trend Towards Applications As a Service

While High Performance Computing (HPC) has been around for decades, it has been largely seen as a small niche for very limited scientific challenges. The cloud computing revolution had the wonderful side effect that everybody can now easily accept that certain tasks are transparently performed elsewhere. Web-based user interfaces enable an application interaction regardless of the actual location of the computation. As such more and more HPC centers offer web-portals to access their systems and applications also offer a web-based front-end, so that the obscure green font on black screen magic of a typical SSH session is hidden from the end user. This enables both new groups to use HPC systems but also provides power users a more error proof and efficient way of using installed applications. This talk showcases how this application as a service mode has changed the computing landscape in a multi-disciplinary research laboratory both from a users and an HPC operators perspective.

Guido Juckeland  
HZDR  
g.juckeland@hzdr.de

Axel Huebl  
Helmholtz-Zentrum Dresden-Rossendorf (HZDR),  
Germany  
a.huebl@hzdr.de

## MS2

### Pipelining Fast Fourier Transform on the OpenPOWER Cluster

Computer systems with accelerators, such as GPUs, have become mainstream for supercomputing. OpenPOWER clusters are the open architecture of supercomputing systems with POWER processors and NVIDIA's GPU which are connect with a new high speed interconnect NVLink. Fast Fourier Transform requires high network bandwidth to transpose arrays between distributed processes and also between GPUs and CPUs, therefore NVLink accelerates array transpose in the FFT calculation. We apply pipelining technique to overlap array exchange and calculation and evaluate parallel FFT on the OpenPOWER Cluster.

Jun Doi  
IBM Research - Tokyo  
doichan@jp.ibm.com

## MS2

### SPIRAL FFT

In this talk we address the question of how to automatically map computational kernels like the fast Fourier transform (FFT) across a wide range of computing platforms to highly efficient code, and prove the correctness of the synthesized code. This addresses two fundamental problems that software developers are faced with: performance portability across the ever-changing landscape of parallel platforms, and verifiable correctness of sophisticated floating-point code. The problem is attacked as follows: We develop a formal framework to capture computational algorithms, computing platforms, and program transfor-

mations of interest, using a unifying mathematical formalism we call operator language (OL). Then we cast the problem of synthesizing highly optimized computational kernels for a given machine as a strongly constrained optimization problem that is solved by a multi-stage rewriting system. Since all rewrite steps are semantics preserving identity operations, our approach allows us to formally prove the equivalence between the kernel specification and the synthesized program. We have implemented this approach as part of the Spiral system where we have formalized a selection of computational kernels from the signal and image processing domain with focus on the FFT, software-defined radio, and robotic vehicle control.

Franz Franchetti  
Department of Electrical and Computer Engineering  
Carnegie Mellon University  
franzf@ece.cmu.edu

## MS2

### Automatic FFT Kernel Generation for CUDA GPUs

FFT computation largely depends on its transform size. To achieve better performance, it is important to generate an FFT kernel for given transform size. In this talk, I will present an auto-tuning FFT library for CUDA GPUs, which generates various candidate FFT kernels.

Akira Nukada  
Tokyo Institute of Technology  
nukada@smg.is.titech.ac.jp

## MS2

### Implementation of Parallel FFTs on Cluster of Intel Xeon Phi Processors

In this talk, we propose an implementation of parallel fast Fourier transforms (FFTs) on cluster of Intel Xeon Phi processors. To obtain peak performance of the Intel Xeon Phi processor, it must be both vectorized and multithreaded. We present an approach that makes use of the MCDRAM of the Knight Landing processor by blocking technique. We also present a computation-communication overlap method that introduces a communication thread with OpenMP. Performance results of FFTs on a cluster of Intel Xeon Phi processors are reported.

Daisuke Takahashi  
Center for Computational Sciences  
University of Tsukuba  
daisuke@cs.tsukuba.ac.jp

## MS3

### NVIDIA's Batched BLAS, Challenges, and C++ Abstractions

Batched BLAS operations have already appeared in cuBLAS with excellent performance properties. In this work, we review applications of NVIDIA's Batched BLAS primitives to tensor computations in machine learning and the efficient computation of FFTs, challenges going forward for vendors attempting to implement the Batched BLAS and BLAS G2 proposals including interleaved formats and group-batching, and ongoing work to implement a C++ BLAS interface that enables more modular, portable, and composable linear algebra. The C++ library adopts the C++17 Standard Library Algorithms extensions for parallel processing. By accepting an execution policy param-

eter, compute environment information can be passed by the user to many backends and composed with other useful library-wide abstractions. These follow closely the C++ BLAS proposal with some key differences.

Cris Cecka  
NVIDIA Corporation  
ccecka@nvidia.com

### MS3

#### Batched BLAS in High-Order Finite Element Methods

Computing with high-order finite element methods often results in a large number of sparse matrix-vector products in an iterative linear solver. For high orders it is more efficient to construct these with a series of relatively small repeated tensor contractions for each finite element, instead of first computing and storing the large sparse matrix. This talk will examine the problem and what parallelism exists that should be exploited on both CPUs and GPUs, discuss optimization strategies, and will go on to explain what is required from the batched BLAS standard routines in order to efficiently apply them to these methods.

Johann Dahm  
IBM  
Johann.Dahm@ibm.com

Steve Rennich  
NVIDIA Corporation  
srennich@nvidia.com

David Medina  
Rice University  
dmed256@gmail.com

Veselin Dobrev, Tzanio Kolev  
Lawrence Livermore National Laboratory  
dobrev1@llnl.gov, kolev1@llnl.gov

### MS3

#### Standardization and Optimization Opportunities for Batching Linear Algebra on Intel Processors

Many high-performance computing and machine learning applications rely on the use of independent dense linear algebra computations on large groups of matrices. As these operations are often performed on relatively small matrices, batching the operations allows exploiting the parallelism available on today's multi- and many-core systems. A need for standardizing batched APIs exists, and we present the competing APIs. We consider batched BLAS and LAPACK subroutines that operate on either the canonical data layout or on a packed, non-canonical layout that allows for cross-matrix SIMD vectorization. We consider the merits of each approach and look at the performance speedup over non-batched subroutines on the latest Intel architectures.

Sarah Knepper, Murat E. Guney, Kazushige Goto, Louise Huot, Arthur Mitrano, Shane Story  
Intel Corporation  
sarah.knepper@intel.com, murat.e.guney@intel.com, kazushige.goto@intel.com, louise.huot@intel.com, arthur.araujo.mitrano@intel.com, shane.story@intel.com

Mawussi Zounon  
University of Manchester

mawussi.zounon@manchester.ac.uk

### MS3

#### Vector-friendly Batched BLAS and LAPACK Kernels : Design and Applications

BLAS/LAPACK operations on groups of small matrices are becoming increasingly popular both in applications from HPC and machine learning. The community is working towards a standard for such "batched linear algebra". We present one implementation of batched BLAS/LAPACK kernels in an open-source Kokkoskernels package. The implementations use a non-canonical compact data layout for vector friendly implementation of the batched kernels. The data layout allows vectorization across multiple small matrices. Furthermore, the implementations are callable within threads, or team of threads for better kernel performance. We demonstrate performance-portable implementation of the batched kernels in KNL and GPU architectures. We also integrate the batched kernels in a line solver and show the resulting speedup. Detailed performance analysis of the kernels show that the kernels result in better vector-utilization.

Siva Rajamanickam, Kyungjoo Kim, Andrew Bradley, Mehmet Deveci  
Sandia National Laboratories  
srajama@sandia.gov, kyukim@sandia.gov, ambradl@sandia.gov, mndevci@sandia.gov

Simon D. Hammond  
Scalable Computer Architectures  
Sandia National Laboratories  
sdhammo@sandia.gov

Christian Trott  
Sandia National Laboratories  
crtrott@sandia.gov

### MS4

#### Large-Scale Applications for Real-World Problems Solving Partial Differential Equations on Heterogeneous GPU Machines

It is necessary to manage the hierarchy of parallel computing to realize large-scale real-world simulations solving partial differential equations. Especially, a GPU has more than 3,000 cores on a chip and we have a fine-grain parallelization for GPU computing and a coarse-grain parallelization for multiple-node MPI computation for the implementation on supercomputers equipped with a large number of GPUs. One of the challenging topics of Computational Fluid Dynamics is a simulation of multiphase flows such as gas-liquid and solid-gas or solid-liquid two-phase flows. They require high-resolution mesh to describe complex interfaces. Another topic is a dynamic load balance for particle simulations in which the particle distribution is dynamically changing in time and space. Several applications running on TSUBAME2.5/3.0 are demonstrated.

Takayuki Aoki  
Tokyo Institute of Technology  
taoki@gsic.titech.ac.jp

### MS4

#### The Universe in a Computer: How Numerical

### Methods are Essential

We will talk about our contribution to a large project with the goal of a self-consistent numerical simulation of the evolution of the universe beginning soon after the Big Bang and ending with the formation of realistic stellar systems like the Milky Way. This is a multi-scale problem of vast proportions. It requires the development of new numerical methods that excel in accuracy, parallel scalability to the processes relevant in galaxy formation. These numerical methods themselves require the development of mathematical theory in order to guarantee the above mentioned requirements. In this talk we shall focus on our contribution to this effort. To this end we present a positivity preserving, entropy consistent numerical flux. This is being used in a finite volume code for ideal magnetohydrodynamics (MHD), which possesses excellent stability properties. Ingredients are: an approximate Riemann solver, extension to multidimensions via a Powell term, second order preserving positivity. Next we describe a discontinuous Galerkin method where these features have been introduced. Numerical implementations in various astrophysical codes will be shown. They show the overall stability of these schemes. This is joint work among others with Francois Bouchut, Praveen Chandrashekar and Volker Springel.

Christian F. Klingenberg  
Würzburg University, Dept. of Mathematics  
Germany  
klingen@mathematik.uni-wuerzburg.de

### MS4

#### Performance Analysis and Tuning of ExaFSA Codes

Due to increase of complexity of an application code, especially in multi-scale and multi-physics simulation, the maintainability of a code would become significantly important. This work keeps the maintainability by separating system-awareness from an application code using a code transformation framework. From the performance analysis of ExaFSA codes, it is figured out that performance tuning is effective. To keep the maintainability of the original code, performance tunings are expressed as rules of code transformation. Thus, application developers can maintain a single version of an application code and perform code transformation considering a target platform. As a result, high performance can be achieved by keeping the maintainability of the code.

Kazuhiko Komatsu  
Tohoku University  
komatsu@tohoku.ac.jp

### MS4

#### Coupling Parallel Simulation Codes: Problems and Solutions

The increasing computational power of the next generation of supercomputers makes us wish to model our applications more accurately by including more physical effects into existing simulations and resolving larger domains including multi-scale effects. Simply more computational power will, however, not be enough. We also need to deal with the increasing complexity of these models. Writing a new tool from scratch for more and more complex multi-physics or multi-scale simulations might soon become impossible. Reuse and coupling of existing simulation programs is nec-

essary, in particular, as years of effort also went into the efficient parallelization of these existing programs. In this contribution, we discuss how existing simulation programs with possibly different domain decompositions can be coupled in a minimally-invasive way without making the scalability of the overall simulation worse than the individual scalability of each program. The discussion encompasses algorithmic as well as technical challenges.

Benjamin Uekermann  
Department of Informatics  
Technical University of Munich  
uekerman@in.tum.de

Hans-Joachim Bungartz  
Technical University of Munich, Department of Informatics  
Chair of Scientific Computing in Computer Science  
bungartz@in.tum.de

Miriam Mehl  
Universität Stuttgart  
miriam.mehl@ipvs.uni-stuttgart.de

Florian Lindner  
University of Stuttgart  
florian.lindner@ipvs.uni-stuttgart.de

Benjamin Rüth  
Department of Informatics  
Technical University of Munich  
rueth@in.tum.de

Amin Totounferoush  
Institute for Parallel and Distributed Systems  
University of Stuttgart  
amin.totounferoush@ipvs.uni-stuttgart.de

### MS5

#### Dynamic Accuracy Schemes for Deep Neural Networks

In this talk we will identify and exploit opportunities for approximate computing approaches in the machine learning context. We will show to accelerate deep learning approaches by leveraging reduced accuracy schemes and enable improvements in terms of energy consumption and performance. We will demonstrate the algorithmic quality of our dynamic mechanisms to accelerate the Back-Propagation process in a multi-GPU context.

Marc Casas  
Barcelona Supercomputing Center  
marc.casas@bsc.es

### MS5

#### To Reduce Numerical Precision to Achieve Higher Accuracy in Weather and Climate Modelling

Resolution and model complexity of weather and climate models are limited by performance of today's supercomputers. If numerical precision can be reduced, we can reduce power consumption and increase performance. However, it is difficult to identify the optimal level of precision that should be used in weather and climate models that show non-linear dynamics. We investigate a reduction of numerical precision in a range of models (from toy models of atmospheric dynamics to full weather forecast systems) and



study approaches to reduce numerical precision to a level that can be justified by information content and model error, including tests that reduce precision with spatial scale or forecast lead time. We find that numerical precision can indeed be reduced significantly in many components of weather and climate models and calculate measurements for savings that are possible if precision is reduced. For example: Performance of a weather forecast can be increase by 40% when single precision is used instead of double precision producing virtually the same quality in results. Power consumption can be reduced by approximately a factor of three when running a spectral dynamical core with 22 bit floating point numbers instead of 64 bits in double precision. If savings are reinvested to increase resolution and model complexity, predictions for weather and climate can be improved.

Peter D. Dueben  
University of Oxford  
peter.dueben@ecmwf.int

### MS5

#### The Transprecision Computing Paradigm and its Impact

Guaranteed numerical precision of each elementary step in a complex computation has been the mainstay of traditional computing systems for many years. This era, fueled by Moore's law and the constant exponential improvement in computing efficiency, is at its twilight: from tiny nodes of the Internet-of-Things, to large HPC computing centers, sub-picoJoule/operation energy efficiency is essential for practical realizations. To overcome the power wall, a shift from traditional computing paradigms is now mandatory. In this talk we discuss and highlight the potential impact that a transprecision computing framework can have on a large range of application scenarios, spanning in the domains of IoT, Big Data Analytics, Deep Learning, and HPC simulations. By combining together into a seamless design transprecision developments in devices, circuits, software tools, and algorithms, we expect to enable major energy efficiency improvements, even when there is no freedom to relax end-to-end application quality of results.

Cristiano Malossi  
IBM Research - Zurich  
acm@zurich.ibm.com

### MS5

#### Exploiting Auxiliary Floating Point Types to Reduce Total Energy Consumption of Applications with Precision Constraints

In most computing domains, the use of standard floating-point (FP) formats is a widespread solution to design applications characterized by a large dynamic range. Focusing on ultra-low-power (ULP) embedded systems, the adoption of smaller-than-32-bit formats is a major opportunity to reduce energy consumption, since it allows to simplify the arithmetic circuitry and to reduce the memory bandwidth required to transfer data between memory and core registers by enabling vectorization. Similar approaches have been already applied in other contexts (e.g., approximate computing), but the proposed techniques are not suitable for applications that provide strict requirements on the precision of final results. In this scenario, research trends are stepping beyond of the concept of approximation itself, towards a novel paradigm called transprecision computing. In this talk we introduce a methodology to explore alter-

native FP formats and we provide a *transprecision programming flow* to enhance existing applications. First, we present an experimental evaluation that highlights the benefits of introducing auxiliary FP formats. Then, we introduce a set of variable-precision software libraries and precision-tuning tools that enable exploration of FP types by leveraging both precision and dynamic range of program variables. Finally, we provide a transprecision programming flow which integrates the results provided by FP explorations into a compiler toolchain.

Giuseppe Tagliavini, Andrea Marongiu, Davide Rossi, Michele Lombardi, Michela Milano  
University of Bologna  
giuseppe.tagliavini@unibo.it, a.marongiu@unibo.it,  
davide.rossi@unibo.it, michele.lombardi2@unibo.it,  
michela.milano@unibo.it

### MS6

#### Uncertainty Quantification Using Tensor Methods and Large Scale HPC

Abstract Not Available At Time Of Publication.

Lars Grasedyck  
RWTH Aachen  
lgr@igpm.rwth-aachen.de

### MS6

#### Efficient Parallelization by Space-Time Discretization and Solution

We present and discuss different space-time approaches for the numerical solution of linear and non-linear time dependent partial differential equations. We first discuss discretization in space-time by means of Galerkin techniques and exploit this approach for adaptive discretization in space and time. We then present multigrid and domain decomposition techniques for the solution of the arising non-linear systems and discuss their properties. This will include in particular multigrid approaches with stabilization in time on the coarser levels. We will first discuss adaptivity in the setting of a space-time approach for the implementation of Fitz-Hugh Nagumo model in electrophysiology, a non-linear parabolic equation. We will show that our space-time approach is well suited for adaptivity, as it avoids a small global time-step, but instead allows for refining around the resulting travelling-wave solution. We then will present examples from computational mechanics, including thermal, non-linear mechanical and fluid systems and will show how the stabilized multigrid method behaves in these cases. In addition we discuss implementation aspects and explain how the Galerkin framework allows for an efficient parallelization of the assembly process. We finally comment on the convergence of Newton's method when applied to the solution of the non-linear space-time systems.

Rolf Krause  
Università della Svizzera italiana, Switzerland  
rolf.krause@usi.ch

Seif Ben Bader  
Institute of Computational Science  
USI, Lugano  
ben.bader.seif@seif.ben.bader@usi.ch

Marco Favino  
Institute of Computational Science

Università della Svizzera italiana  
marco.favino@usi.ch

Christian Hesch  
University of Siegen  
hesch christian |christian.hesch@uni-siegen.de|

Alessio Quaglino  
Institute of Computational Science  
Università della Svizzera italiana  
alessio.quaglino@usi.ch

## MS6

### Scalable Shape Optimization Methods for Structured Inverse Modeling Using Large Scale HPC

Abstract Not Available At Time Of Publication.

Volker H. Schulz  
University of Trier  
Department of Mathematics  
Volker.Schulz@uni-trier.de

## MS6

### Parallel Adaptive and Robust Multigrid

Abstract Not Available At Time Of Publication.

Gabriel Wittum  
KAUST  
Wittum@gcsc.uni-frankfurt.de

## MS7

### Title Not Available At Time Of Publication

Abstract Not Available At Time Of Publication

Frank Cappello  
Argonne National Laboratory  
cappello@mcs.anl.gov

## MS7

### Data Analytics and Machine Learning for Supercomputer Datacenter Design

In this talk, we will present recent results from data analytics and machine learning based on field data from some of the most powerful supercomputers in the world. These insights are used by Los Alamos National Laboratory to drive policies on purchasing new systems, maintaining systems in production, and driving future system / component design with semiconductor chip vendors. Subsystems to be discussed include supercomputer memory, storage, and job schedulers.

Nathan A. DeBardeleben, Elisabeth Baseman  
Los Alamos National Laboratory  
ndebard@lanl.gov, lissa@lanl.gov

## MS7

### Resilience Design Patterns: A Structured Approach to Resilience at Extreme Scale

Reliability is a serious concern for future extreme-scale high-performance computing (HPC). While the HPC community has developed various resilience solutions, the solution space remains fragmented. With this work, we de-

velop a structured approach to the design, evaluation and optimization of HPC resilience using the concept of design patterns. We identify the problems caused by faults, errors and failures in HPC systems and the techniques used to deal with these events. Each well-known solution that addresses a specific resilience challenge is described in the form of a pattern. We develop a catalog of such resilience design patterns, which may be used by system architects, system software and tools developers, application programmers, as well as users and operators as essential building blocks when designing and deploying resilience solutions. We also develop a design framework that enhances a designer's understanding the opportunities for integrating multiple patterns across layers of the system stack and the important constraints during implementation of the individual patterns. It is also useful for designing mechanisms and interfaces to coordinate flexible fault management across hardware and software components. The resilience patterns and the design framework also enable exploration and evaluation of design alternatives and support optimization of the cost-benefit trade-offs among performance, protection coverage, and power consumption of resilience solutions.

Saurabh Hukerikar, Christian Engelmann  
Oak Ridge National Laboratory  
hukerikarsr@ornl.gov, engelmanncc@ornl.gov

## MS7

### Accurate and Portable Compiler-based Fault Injection in HPC Programs

Quantifying the effect of soft errors and the effectiveness of resilience techniques requires portable and accurate fault injection (FI) methods. Compiler-based FI frameworks have become popular in resilience studies as they can be ported to multiple platforms and are easy to combine with static error-propagation analysis. Most existing compiler-based FI tools, however, fail to realistically inject faults in all locations of a program (memory and registers) since they operate on a high-level representation of the program; this can result in significant inaccuracies when error manifestations (e.g., crashes and silent corruptions) are quantified. In this work, we present a compiler-based FI framework, called REFINE, that addresses the existing problems of current frameworks by performing injections at the backend of a compiler, rather than in the intermediate-representation where most tools operate. We demonstrate that this approach is comparable in accuracy with respect to low-level binary FI while maintaining the portability benefits of compiler-based FI.

Ignacio Laguna  
Lawrence Livermore National Laboratory  
ilaguna@llnl.gov

Giorgis Georgakoudis  
Queen's University of Belfast  
g.georgakoudis@qub.ac.uk

Dimitrios S. Nikolopoulos  
Queen's University of Belfast  
Belfast, NI  
d.nikolopoulos@qub.ac.uk

Martin Schulz  
Lawrence Livermore National Laboratory

schulzm@llnl.gov

**MS8**

### Improving Parallel Structured Multigrid Methods with Block Smoothers

While multigrid methods show an excellent scaling behavior depending on the number of cores only logarithmically, on large-scale supercomputers this logarithmic dependency is visible. To reduce influence of this dependence, the amount of work being carried out on the coarse levels has to be reduced. This can be achieved by applying aggressive coarsening, effectively reducing the number of coarse levels. While this is easily possible in a geometric multigrid setting, the overall performance of the method will deteriorate, as the size of the coarse space is reduced. To retain a good convergence rate, the smoothing procedure has to be improved. We propose to use block smoothers to accomplish this, at the same time this results in a higher locality of the operations performed and thus in a better exploitation of modern computer architectures.

Matthias Bolten

Bergische Universität Wuppertal  
Fakultät für Mathematik und Naturwissenschaften  
bolten@math.uni-wuppertal.de

**MS8**

### Multigrid Smoothers for the Stokes Problem

Multigrid methods are efficient iterative solvers for the solution of partial differential equations (PDEs). The efficiency of multigrid methods is due to the combination of suitable smoothers with a coarse grid correction. As one of the two key ingredients smoothers have a significant impact on the performance of multigrid solvers. Most of the literature on smoothers in multigrid methods is concerned with scalar PDEs, only. Systems are considered less often. In this talk, we present a comparative study of several smoothers for multigrid methods for the solution of the Stokes equations. Besides the commonly used Vanka smoother and the Braess-Sarazin smoother, we also consider a non-overlapping variant of the Vanka smoother. While the latter is computationally cheaper, the convergence depends much more on the implementation than that of the overlapping method. We consider discretization using appropriate finite elements as well as by finite differences on staggered grids. A comparison including the computational cost and the convergence properties of the different methods will be presented.

Lisa Claus

Bergische Universität Wuppertal  
lisa.claus@math.uni-wuppertal.de

**MS8**

### The Analysis for Robustness of Sa-Amg Method by Extraction of Near-Kernel Vectors

The smoothed aggregation algebraic multigrid (SA-AMG) method is among the fastest solver for large-scale linear equation  $Ax = b$ . The SA-AMG method achieves good convergence by interpolating a long wave component efficiently. To interpolate this, this method generates small-sized matrices from the original matrix problem. Moreover, the convergence can be further improved by setting near-kernel vectors. The near-kernel vector is a vector  $p$  that is satisfied  $Ap \approx 0, p \neq 0$ . Almost all previous works

use the same number of near-kernel vectors at all levels. In the present work, we propose the extraction method of near-kernel vectors at coarser levels. We evaluate the performance of the solver that extracts near-kernel vectors and add them at each level. By using this method, the performance is improved. But, setting many near-kernel vectors did not always improved. This is a trade-off between the increasing cost and the improvement of the convergence by using near-kernel vectors at coarser levels. And also, we evaluate the method that estimates the suitable number of near-kernel vectors. This method can estimate the suitable number of near-kernel vectors approximately.

Naoya Nomura

The University of Tokyo  
naoya\_nomura@mist.i.u-tokyo.ac.jp

**MS8**

### Structured Multigrid at Extreme Scales

Large-scale high-performance computing systems continue to challenge the parallel scalability of multilevel solvers. Algebraic multigrid methods are robust for a variety of problems, yet data locality and the cost of data movement in modern architectures is motivating the need to exploit structure in the problem. Robust logically structured variational multigrid methods, such as Black Box Multigrid (BoxMG), maintain structure throughout the multigrid hierarchy. Locality is thus retained at coarse levels, yet the parallel scalability is limited by coarse-grid problems where the overhead in communication dominates computation. We present an algorithm for redistributing coarse-grid problems through incremental agglomeration. A predictive performance model is used to provide effective redistribution decisions for structured multilevel solvers. The parallel scalability of this approach is demonstrated on two large-scale computing systems, with solves over 500K+ cores.

Luke Olson, Andrew Reisner

University of Illinois Urbana-Champaign  
lukeo@illinois.edu, lukeo@illinois.edu

David Moulton

Los Alamos National Laboratory  
Applied Mathematics and Plasma Physics  
moulton@lanl.gov

**MS9**

### High-performance Dense Tucker Decomposition on GPU Clusters

The Tucker method is one of the most popular tensor decomposition algorithms currently in use, and its computation on dense data sets is largely dominated by matrix-matrix multiplication (GEMM). As such, GPUs would be highly suited for accelerating dense Tucker decomposition, and should yield a significantly higher performance than traditional CPUs. We methodically study and implement highly optimized dense Tucker decomposition on a cluster of multi-GPU nodes. First, our study shows that the current data partitioning method is ill-suited for GPU acceleration, as it reduces the matrix size, thereby reducing the achievable performance on GPUs. Secondly, in order to maintain a large matrix size to achieve high GPU efficiency, redistribution of the tensor data is required for every mode. This requires costly gather/scatter and MPI communication, particularly when the number of tensor modes is large. Lastly, the SVD operation begins to dominate the total ex-

ecution time as the number of nodes grows. We propose the following optimizations to mitigate these bottlenecks: slice block partitioning, tensor re-use, and randomized SVD algorithm. For a 3D tensor of dimension 3000x3000x3000 and rank of 300 (a 216 GB data set), our implementation utilizing four GPUs per node achieves up to 14.4x speedup on a single node, and up to 4x speedup when scaled to 64 nodes, over an implementation running on two CPUs per node.

Jee Choi  
IBM TJ Watson  
jwchoi@us.ibm.com

Xing Liu  
IBM  
xing.research@gmail.com

## MS9

### Computing Sparse Tensor Decompositions using Dimension Trees

Tensor decompositions have increasingly been used in many applications domains including computer vision, signal processing, graph analytics, healthcare data analysis, recommender systems, and many other related machine learning problems. In many of these applications, the tensor representation of data is big, sparse, and of low-rank, and an efficient computation of tensor decompositions is indispensable for analyzing datasets of massive scale. Two major computational kernels involved at the core of the iterative tensor decomposition algorithms are tensor-times-vector (TTV) and -matrix (TTM) multiplications. In this talk, we propose a novel algorithmic framework along with an associated sparse tensor data structure that enables an efficient computation of these two operations for high dimensional sparse tensors. With this framework, we enable storing and reusing partial TTV and TTM results, thereby reducing the cost per iteration to  $O(N \log N)$  TTM/TTV calls from  $O(N^2)$  in the traditional scheme. Second, we propose an efficient tensor storage scheme representing an  $N$ -dimensional tensor using  $O(N \log N)$  index arrays, and enabling the partial computations in this framework. Third, we propose shared memory parallelization of TTV and TTMs using this sparse data representation. With all these optimizations, we up to 6x speedup over the state of the art implementations using 32 dimensional sparse tensors.

Oguz Kaya  
INRIA Bordeaux  
oguz.kaya@outlook.com

Bora Ucar  
LIP, ENS-Lyon, France.  
Bora.Ucar@ens-lyon.fr

## MS9

### A Sparse Tensor Algebra Compiler

Tensor decomposition algorithms often build on basic tensor algebra kernels, such as tensor contractions and the matricized tensor times Khatri-Rao product. When the tensors are sparse, which means most components are zeros, these kernels must be written to work on sparse tensor formats. There are many such formats, many possible tensor expressions, and many target platforms. It is therefore hard for library developers to implement and optimize all the necessary kernels. I will present the Tensor Alge-

bra Compiler (taco), which automatically generates dense, sparse, and mixed linear and tensor algebra kernels. It is implemented as a C++ library that compiles and executes any tensor expression on tensors stored with user-defined formats, and has performance competitive with hand-optimized libraries, while supporting far more tensor computations. I will also discuss taco's capabilities for generating parallel kernels. For more information, see [tensor-compiler.org](http://tensor-compiler.org).

Fredrik Kjolstad  
Massachusetts Institute of Technology  
fred@csail.mit.edu

Shoaib Kamil  
Adobe Systems Inc  
kamil@adobe.com

Stephen Chou  
Massachusetts Institute of Technology  
s3chou@csail.mit.edu

David Lugato  
CEA  
david.lugato@cea.fr

Saman Amarasinghe  
Massachusetts Institute of Technology  
saman@csail.mit.edu

## MS9

### HiCOO: Hierarchical Storage of Sparse Tensors

This paper proposes a new storage format for sparse tensors, which we call Hierarchical COordinate (HICOO; pronounced: haiku). It derives from coordinate (COO) format, arguably the de facto standard for general sparse tensor storage. HICOO improves upon COO by compressing the indices in units of sparse tensor blocks, with the goals of preserving the mode-agnostic simplicity of COO while reducing the bytes needed to represent the tensor and promoting data locality. We evaluate HICOO by implementing a single-node, multicore-parallel version of the matricized tensor-times-Khatri-Rao product (MTTKRP) operation, which is the most expensive computational core in the widely used CANDECOMP/PARAFAC decomposition (CPD) algorithm. This MTTKRP implementation achieves up to  $5.3\times$  ( $2.5\times$  on average) speedup over COO format and up to  $4.3\times$  ( $1.5\times$  on average) speedup over another state-of-the-art format, compressed sparse fiber (CSF). When used within CPD, we also observe speedups against COO- and CSF-based implementations.

Jiajia Li, Jimeng Sun  
College of Computing  
Georgia Institute of Technology  
jiajiali@gatech.edu, jsun@cc.gatech.edu

Richard Vuduc  
Georgia Institute of Technology  
richie@cc.gatech.edu

## MS10

### Rethinking Performance in HPC Systems: Monitoring, Analytics, and Resource Management

As the HPC applications and systems evolve, users have started observing significant performance variations across

runs—even when running the exact same application with the same configuration. A number of these variations result from contention in shared resources. Root-causing these variations and preventing them with smarter resource management methods are therefore becoming essential to reach the ambitious exascale computing goals. This talk will first introduce an automated, machine learning based framework that is capable of diagnosing known system and application anomalies that create performance variations. It will then demonstrate new observations on how task allocation and mapping can make a substantial difference in the performance of new HPC machines with dragonfly networks. Finally, the talk will present insights on designing new heuristics for managing resources of future HPC systems. Results on real-world HPC machines show highly promising results on both accurate diagnosis of performance anomalies and on improving performance via better task management.

Ayşe Coskun  
Boston University  
Electrical and Computer Engineering Department  
acoskun@bu.edu

#### MS10

##### **Energy and Time-to-Solution Optimization with GEOPM**

This talk discusses the role of the Global Extensible Open Power Manager (GEOPM) runtime framework in implementing a scalable and flexible stack for resource and power management in HPC systems. It discusses the importance of adapting resource and power management decisions on-the-fly through application profiling. Finally, this talk overviews the long-term vision for GEOPM which includes coordinating optimization of software parameters in the OpenMP, MPI, and application layers with optimization of hardware control knob settings.

Jonathan Eastep  
Intel Corporation  
jonathan.m.eastep@intel.com

#### MS10

##### **The Interplay of Workflow Execution and Resource Provisioning**

This talk will examine issues of workflow execution, in particular using the Pegasus Workflow Management System, on distributed resources and how these resources can be provisioned ahead of the workflow execution. Pegasus was designed, implemented and supported to provide abstractions that enable scientists to focus on structuring their computations without worrying about the details of the target cyberinfrastructure. To support these workflow abstractions Pegasus provides automation capabilities that seamlessly map workflows onto target resources, sparing scientists the overhead of managing the data flow, job scheduling, fault recovery and adaptation of their applications. In some cases, it is beneficial to provision the resources ahead of the workflow execution, enabling the re-use of resources across workflow tasks. The talk will examine the benefits of resource provisioning for workflow execution.

Rafael Ferreira da Silva  
USC Information Sciences Institute

rafsilva@isi.edu

#### MS10

##### **Enabling Hierarchical Scheduling in Next-Generation HPC Systems with Flux**

The landscape of supercomputing systems is changing as we move from the petascale era to the exascale era. Resources as well as workload types on shared clusters are becoming more complex than ever. Simultaneously managing diverse resources and workload types at large scales poses several scheduling optimization challenges. Despite this, the resource management and job scheduling software that is used in production supercomputers remains stuck with the old approaches. In this talk, I will address some key challenges in resource management for next-generation supercomputers: scalability, flexibility for flow resources (such as power, network and I/O), and workload customization. I will present Flux, a next-generation production resource manager under development at Lawrence Livermore National Laboratory, which is based on a fully hierarchical scheduling model. Flux provides a framework for addressing the aforementioned challenges, and looks beyond traditional approaches that have focused primarily on using centralized models and static resource allocations.

Tapasya Patki  
Lawrence Livermore National Laboratory  
patki1@llnl.gov

#### MS11

##### **Fast Finite Element Analyses for Electric Machines by Using Massively Parallel Processing**

We present the parallel performance of massively parallel processing in finite element electromagnetic field analyses (FEA). The domain decomposition technique is used for parallelization, and the Message Passing Interface (MPI) library is used for processing communications. These techniques are very effective for speeding up FEA, and results show over a several tens of times speed increase compared with the sequential solver. This speed increase contributes to designing electric machines.

Kazuki Semba  
JSOL  
semba.kazuki@jsol.co.jp

Hirokatsu Katagiri, Tatsuya Asanuma, Masahiko Miwa,  
Hiroyuki Sano, Takashi Yamada  
JSOL Corporation  
semba.kazuki@jsol.co.jp, semba.kazuki@jsol.co.jp,  
semba.kazuki@jsol.co.jp, semba.kazuki@jsol.co.jp,  
semba.kazuki@jsol.co.jp

#### MS11

##### **Analysis of Rotating Machines by Hierarchical Domain Decomposition Method**

Rotating machines such as an electric generator or motors is representative. Such devices have become essential in our lives; thus, the development of an efficient rotating machine has become necessary to reduce cost and environmental loading. Designing rotating machines includes electromagnetic analysis, which takes substantial time; thus, methods for reducing the time steps, such as TP-EEC and TDC, and for reducing the computation time, such as the parallel iterative method, have been proposed. However, these



methods cannot utilize massively parallel computers efficiently, and still take substantial time to analyze a rotating machine. However, for non-linear magnetostatic problems, time-harmonic eddy current problems, and high-frequency electromagnetic problems, problems with billions of degrees of freedom have been solved efficiently on the massively parallel computers using the hierarchical domain decomposition method (HDDM). Nevertheless, in cases that involve a moving body, the HDDM with static domain decomposition has not attained sufficient parallel efficiency. Moreover, the cost of dynamic domain decomposition is substantial. In this presentation, a new domain decomposition technique for the hierarchical domain decomposition method for analyzing devices including moving bodies efficiently on massively parallel computers is described. Then, the computational efficiency of the proposed method is shown through analyses of rotating machine models.

Shin-ichiro Sugimoto  
Tokyo University of Science, Suwa  
sugimoto@rs.tus.ac.jp

#### MS11

##### **Parallel-in-Time Magnetic Field Analysis of Electric Machines Based on Time-Periodic Finite-Element Method and Time-Periodic Explicit Error Correction Method**

In this study, two types of parallel-in-time finite-element methods are compared in magnetic field analyses of practical electric machines. One is called parallel time-periodic finite-element method (TPFEM) and the other is called parallel time-periodic explicit-error-correction (TP-EEC) method. The parallel TPFEM provides large granularity in parallel computing by treating all nonlinear systems of equations at every time step for a period simultaneously. It has a great advantage in highly parallel computation environment because of its good scalability even in the small-scale problems. In contrast, the transient calculations are performed independently in each process in the parallel TP-EEC method. By solving the auxiliary system of linear equations derived from the polyphase TP-EEC method, poorly-converged error components are extracted. As a consequence, the initial values in each process can be corrected and steady-state solutions can be successfully obtained. The parallel TPFEM is useful in highly parallel computation environments, especially when as many processes as possible can be used compared with the number of time steps per (half) period. On the other hand, the parallel TP-EEC method is suitable for problems with a large number of time steps. In this study, the parallel performance of these two approaches are discussed in the magnetic field analysis of an interior permanent magnet synchronous motor driven by pulse-width-modulation inverter.

Yasuhito Takahashi  
Doshisha University  
ytakahashi@mail.doshisha.ac.jp

#### MS11

##### **Parallelization of Preconditioned Krylov Subspace Method Using Block-Multicolor Ordering in Linear Systems Derived from Finite Element Method**

Fast computation of a linear system is essential for further advance in electromagnetic field analysis based on the finite element method (FEM). The parallelization of preconditioned Krylov subspace method is suitable to reduce

the elapsed time. The multicolor (MC) ordering has been applied to parallelization of the forward and backward substitution. However, because the deterioration of cache-hit ratio is occurred owing to the enlargement of bandwidth, the elapsed time using MC would be longer than the case with block preconditioner. To realize further speed-up, the block MC supported by a level structure arising in the reverse Cuthill-McKee (RCM) ordering was newly proposed by the authors. The name of proposed method is abbreviated as PLES. The PLES ordering is implemented through the following steps: 1) the level structure is constructed by RCM ordering; 2) the new block number are determined from the level structure so that the number of unknowns in one block can be uniform; and 3) coloring and renumbering process is performed. When the PLES is used as matrix ordering, the bandwidth could be reduced in comparison with MC. In this paper, the validity of PLES ordering is demonstrated in the parallelized incomplete-Cholesky-preconditioned conjugate gradient (ICCG) method. Furthermore, the affinity of three parallelization techniques (block IC with RCM, MC, PLES) on real symmetric linear systems derived from the edge-based FEM is particularly presented.

Tomonori Tsuburaya  
Fukuoka University  
tsuburaya@fukuoka-u.ac.jp

Yoshifumi Okamoto  
Hosei University  
okamotoy@hosei.ac.jp

#### MS12

##### **Petascale Earthquake Simulation Using High Order Methods**

Earthquake simulation requires to cover multiple scales, as the triggering dynamic rupture processes occur on a meter scale while generated seismic waves may travel over 100s or 1000s of kilometers. Petascale supercomputers nowadays allow the simulation of highly detailed models - we will present the recent simulation of the Sumatra 2004 earthquake using the earthquake simulation software SeisSol as an example. High order discontinuous Galerkin methods have been shown to provide a good combination of high accuracy and efficiency of the numerical scheme with large enough arithmetic intensity to allow compute-bound performance on latest supercomputing architectures. We will discuss recent advances on highly optimized and scalable DG algorithms in the SeisSol software and in ExaHyPE, a new engine for hyperbolic PDE systems.

Michael Bader  
Technical University of Munich  
bader@in.tum.de

#### MS12

##### **Advanced Load-Balancing Techniques for Massive Parallel Flood Simulations**

As current events reveal, the simulation and accurate prediction of massive scale natural disasters becomes increasingly important to ensure the safety of large populations. Within our special interest of high-performance computing for flooding events, we propose a new strategy for a decentral management of the simulation domain, allowing the computation of even city sized areas. Key characteristic of this approach is a domain decomposition into individual sub-domains without global domain knowledge and a

diffusion load-balancing.

Christoph M. Ertl  
Technical University of Munich  
Chair for Computation in Engineering  
christoph.ertl@tum.de

Nevena Perovic  
Chair for computation in engineering  
TU München  
nevena.perovic@tum.de

Bobby Minola Ginting  
Technical University of Munich  
Chair for Computation in Engineering  
bobbyminola.ginting@tum.de

Ralf-Peter Mundani  
TUM, Faculty of Civil, Geo and Environmental  
Engineering  
Chair for Computation in Engineering  
mundani@tum.de

Ernst Rank  
Computation in Engineering  
Technische Universität München  
ernst.rank@tum.de

## MS12

### Large Scale Numerical Simulations of Geothermal Potentials Using a Distributed Parallel Data Structure

Severe climate changes in the last decades illustrate that a paradigm shift towards regenerative energy generation is necessary. Geothermal energy in combination with heat pumps pose a huge potential for domestic heat generation. However, the simulation of geothermal potential is a non-trivial task due to problems ranging from large domain sizes to settings of boundary conditions. This presentation highlights possibilities of applying a massive parallel computational approach for evaluating geothermal potentials in urban districts.

Jérôme Frisch  
Institute of Energy Efficiency and Sustainable Building  
E3D  
RWTH Aachen University  
frisch@e3d.rwth-aachen.de

Ralf-Peter Mundani  
TUM, Faculty of Civil, Geo and Environmental  
Engineering  
Chair for Computation in Engineering  
mundani@tum.de

## MS12

### Large Eddy Simulation of Thermal Atmospheric Environment in Urban Boundary Layer

Since urban atmospheric environment is strongly controlled by multi-scale flow dynamics, simulation of the urban atmospheric boundary layer requires fine grid spacing and huge computational domain. We accomplished to simulate an urban atmospheric boundary layer using a large eddy simulation model running on TSUBAME super computing system. Flow characteristics within and above a building

canopy were successfully examined.

Atsushi Inagaki  
Tokyo Institute of Technology  
inagaki.a.ab@m.titech.ac.jp

Naoyuki Onodera  
Center for Computational Science & e-Systems  
Japan Atomic Energy Agency  
onodera.naoyuki@jaea.go.jp

Manabu Kanda, Takayuki Aoki  
Tokyo Institute of Technology  
kanda.m.aa@m.titech.ac.jp, taoki@gsic.titech.ac.jp

## MS13

### Fast Fourier Transforms (fft)

This talk will address a broad introductory about the Fast Fourier Transforms starting from its mathematical concept and how it evolved to become an essential algorithm that is widely used in scientific applications. It has been considered as one of the great computational developments over the years. A quick overview about the FFTW (or the Fast Fourier Transform in the West) software library packager will be addressed and some other FFT packages will be highlighted.

Samar A. Aseeri  
Computational Scientist at KAUST  
Extreme Computing Research Center (ECRC)  
samar.aseeri@kaust.edu.sa

## MS13

### Use of the Fast Fourier Transform in Solving Partial Differential Equations

The Fourier transform is used in the solution of partial differential equations due to its low operation count and high accuracy when calculating derivatives. The efficiency of this method is discussed for a number of semi-linear partial differential equations, with particular focus on nonlinear Schrödinger equations and the Navier-Stokes-Boussinesq equations which model convection. Efficiency of time stepping procedures and comparison with other numerical methods for solving the same equations on parallel computers will also be discussed.

Benson K. Muite  
Department of Mathematics  
University of Michigan  
benson.muite@ut.ee

## MS13

### Parallel Fast Gauss Transform

Abstract: We will discuss a fast adaptive parallel algorithm to compute the “Gaussian-type” transforms. The parallel time complexity estimates for our algorithms are  $\mathcal{O}(N/n_p)$  for uniform point distributions and  $\mathcal{O}(N/n_p \log N/n_p + n_p \log n_p)$  for nonuniform distributions using  $n_p$  CPUs. We incorporate a plane-wave representation of the Gaussian kernel which permits diagonal translation. We use parallel octrees and a new scheme for translating the plane-waves to efficiently handle nonuniform distributions. We will discuss several applications.

Shravan Veerapaneni

Department of Mathematics  
University of Michigan  
shravan@umich.edu

### MS13

#### Implementation of OpenFFT and Its Application to Industrial Problems

The fast Fourier transform (FFT) is a primitive kernel in numerous fields of science and engineering. The FFT is usually performed with a large data set in multiple dimensions, making it a computationally expensive calculation. OpenFFT is an open-source parallel package for 3-D FFTs, built on a communication-optimal domain decomposition method for achieving minimal volume of communication. The development of OpenFFT and its applications in some problems will be presented and discussed in the presentation.

Truong Vinh Truong Duy  
Nissan ARC, Ltd.  
duy-truong@nissan-arc.co.jp

### MS14

#### Batched Triangular Dense Linear Algebra for Very Small Matrices on GPUs

Batched dense linear algebra (DLA) kernels are ubiquitous in scientific applications, like tensor contractions in deep learning and data compression in hierarchical low rank matrix approximation. Batch calls remove the expensive overhead of multiple API calls while increasing the occupancy of the underlying hardware. We describe the design and performance of a class of batched triangular DLA kernels on very small data sizes using NVIDIA GPUs. By deploying recursive formulations, stressing the register usage, maintaining data locality, and reducing threads synchronization, thanks to CUDA shuffle instructions, the new batched kernels outperform existing state-of-the-art implementations on both x86 CPU and GPU architectures.

Ali M. Charara, David E. Keyes  
KAUST  
ali.charara@kaust.edu.sa, david.keyes@kaust.edu.sa

Hatem Ltaief  
King Abdullah University of Science & Technology (KAUST)  
hatem.Ltaief@kaust.edu.sa

### MS14

#### Batched DGEMM Operations in Density Matrix Renormalization Group

The Lanczos algorithm for computing the lowest eigenvector of the Hamiltonian matrix is a computational kernel in DMRG (Density Matrix Renormalization Group) for modeling correlated materials. This matrix is block partitioned where each submatrix is a sum of Kronecker products of various sizes. The Kronecker products ( $\text{vec}(Y) = \text{kron}(A, B) * \text{vec}(X) = B * X * A^t$ ) can be implemented as DGEMM operations. This is a use case for batch DGEMM of different matrix sizes on GPU and multi-core CPUs. Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725. Notice: "This manuscript has

been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>)."

Eduardo F. D'Azevedo  
Oak Ridge National Laboratory  
Computer Science and Mathematics Division  
dazevedoef@ornl.gov

Arghya Chatterjee  
School of Computer Science  
Georgia Institute of Technology  
arghya@gatech.edu

Wael R. Elwasif  
Oak Ridge National Laboratory  
elwasifwr@ornl.gov

Oscar Hernandez  
Oak Ridge National Lab  
oscar@ornl.gov

Gonzalo Alvarez  
Computational Sciences & Engineering Division  
Oak Ridge National Laboratory  
alvarezcampg@ornl.gov

### MS14

#### Magma Batched Computations: Current Development and Trend

There's an increasing demand for accelerated independent computations on tensors and many small matrices. Although common, these workloads cannot be efficiently executed using standard linear algebra libraries. To fill the gap, we developed the MAGMA Batched library that achieves dramatically better performance by repetitively executing the small operations in "batches." We'll describe a methodology on how to develop high-performance BLAS, SVD, factorization, and solvers for both large- and small-batched matrices. We'll also present the current state-of-the-art implementations and community efforts to standardize an API that extends BLAS for Batched computations.

Azzam Haidar  
Department of Electrical Engineering and Computer Science  
University of Tennessee, Knoxville  
haidar@icl.utk.edu

Stan Tomov, Ahmad Ahmad  
University of Tennessee  
tomov@eecs.utk.edu, ahmad@icl.utk.edu

Jack J. Dongarra  
University of Tennessee, Oak Ridge National Laboratory, USA

dongarra@icl.utk.edu

#### MS14

##### **CuHinesBatch, a Cuda Implementation to Solve Batches of Hines Systems. Human Brain Project**

The simulation of the behavior of the Human Brain is one of the most important challenges today in computing. The main problem consists of finding efficient ways to manipulate and compute the huge volume of data that this kind of simulations need, using the current technology. In this sense, this work is focused on one of the main steps of such simulation, which consists of computing the Voltage on neurons morphology. This is carried out using the Hines Algorithm. Although this algorithm is the optimum method in terms of number of operations, it is in need of non-trivial modifications to be efficiently parallelized on NVIDIA GPUs. We proposed several optimizations to accelerate this algorithm on GPU-based architectures, exploring the limitations of both, method and architecture, to be able to solve efficiently a high number of Hines systems (neurons). Each of the optimizations are deeply analyzed and described. To evaluate the impact of the optimizations on real inputs, we have used 6 different morphologies in terms of size and branches. Our studies have proven that the optimizations proposed in the present work can achieve a high performance on those computations with a high number of neurons, being our GPU implementations about 4 and 8 faster than the OpenMP multicore implementation (16 cores), using one and two K80 NVIDIA GPUs respectively. Also, it is important to highlight that these optimizations can continue scaling even when dealing with number of neurons.

Pedro Valero Lara  
Barcelona Supercomputing center  
pedro.valero@bsc.es

#### MS15

##### **Implementation Techniques of Domain-Specific Languages for Parallel Solvers**

This presentation will introduce our study on implementation techniques of domain-specific languages (DSLs) embedded in their host languages. Such a DSL is often called library-based DSLs since they can be regarded as normal libraries written in the host languages. These libraries provide their functionalities through language-like programming interfaces and hence they are preferable from the users' perspectives. However, they have been known as being less efficient than stand-alone DSLs. An implementation technique called deep embedding mitigates this drawback but makes the DSL implementation complicated. We are developing a better implementation technique through development of our parallel-solver DSL.

Shigeru Chiba  
Department of Creative Informatics  
University of Tokyo  
chiba@acm.org

#### MS15

##### **Parallel Implicit DG Algorithms Without Linear System Inversion. Applications to Plasma Physics**

The Palindromic Discontinuous Galerkin (PDG) method is a general approach for solving hyperbolic systems of conservation laws. In many applications the time scale of

the interesting phenomena is very different from the time scale imposed by the explicit CFL condition. The PDG method is a general implicit (but matrix-free) high order method for approximating systems of conservation laws. It is unconditionally stable and has the complexity of an explicit scheme. It relies on a vectorial kinetic interpretation of the conservation laws proposed in [Aregba, Natalini, Discrete kinetic schemes for multidimensional systems of conservation laws, 2000]. The kinetic system is approximated with an asymptotic-preserving high order DG method. The method is well adapted to parallel optimizations. We will review the task-based implementation of the method, based on the StarPU runtime library, and some applications to fluid mechanics and plasma physics. Some results of the presentation are also given in [Badwaik & *al.*, Task-based parallelization of an implicit kinetic scheme, 2017].

Philippe Helluy  
Scientific Computing  
University of Strasbourg  
philippe.helluy@unistra.fr

#### MS15

##### **Generation of Highly Parallel Multigrid Solvers for CFD Applications**

Recent advances in computer hardware technology make more and more realistic simulations possible; however, this progress has a price. On the one hand, the growing complexity of the physical and mathematical models requires the development of new and efficient numerical methods. On the other hand, the trend towards heterogeneous and highly parallel architectures increases the programming effort necessary to implement, develop, and maintain these models. ExaStencils addresses these issues by providing a multi-layered domain specific language that enables the users to formulate their problems on different levels of abstraction. From these formulations efficient implementations can be generated and optimized automatically since the data layout is known. First, we restricted ourselves to regular grids and then the data layout was step by step extended to be able to handle more complex grids like staggered grids, locally refined grids near the boundaries or currently also certain kinds of block-structured grids. In addition we started to look also on more complex applications in CFD and geosciences. In this talk we show that our approach enables us to obtain high computational performance at low effort to implement new methods and applications. Thus, it becomes possible to generate scalable and efficient code also for non-trivial applications.

Harald Koestler  
University Erlangen-Nürnberg  
harald.koestler@fau.de

Sebastian Kuckuk  
Universitaet Erlangen-Nuernberg, Germany  
sebastian.kuckuk@fau.de

#### MS15

##### **Simulating Cosmic Structure Formation**

We present the results from a series of large-scale numerical simulations of self-gravitating systems. We integrate directly the collisionless Boltzmann equation in a six-dimensional phase-space. Our numerical scheme is based on the 5th- and 7-th order monotonicity and positivity preserving scheme for advection equations. The latter feature

is crucial for our simulations that treat the velocity distribution function explicitly. We have successfully performed simulations on a 128 to the power 6th grid. In this talk, we show the results of a test suite and also compare some of them with the results of conventional particle simulations. Finally, we show a particular application of the large-scale distribution of cosmic relic neutrinos, whose results will be relevant to a major goal of the next generation cosmology surveys.

Naoki Yoshida

Department of Physics

University of Tokyo

naoki.yoshida@phys.s.u-tokyo.ac.jp

## MS16

### Exploiting Extended Hardware Margins to Improve Energy Efficiency

Modern processors are inherently heterogeneous, due to manufacturing variability. Even within the same chip, different cores have different minimum error-free operating points in terms of voltage and frequency. Hardware manufacturers typically follow a conservative approach, specifying the same operating point for all components of the same family, plus guardbands to guarantee error-free operation under adverse combinations of conditions. In this work we discuss how we can embrace this heterogeneity in the quest for more energy-efficient computing. We present characterization results which indicate the extent of operating margins present in commercially available parts. Then, we introduce policies for just-right system configuration for power efficiency, problematic resources isolation, and reliability-aware scheduling. Finally, we characterize the sensitivity to errors and the criticality of data structures and code of the operating system. Such efforts are necessary to support next-generation HPC, datacenters, internet and applications. These systems will unavoidably have to tolerate errors, due to the extreme scale of devices, the operation in sub-optimal environmental and power conditions and, potentially, the operation with less pessimistic voltage / frequency margins to meet stringent power limitations.

Christos D. Antonopoulos

Department of Computer Science

College of William and Mary

cda@e-ce.uth.gr

## MS16

### Adjoint Methods for Stochastic Approximate Computing

We consider numerical simulation programs implementing multivariate (without loss of generality) scalar functions  $y = f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ . For given mean and variance (potentially higher-order moments) of randomly distributed inputs  $\mathbf{x}$  propagation of second- (potentially higher-) order Taylor expansions yields approximations of the moments of the correspondingly random result  $y$ . Specification of a confidence interval for the result yields an inverse problem for the corresponding argument intervals. Solutions to this inverse problem for given inputs yield regions within the input space for which the corresponding approximate simulation results can be predicted statically with the given level of confidence. First- and second-order algorithmic adjoints are used to approximate the moments as well as the solution of the inverse problem by a nonlinear programming method. The evaluation of second- and higher-order

moments can make good use of massive parallelism in most relevant cases. We outline the idea behind our approach illustrated by preliminary numerical results.

Uwe Naumann, Jens Deussen, Jonathan Hüser

RWTH Aachen University

Software and Tools for Computational Engineering

naumann@stce.rwth-aachen.de,

deussen@stce.rwth-

aachen.de, hueser@stce.rwth-aachen.de

## MS16

### New Approaches to Memory Reliability Management for Big Data Workloads

Abstract Not Available At Time Of Publication.

Dimitrios S. Nikolopoulos

Queen's University of Belfast

Belfast, NI

d.nikolopoulos@qub.ac.uk

## MS16

### Adaptive Precision Preconditioning for Sparse Linear Systems

The computational performance and energy consumption of current platforms, when executing practical scientific applications and big data analytic computations, is strongly limited by the cost of memory accesses. Therefore, innovative techniques to reduce communication, synchronization and energy usage are in urgent need in all areas of scientific computing in general, and in the sparse solution of linear systems in particular. Concretely, reducing the volume of data that is transferred between memory and the floating-point units, at each iteration of a preconditioned CG (PCG) method, is crucial to improve the computational performance and energy efficiency of this solver on current architectures. When a block-Jacobi preconditioner is combined with the CG method, most of the accesses to memory are due to the application of the preconditioner that occurs at each iteration of the solve. To tackle the cost of this stage, in this talk we will describe a parallel version of PCG that selectively employs different levels of precision to reduce communication during the application of a block-Jacobi preconditioner. We note that the benefits which can be expected from the application of an adaptive-precision preconditioner are orthogonal (and, therefore, additive) to those that can be obtained by techniques such as mixed precision+iterative refinement and specialized sparse data layouts such as CSB.

Enrique S. Quintana-Ortí

Universidad Jaime I

Spain

quintana@icc.uji.es

Hartwig Anzt

Innovative Computing Lab

University of Tennessee

hartwig.anzt@kit.edu

Jack J. Dongarra

University of Tennessee, Oak Ridge National Laboratory, USA

dongarra@icl.utk.edu

Goran Flegar

Universitat Jaume I



flegar@uji.es

#### MS17

##### Modeling Programming Languages and Potential Applications for HPC

Modeling programming languages, tools and frameworks is an essential part of complex software architectures. Modeling programming languages allows for advanced validated source code transformations that can replace manual implementations in several areas which are relevant for high-level abstractions as well as HPC related applications. An example for this type of transformations are domain specific optimizations. While modern compilers are very good at low-level optimizations, they lack the capability to perform optimizations that are based on domain knowledge and not directly reflected in the source language or the target architecture. In this presentation we will discuss how modeling frameworks can be used to highly augment manual implementations. We present our novel modeling framework VMF and discuss potential applications for native code and HPC. Creating models of tools and programming languages enables algorithms to process and transform programs and domain specific knowledge in the same way as algorithms process regular data.

Michael Hoffer  
G-CSC, Uni Frankfurt  
michael.hoffer@gcsc.uni-frankfurt.de

#### MS17

##### Robust and Adaptive Non-Linear Solvers

Abstract Not Available At Time Of Publication.

Arne Nägel  
Goethe-Center for Scientific Computing  
Goethe-University Frankfurt  
naegel@gcsc.uni-frankfurt.de

#### MS17

##### Meshing and Adaption of Unstructured Grid Hierarchies for Large Scale Parallel Multigrid Computations

The generation of detailed three dimensional meshes for the simulation of complex physical processes is crucial to capture important properties of the underlying domains and to reach a satisfying accuracy. At the same time this level of detail poses high demands both on suitable hardware and numerical solver efficiency. Parallel multigrid methods have been shown to exhibit near optimal weak scalability for massively parallel computations in various applications. Furthermore, using adaptive refinement and adaptive multigrid methods, the required CPUh to reach a solution can be substantially reduced. Especially in massively parallel environments, mesh generation, hierarchy creation, and grid distribution must go hand in hand to achieve maximal scalability. To this end we present a general, interweaved approach to parallel grid creation and hierarchical redistribution and show its applicability and scalability in several application areas.

Sebastian Reiter  
Goethe-Center for Scientific Computing  
Goethe-Universität Frankfurt

sebastian.reiter@gcsc.uni-frankfurt.de

#### MS17

##### Scaling Studies for Large Scale Multigrid Computations

The present-day supercomputers with hundred thousands of cores provide an enormous computing power and thus allow for large-scale simulations of complex physical models on highly resolved spatial geometries. Using grid-based finite-element and finite-volume discretization methods, we focus on questions ranging from biology to subsurface flow and solve the discretized equations employing the multigrid methods on massively parallel architectures. We present the employed parallelization strategy and its implementation together with empirical scaling results showing the efficiency and applicability of our multigrid implementation on massively parallel architectures.

Andreas Vogel  
University of Frankfurt, Germany  
a.vogel@rub.de

Sebastian Reiter  
Goethe-Center for Scientific Computing  
Goethe-Universität Frankfurt  
sebastian.reiter@gcsc.uni-frankfurt.de

Arne Nägel  
Goethe-Center for Scientific Computing  
Goethe-University Frankfurt  
naegel@gcsc.uni-frankfurt.de

Gabriel Wittum  
Extreme Computing Research Center  
King Abdullah University of Science and Technology  
gabriel.wittum@kaust.edu.sa

#### MS18

##### Pattern-based Modeling of Fail-stop and Soft-error Resilience for Iterative Linear Solvers

The reliability of high-performance computing (HPC) platforms is among the most critical challenges as systems continue to increase component counts, while the individual component reliability decreases and software complexity increases. While most resilience solutions are designed to address a specific fault model, HPC applications must contend with extremely high rates of faults from various sources with different levels of severity. Therefore, resilience for extreme-scale HPC systems and their applications requires an integrated approach, which leverages detection, containment and mitigation capabilities from different layers of the HPC environment. With this work, we propose an approach based on design patterns to explore a multi-level resilience solution that addresses silent data corruptions and process failures. The structured approach enables evaluation of the key components of a multi-level resilience solution using pattern performance models and systematically integrating the patterns into a complete solution by assessing the interplay between the patterns. We describe the design steps to develop a multi-level resilience solution for an iterative linear solver application that combines algorithmic resilience features of the solver with the fault tolerance primitives provided by ULFM MPI. Our results demonstrate the viability of designing HPC applications capable of surviving simultaneous injection of hard

and soft errors in a performance efficient manner.

Rizwan Ashraf, Christian Engelmann, Saurabh Hukerikar  
Oak Ridge National Laboratory  
ashrafra@ornl.gov, engelmannc@ornl.gov, huk-  
erikarsr@ornl.gov

## MS18

### Recent Advances on the Algorithm-based Fault Tolerance of the Sakurai-Sugiura Eigensolver

The Sakurai-Sugiura method (SSM) is a contour integral eigensolver for computing eigenvalues in a certain region and corresponding eigenvectors of generalized and nonlinear eigenvalue problems. Recent years, an inherent fault tolerance of SSM has been theoretically investigated. In this talk, we present techniques for efficiently utilizing the fault tolerance of the method.

Yasunori Futamura  
University of Tsukuba  
futamura@cs.tsukuba.ac.jp

Akira Imakura  
Department of Computer Science  
University of Tsukuba  
imakura@cs.tsukuba.ac.jp

Tetsuya Sakurai  
University of Tsukuba  
sakurai@cs.tsukuba.ac.jp

## MS18

### An Algorithm-Based Fault-Tolerant Framework for High-dimensional PDEs

We present our framework to solve high-dimensional PDEs based on the Fault-Tolerant Combination Technique (FTCT). The FTCT offers two main features: mitigating the curse of dimensionality by combining many coarse grids to approximate fine grids, and fault tolerance by excluding failed results. The FTCT does not rely on recomputing, which comes at a small cost in accuracy. Experiments show that the FTCT adds little overhead, small decrease of accuracy, and scales up to 180k cores.

Michael Obersteiner, Alfredo Parra Hinojosa  
Technical University of Munich  
oberstei@in.tum.de, hinojosa@in.tum.de

Mario Henne  
University of Stuttgart  
mario.heene@ipvs.uni-stuttgart.de

Hans-Joachim Bungartz  
Technical University of Munich, Department of Informatics  
Chair of Scientific Computing in Computer Science  
bungartz@in.tum.de

Dirk Pflüger  
University of Stuttgart  
Dirk.Pflueger@ipvs.uni-stuttgart.de

## MS18

### Algorithm Based Fault Recovery of Adaptively Re-

### finer Parallel Multilevel Grids

We will present a parallel adaptive multigrid method that uses dynamic data structures to store a nested sequence of meshes and the iteratively evolving solution. After a fault, the data residing on the faulty processor will be lost. However, with suitably designed data structures, the neighbouring processors contain enough information so that a consistent mesh can be reconstructed in the faulty domain with the goal of resuming the computation without having to restart from scratch.

Linda Stals  
Australian National University  
linda.stals@anu.edu.au

## MS19

### Horizontal Vectorization of Block Krylov Solvers

Many applications, like inverse problems, optimization or uncertainty quantification, require solving the same linear system for many different right-hand sides. Block Krylov methods allow solving these systems simultaneously. We discuss an efficient implementation of such Block Krylov methods using horizontal vectorization on modern HPC systems SIMD units. By solving for many right hand sides simultaneously, we increase the arithmetic intensity and by this the overall performance. Using generic C++ algorithms offers an attractive way to implement the block variants of the solvers. We employ the vectorized solvers to inverse modeling in EEG/MEG source analysis. Oversubscription of the vectorization units additionally improves the bandwidth usage, so that a nearly optimal speedup compared to the scalar version can be obtained.

Christian Engwer, Fahlke Jorrit, Andreas Nüßing  
INAM, University of Münster  
christian.engwer@uni-muenster.de, jorrit.fahlke@uni-  
muenster.de, andreas.nuessing@uni-muenster.de

## MS19

### Iterative Parallel Methods for Deriving ILU-Type Preconditioners

We consider linear systems of equations related to sparse matrices  $A$ . Incomplete LU decomposition usually leads to a good preconditioner. Here, we will present iterative and converging methods for computing ILU and MILU that are also easy to parallelize. The convergence speed of these iterative methods depends strongly on the condition number of  $A$ . So fast converging iterative ILU and MILU methods are necessary in this case. Therefore, we also consider Newton's method for computing ILU and MILU. For solving the resulting sparse triangular systems on a parallel computer we use the Jacobi method with an incomplete sparse approximate inverse preconditioner. Furthermore, the Jacobi iteration can be accelerated by using the Euler expansion. In view of the triangular structure of  $L$  and  $U$ , the accelerated Jacobi iteration is guaranteed to converge fast.

Thomas K. Huckle  
Institut fuer Informatik  
Technische Universitaet Muenchen  
huckle@in.tum.de

Hartwig Anzt  
Innovative Computing Lab  
University of Tennessee

hartwig.anzt@kit.edu

Matthias Bolten  
Bergische Universität Wuppertal  
Fakultät für Mathematik und Naturwissenschaften  
bolten@math.uni-wuppertal.de

#### MS19

##### **A Recursive Multilevel Trust Region Method for Phase-field Models of Fracture**

The numerical simulation of failure mechanism in solids is computationally challenging, as crack-paths with a possibly complex topology have to be resolved. Phase-field models for fracture overcome these difficulties by regularizing the sharp crack interfaces by means of a diffusive damage model. The resulting coupled systems of non-linear PDEs demand high-resolution meshes for their discretization, as the damage parameter has to be resolved sufficiently accurate. This leads to large-scale nonlinear systems, which have to be solved in each time step. In order to obtain an overall optimal solution method, we employ a recursive multilevel trust region algorithm. This method tackles the nonlinearity directly on each level. This is achieved by introducing level dependent nonlinear objective functions, whose minimization can yield good coarse level corrections for the fine level problem. In order to build efficient coarse levels models for coupled phase-field models, we make use of operator splitting techniques. The presented numerical examples will show that multilevel-based optimization techniques lead to highly efficient solution methods. We will demonstrate this by analyzing convergence results and scaling behaviour.

Alena Kopanicakova  
Università della Svizzera italiana  
kopana@usi.ch

Carola Bilgen, Kerstin Weinberg  
Uni Siegen  
carola.bilgen@uni-siegen.de,  
kerstin.weinberg@uni-siegen.de

Rolf Krause  
Università della Svizzera italiana, Switzerland  
rolf.krause@usi.ch

#### MS19

##### **Parallel Iterative Methods in the ELPA Eigensolver**

The solution of symmetric eigenproblems plays a key role in many computational simulations. Generalized eigenproblems are transformed to a standard problem. This transformation has the drawback that for banded matrices in the generalized eigenproblem the banded structure is not preserved. The matrix of the standard eigenproblem will generally be a full matrix. We followed the ideas of the Group of Lang (University of Wuppertal) who modified Crawford's algorithm and implemented an iterative procedure to the ELPA project which keeps the bandwidth. By keeping the banded structure we save one reduction step on the matrix and one backtransformation step for the eigenvectors. This provides a good speedup compared to the standard transformation procedure with Cholesky factorization.

Michael Rippl  
Technische Universität München

Germany  
ripplm@in.tum.de

#### MS20

##### **Role of Tensors in Machine Learning**

Tensors are higher order extensions of matrices that can incorporate multiple modalities and encode higher order relationships in data. Tensors play a significant role in machine learning through (1) tensor contractions, (2) tensor sketches, and (3) tensor decompositions. I will show that tensor contractions are an effective replacement for fully connected layers in deep learning architectures and result in significant space savings. Tensor contractions present rich opportunities for hardware optimizations through extended BLAS kernels. We propose a new primitive known as StridedBatchedGEMM in CuBLAS 8.0 that significantly speeds up tensor contractions, and avoids explicit copy and transpositions. Tensor sketches are extensions of the popular count sketches for vectors and provide succinct representations in multi-modal tasks such as visual question and answering. Lastly, I will present analysis on how tensor decompositions can efficiently learn latent variable models with guarantees. These functionalities will be demonstrated on Jupyter notebooks on AWS using Tensorly package and has the Apache Mxnet backend for efficient multi-GPU scaling. It uses the new gluon interface with both declarative and imperative programming, which provides both ease of programming, as well as high efficiency.

Anima Anandkumar  
EECS Department  
University of California, Irvine  
animakumar@gmail.com

#### MS20

##### **Shared-Memory Parallelization of MTTKRP for Dense Tensors**

The matricized-tensor times Khatri-Rao product (MTTKRP) is the computational bottleneck for algorithms computing CP decompositions of tensors. In this paper, we develop shared-memory parallel algorithms for MTTKRP involving dense tensors. The algorithms cast nearly all of the computation as matrix operations in order to use optimized BLAS subroutines, and they avoid reordering tensor entries in memory. We benchmark sequential and parallel performance of our implementations, demonstrating high sequential performance and efficient parallel scaling. We use our parallel implementation to compute a CP decomposition of a neuroimaging data set and achieve a speedup of up to 7.4 times over existing parallel software.

Koby B. Hayashi, Grey Ballard, Tobia Michael, Yuijie Jiang  
Wake Forest University  
hayakb13@wfu.edu, ballard@wfu.edu, tobiamj@wfu.edu, jiany14@wfu.edu

#### MS20

##### **Accelerating the Tucker Decomposition with Compressed Sparse Tensors**

The Tucker decomposition is a higher-order analogue of the singular value decomposition and is a popular method of performing analysis on multi-way data (*tensors*). Computing the Tucker decomposition of a sparse tensor is demanding in terms of both memory and computational re-

sources. The primary kernel of the factorization is a chain of tensor-matrix multiplications (TTMc). State-of-the-art algorithms accelerate the underlying computations by trading off memory to memoize the intermediate results of TTMc in order to reuse them across iterations. We present an algorithm based on a compressed data structure for sparse tensors and show that many computational redundancies during TTMc can be identified and pruned without the memory overheads of memoization. In addition, our algorithm can further reduce the number of operations by exploiting an additional amount of user-specified memory. We evaluate our algorithm on a collection of real-world and synthetic datasets and demonstrate up to  $20.7\times$  speedup while using  $28.5\times$  less memory than the state-of-the-art parallel algorithm.

Shaden Smith, George Karypis  
University of Minnesota  
shaden@cs.umn.edu, karypis@cs.umn.edu

## MS20

### The SPLATT Toolkit on the K Computer

Tensors are used to represent data with high dimensionality in many applications, such as web searching, item and tag recommendations, signal and image processing. In these applications, tensor decomposition algorithms are generally applied to find latent relations or predict missing elements in data through operations on a low rank form of tensors. Currently, the CANDECOMP/PARAFAC decomposition (CPD) is one of the prominent tensor decomposition formulations, and is commonly used to formulate a tensor as the summation of rank-one tensors. In this research, the SPLATT, which is a software toolkit developed by the University of Minnesota, and providing a novel data structure and memory and operation-efficient algorithms to carry out the CPD of sparse tensors, is investigated and evaluated on the K computer, and its potentials to be transplanted on the K computer is explored.

Yiyu Tan  
RIKEN Advanced Institute for Computational Science  
tan.yiyu@riken.jp

Toshiyuki Imamura  
RIKEN Advance Institute for Computational Science  
imamura.toshiyuki@riken.jp

## MS21

### COMPSs-Mobile: Parallel Programming for Mobile Cloud Computing

Mobile Cloud Computing (MCC) brings together the interacting and interconnecting capabilities of mobile devices with the "infinite" computing power of the Cloud. COMPSs is a programming model that aims to ease the development of parallel applications running on distributed infrastructures such as clusters or clouds; however, the current runtime system, responsible for the automatic parallelization of the application, does not fit well on MCC environments. Due to mobility, mobile applications are restrained by the limited battery lifetime and network issues are likely to appear. Here, we present the redesigned COMPSs runtime toolkit to enhance the distributed execution of applications on mobile devices.

Francisco Javier Conejero  
Barcelona Supercomputing Center

francisco.conejero@bsc.es

## MS21

### Robust Middleware Services for In-Situ Workflows

Dramatically increasing data volumes and data management/IO costs have led to in-situ approaches data processing and analysis, and resulting in-situ workflows. As a result, robust and scalable middleware services for in-situ workflows that target emerging architectures with deep and complex memory hierarchies and can address extreme scales challenges such as reliability, are becoming essential. In this talk, we explore staging based robust middleware services for in-situ workflows, based on the DataSpace project.

Pradeep Subedi  
Rutgers University  
pradeep.subedi@rutgers.edu

## MS21

### Cache Partitioning as an Extension to Co-Scheduling: Can Applications Benefit from It?

Novel many-core architectures are able to dynamically partition a processor's last level cache memory between cores. This cache partitioning can be efficiently utilised for high performance computing by co-scheduling several (usually two) HPC applications such that they do not disturb each other when accessing the last level cache. This talk will provide first results on using cache partitioning through Intel®'s Cache Allocation Technology (CAT) and how these can be used for co-scheduling strategies in HPC.

Carsten Trinitis  
TU Munich  
carsten.trinitis@in.tum.de

## MS21

### Approaches to the Power Consumption Problem on the K Computer

Recently, HPC systems have become larger and the power consumption has become one of the most important issues. The K computer has also been confronted with this problem. Some applications have consumed the electric power more than our assumption based on LINPACK, and we faced on how to reduce the power consumption in the operation of K computer. In this talk, I'll introduce our approaches to this problem.

Atsuya Uno  
RIKEN Advanced Institute for Computational Science  
uno@riken.jp

## MS22

### Parallelization of Variable Preconditioned Krylov Subspace Method for Linear System Obtained from Meshless Approaches

Recently, a clock frequency of CPU has gone as far as it can go, accelerators such as Graphics Processing Unit (GPU) are adopted for high performance computing calculations. In GPU programming, the simulation code should be parallelized by CUDA or API such as Message Passing Interface (MPI). However, GPU programming cost using CUDA becomes very high. Thus, the algorithm of the method should be simple. GPU for general purpose calculation becoming

cheaper and easy to construct high-performance computer cluster. Thus, the parallelization technique should actively be adopted for the numerical simulation to fulfill the huge scale simulation, and the high performance numerical investigations of electromagnetic field using GPU increase recently. The linear system obtained by eXtended Element-Free Galerkin (XEFG) has a same difficulty as a saddle point problem. Thus, the system becomes ill-posed problem as increasing a dimension size of the system. Furthermore, the system is included zero diagonal elements, and stationary iterative methods such as Successive Over Relaxation (SOR) method is difficult to adopt for the solver. The purpose of the present study is to implement the mixed precision Variable Preconditioned (VP) Krylov subspace method on GPU, and the performances of a GPU device are investigated. In addition, the linear system obtained by XEFG is employed for the evaluations.

Soichiro Ikuno

Tokyo University of Technology  
ikuno@stf.teu.ac.jp

Taku Itoh

Nihon University  
itoh.taku@nihon-u.ac.jp

## MS22

### Performance Evaluation of Tiled 3D FDTD Solver on Recent Multicore Processors

The 3D FDTD method is one of standard simulation methods for electromagnetic waves. We have developed a multi-threaded tiled 3D FDTD solver on a recent multi-core processor. The solver involves a time-space tiling technique in which tile-level parallelization can be exploited. On a problem of  $200^3$  grid points, the developed solver attained 1.88 fold speedup compared with the naively implemented solver. Moreover, numerical tests show that the tile-level parallelization is more effective than the intra-tile parallelization.

Takeshi Iwashita, Takeshi Fukaya

Hokkaido University  
iwashita@iic.hokudai.ac.jp, fukaya@iic.hokudai.ac.jp

## MS22

### Geometric Block Diagonal Preconditioning Technique for Electromagnetic Finite Element Analysis Using Voxel Mesh

Recently, we proposed geometric block diagonal preconditioning technique for large matrix equations arising in electromagnetic finite element computation using a voxel mesh. In this presentation, the concept and theoretical aspect of the new preconditioning technique are discussed. Our technique can be applied to both A- and A-phi formulations and the effectiveness of the technique is demonstrated in some test applications including magnetostatic, eddy-current, and full wave problems.

Takeshi Mifune

Department of Electrical Engineering  
Kyoto University  
mifune@fem.kuee.kyoto-u.ac.jp

## MS22

### An Ultrafast Multi-GPU Eigensolver for Band Structures of Three-Dimensional Photonic Crystals

## tals

Photonic crystals are made up of space-dependent dielectric materials with periodic structures. Some particular photonic crystals lead to full band gaps. Various novel physical studies and promising applications innovations focus on these full band gap structures. Such studies and innovations rely on the simulations of three-dimensional photonic crystals. However, it is very time-consuming to solve the corresponding Maxwell equations numerically. To overcome these numerical challenges, we have developed efficient eigenvalue solvers based on the null-space free techniques and implemented the proposed algorithms on multiple-GPU systems. In addition to the descriptions of the algorithms, we will demonstrate the numerical results to illustrate the efficiency of the proposed algorithms and the parallel eigensolver. This is a joint work with Tsung-Ming Huang at National Taiwan Normal University and Wen-Wei Lin at National Chiao Tung University.

Weichung Wang

National Taiwan University  
Institute of Applied Mathematical Sciences  
wwang@ntu.edu.tw

## MS23

### The Flexible Computational Science Infrastructure (FLESCSI) Programming System: Overview & Productivity

Abstract Not Available At Time Of Publication.

Ben Bergen

Los Alamos National Laboratory  
bergen@lanl.gov

## MS23

### Way of the DARMA: Sequential Semantics for Deterministic-by-default Task Parallelism

Over the past half century, imperative programming models and languages have dominated high-performance scientific computing. A key reason for this dominance is that imperative semantics are a nice match for both the typical means of expressing scientific algorithms and for the underlying execution of those algorithms on simple, sequential processors. While the former of these remains an important factor in programming model design for many application domains, the latter is based on an increasingly flawed picture of modern hardware. The challenge for next-generation programming models is thus to facilitate the extraction of concurrency and asynchrony needed for new architectures without sacrificing key features that make imperative programming useful and efficient for scientific computing applications. DARMA is a programming models research effort to develop a set of abstractions and semantics that naturally extract a useful concurrency specification from applications written with familiar and productive semantics. DARMA abstractions are data-centric and deterministic-by-default, meaning that the program result is independent execution order within the constraints imposed by the model. The use of zero-overhead (or low-overhead) implementations of these abstractions has already enabled performance and scalability comparable to more traditional models, even in the early stages of the project.

David S. Hollman

Sandia National Labs



dshollm@sandia.gov

## MS23

### Charm++: Interoperation and Support for Higher Level Abstractions

Abstract Not Available At Time Of Publication.

Phil Miller

Charmworks

University of Illinois Urbana-Champaign

phil@hpccharm.com

## MS23

### Scaling Implicitly Parallel Programs to a Thousand Nodes with Regent

State of the art high-performance scientific applications are typically written in low-level, explicitly parallel programming models that require users to manually manage parallelism and data movement in their applications. Implicit parallelism is an attractive alternative in which the programming system takes responsibility for scheduling parallelism and data movement across the machine. However, traditional implementations of implicit parallelism suffer from substantial limitations: static, compiler-based implementations restrict the programming model to exclude dynamic features, while dynamic, runtime-based approaches incur overheads that can limit scalability. Regent is a programming language designed to enable a hybrid static and dynamic analysis of implicit parallelism. Regent programs are composed of tasks, and program data is stored in regions. A Regent implementation is responsible for automatically discovering parallelism in an application by analyzing the region usage of executed tasks in program order. We present an optimizing compiler for Regent that transforms implicitly parallel programs into efficient explicitly parallel code via a combination of static and dynamic program analysis. We measure the performance and scalability of several Regent programs on large supercomputers and demonstrate that optimized Regent programs perform comparably to manually optimized explicitly parallel programs.

Elliott Slaughter

SLAC National Accelerator Laboratory

slaughter@cs.stanford.edu

## MS24

### A Mixed Precision Semi-Lagrangian Method for Modern Hardware Architectures

Obtaining approximate solutions to the Vlasov equation is important for many applications in plasma and astrophysics. The main difficulties in conducting such simulations is the high-dimensional phase space and the nonlinear nature of the problem. In addition, it is desirable to conserve a range of physically important invariants. Most algorithms that have been employed previously introduce global communication and are thus not very well suited for modern HPC systems. In this talk we consider a Hamiltonian splitting in conjunction with a semi-Lagrangian discontinuous Galerkin (sLdG) method. The sLdG method is not only local and conserves charge and current exactly, it also shows excellent long time behavior with respect to energy conservation. We propose a mixed-precision implementation that exploits the structure of the approximation space and succeeds in reducing memory as well as in

increasing performance. Numerical results on graphic processing units and on the Intel Xeon Phi as well as scaling studies on clusters will be discussed. We further show how some of these ideas can be extended to different partial differential equations (such as the Kadomtsev-Petviashvili equation). This talk is based on [L. Einkemmer, A. Ostermann, arXiv:1701.05602], [L. Einkemmer, J. Plasma Phys. 83:2 (2017)], [L. Einkemmer, High Performance Computing and Simulation (HPCS), 2016 International Conference on], and [N. Crouseilles, L. Einkemmer, E. Faou, J. Comput. Phys. 283 (2015)].

Lukas Einkemmer

University of Innsbruck

lukas.einkemmer@uibk.ac.at

## MS24

### Toward Modern Scientific Computations Based on Structure-Preserving Methods

In last several decades scientific computations have experienced rapid developments both in theoretical numerical analysis and high-performance computing. In the former, the concept of structure-preserving methods have proved its power for better numerics, while in the latter various techniques for parallelization, communication, and implementation have been devised for faster computations. Despite each success in its own research field, they both have come to a certain maturity, and it seems some new breakthrough is hoped. In this talk, we discuss some new possibilities in combining the two research fields, which we feel a promising new research direction.

Takayasu Matsuo

Graduate School of Information Science and Technology,  
the University of Tokyo

matsuo@mist.i.u-tokyo.ac.jp

Yuto Miyatake

Nagoya University

miyatake@na.nuap.nagoya-u.ac.jp

## MS24

### Structure-Preserving Model Reduction of Forced Hamiltonian Systems

This work proposes the proper symplectic decomposition (PSD) method for forced Hamiltonian systems. As an analogy to the proper orthogonal decomposition (POD), PSD is designed to build a symplectic subspace to fit empirical data. Our aim is two-fold. First, to achieve computational savings for large-scale Hamiltonian systems with external forces. Second, to simultaneously preserve the symplectic structure and the forced structure of the original system. Corresponding to the local form of d'Alembert's principle, we propose the symplectic Galerkin projection method to reconstruct low-dimensional systems. In a special case when the external force is described by the Rayleigh dissipative function, the proposed method automatically preserves the dissipativity, boundedness, and stability of the original system. The stability, accuracy, and efficiency of the proposed method are illustrated through numerical simulations of wave equations.

Liqian Peng

Extreme Scale Data Science & Analytics Department  
Sandia National Laboratories

lpeng@sandia.gov

## MS24

### Energy-Preserving Parareal Algorithm for the Hamilton Equation

Energy-preserving integrators are integrators that preserve the energy. These integrators are indispensable for long time simulations performed in, for example, celestial mechanics. Because in such cases the degree of parallelism in the spatial direction is possibly small, parallelization in the temporal direction is of importance. In this talk, we propose a parareal algorithm where the both fine and coarse solvers are energy-preserving by employing a recently proposed explicit energy-preserving method as the coarse solver.

Ai Ishikawa, Takaharu Yaguchi, Mitsuo Yokokawa  
Kobe University  
a-ishikawa@stu.kobe-u.ac.jp, yaguchi@pearl.kobe-u.ac.jp,  
yokokawa@port.kobe-u.ac.jp

## MS25

### The ELSI Infrastructure for Large-Scale, Massively Parallel Electronic Structure Theory

Electronic structure theory, particularly Kohn-Sham density functional theory (KS-DFT), is a "sweet spot" for realistic atomistic simulations of chemical systems and materials. For "small" systems (up to approx. 1,000 atoms, depending on the approach), a traditional linear eigenvalue solution (cubic scaling with system size), embedded in a non-linear fixed-point iteration scheme, is often computationally most efficient to solve for the Kohn-Sham eigenvectors, eigenvalues, density matrix, and ultimately the total energy and other predicted observables. Larger systems, however, require lower-scaling solutions, which are often restricted in applicability and not equally efficient in all scenarios. This lecture introduces the concepts and present status of the "ELSI" ELectronic Structure Infrastructure [Yu et al., Comp. Phys. Commun. (2017), DOI: 10.1016/j.cpc.2017.09.007], a software layer designed to unify access to different solver strategies and their open-source implementations. ELSI implements efficient access to three different solvers, ELPA (dense linear algebra based eigenvalue solution), libOMM (cubic scaling general solution for the density matrix), and PEXSI (quadratic scaling or below for KS-DFT), as well as efficient format conversion strategies for dense and sparse distributed matrix storage formats, and more. A performance comparison of all three solvers towards large systems is presented, and future strategies to seamlessly switch between them are discussed.

Volker Blum  
Duke University  
Department of Mechanical Engineering and Materials Science  
volker.blum@duke.edu

## MS25

### A Conjugate Gradient Method for Electronic Structure Calculations

In this talk, we will introduce a conjugate gradient method for electronic structure calculations. We propose a Hessian based step size strategy, which together with three orthogonality approaches yields three algorithms for computing the ground state energy of atomic and molecular systems. Un-

der some mild assumptions, we prove that our algorithms converge locally. It is shown by our numerical experiments that the conjugate gradient method is efficient. This is a joint work with Zhuang Liu, Liwei Zhang, and Aihui Zhou.

Xiaoying Dai  
Institute of Computational Mathematics  
Chinese Academy of Sciences, Beijing, China  
daixy@lsec.cc.ac.cn

## MS25

### Eigenstate Calculations using Sakurai-Sugiura Method with the Large-scale Electronic Structure Calculation Code CONQUEST

Abstract Not Available At Time Of Publication.

Ayako Nakata  
National Institute for Materials Science  
nakata.ayako@nims.go.jp

Tsuyashi Miyazaki  
National Institute for Materials Science  
miyazaki.tsuyoshi@nims.go.jp

Yasunori Futamura  
University of Tsukuba  
futamura@cs.tsukuba.ac.jp

Tetsuya Sakurai  
Department of Computer Science  
University of Tsukuba  
sakurai@cs.tsukuba.ac.jp

## MS25

### Parallel Computation of the Single Particle Density Matrix for Electronics Structure Calculations

Computing the single particle density matrix (SPDM) for electronics structure calculations can be quite expensive, typically requiring  $O(N^3)$  operations. We present our efforts to efficiently compute the SPDM within the context of a linear scaling algorithm for electronics structure calculations, using an intermediate basis set of localized orbitals and by a series of "local" solves.

Daniel Osei-Kuffuor  
Lawrence Livermore National Laboratory  
oseikuffuor1@llnl.gov

## MS26

### Exascale Deep Learning for Text Comprehension

We present an overview of how deep learning approaches are being used for understanding clinical pathology reports. In particular, we describe newer algorithmic approaches to simultaneously extract multiple tumor locations, histological grade and biomarker information automatically from clinical reports. We demonstrate how simple improvements to the algorithm can allow for scaling on emerging computing platforms. Additionally, we describe some of our newer approaches that we have been developing for optimizing hyperparameters in the context of designing better performing deep networks.

Arvind Ramanathan, Shang Gao, Michael Young, John Qiu, James Blair Christian, Hong-Jun Yoon, Georgia

Tourassi  
Oak Ridge National Laboratory  
ramanathana@ornl.gov, gaos@ornl.gov,  
youngmt1@ornl.gov, qiuix@ornl.gov, christianjb@ornl.gov,  
yoonhj@ornl.gov, tourassig@ornl.gov

Paul Fearn  
National Cancer Institute  
paul.fearn@nih.gov

## MS26

### Challenges in Training Large Neural Networks for Natural Language Processing

Deep neural networks in computer vision has achieved superior performance, especially when more layers are used. Recently, with the help of residual connection, neural networks in Natural Language Processing (NLP) are also becoming deep. However, we are facing multiple challenges to train a large NLP model. In particular, the challenges mainly come from two sources: the increased amount of computation and number of parameters. Although one can reduce the impact of high computational cost by training the models on a multi-GPU system, the increased number of parameters makes it difficult to synchronize the gradients. In this talk, we will introduce some techniques that enables fast computation and reduces the model size. A discussion will be given for boosting the speed of multi-GPU training for large NLP models.

Raphael Shu  
University of Tokyo  
shu@nlab.ci.i.u-tokyo.ac.jp

Hideki Nakayama  
The University of Tokyo  
nakayama@ci.i.u-tokyo.ac.jp

## MS26

### Accelerated Deep Learning Advances in Fusion Energy Science Predictions

Deep learning methods are increasingly deployed in many scientific and industrial domains. For example, time-urgent and very challenging problem facing the development of a fusion energy reactor is the need to reliably predict and avoid large-scale major disruptions in magnetically-confined tokamak systems such as the EUROfusion Joint European Torus (JET) today and the burning plasma ITER device in the near future. Significantly improved methods of prediction are required to provide sufficient advanced warning for disruption avoidance or mitigation strategies to be effectively applied before critical damage can be done to ITER – a ground-breaking \$25B international burning plasma experiment with the potential capability to exceed breakeven fusion power by a factor of 10 or more. This formidable task demands accuracy beyond the reach of hypothesis-driven /first-principles extreme-scale computing (HPC) simulations that dominate current research and development in the field. Recent HPC-relevant advances in the deployment of deep learning recurrent nets have been demonstrated in scaling studies of Princeton's new Deep Learning Code – "FRNN (Fusion Recurrent Neural Net) Code on modern GPU systems.

William Tang  
Princeton University  
PPPL  
tang@pppl.gov

Julian Kates-Harbeck  
Harvard University  
juliankh91@gmail.com

Alexey Svyatkovskiy  
Princeton University  
alexey@princeton.edu

## MS26

### Scalable Deep Learning for Scientific Data Sets

In this presentation we will talk about how high performance computing is enabling the scientists to train larger, deep neural network models more quickly and how that can accelerate the the role of scientific discovery. Additionally, we will present our approach for developing a distributed memory scalable deep learning toolkit optimized for HPC systems.

Brian Van Essen  
Lawrence Livermore National Laboratory  
vanessen1@llnl.gov

## MS27

### A Framework for Large-Scale, High-Performance Lattice Boltzmann Simulations in Complex Geometries

The waLBerla software framework provides efficient parallel methods that enable the simulation of flows on adaptive meshes in complex domains. The lattice Boltzmann method has proven to be a suitable approach to scalable computational fluid dynamics. Our implementation is carefully designed for large-scale applications and has enabled simulations of up to  $10^{12}$  lattice cells. We will discuss data structures, algorithms, and novel software engineering techniques based on automatic program transformations that additionally help to achieve excellent single-core and intra-node performance. We also present experiments on strong and weak scalability at large scale.

Christian Godenschwager  
University of Erlangen-Nuremberg  
christian.godenschwager@fau.de

Martin Bauer  
University of Erlangen-Nuremberg  
Department of Computer Science  
martin.bauer@fau.de

Florian Schornbaum  
University of Erlangen-Nuremberg  
florian.schornbaum@informatik.uni-erlangen.de

Ulrich Rüde  
Friedrich-Alexander-University Erlangen-Nuremberg,  
Germany  
ulrich.ruede@fau.de

## MS27

### An Octree-Based AMR Lattice Boltzmann Method for Multi-GPU Aerodynamics Simulations

Lattice Boltzmann Method (LBM) is one of promising methods for large-scale aerodynamics simulations especially for GPU computing. The area near complex bodies should be locally refined to assign high resolution mesh for turbulent flow computations in the way of Large Eddy

Simulation (LES). An octree-based Adaptive Mesh Refinement (AMR) has been introduced and we optimize the implementation by using the C++ template for automatic GPU-kernel generation. In order to improve the numerical stability and the accuracy, the Multiple-relaxation-time model or cumulant model of LBM, Interpolated Bounce-back boundary condition have been employed. Furthermore we have implemented a multi-GPU computation and perform turbulent flow simulations with Reynolds numbers of more than 100,000 for complex shape bodies.

Yuta Hasegawa, Takayuki Aoki  
Tokyo Institute of Technology  
hasegawa@sim.gsic.titech.ac.jp, taoki@gsic.titech.ac.jp

## MS27

### Extending Existing Simulation Software Packages with Spatial Adaptivity in a Minimal-Invasive Way

We present our approach for minimally invasive integration of dynamically-adaptive tree-structured grids in established large simulation codes for the example of the molecular dynamics (MD) simulation code ESPResSo. It includes an implementation of the lattice-Boltzmann method (LBM) generating a background flow for a molecular ensemble. We port ESPResSo's originally regular grids for short-range MD and the LBM to an extended version of p4est, a well-known and scalable library for tree-structured Cartesian grids. Our contribution to p4est allows simpler integration into existing applications. As short-range MD and LBM do not use the same grid resolution, we use independent p4est instances for each algorithm and describe our approach for reducing communication in the coupling scheme. Whereas the LBM grid changes dynamically, we stick to a regularly discretized grid for the short-range MD. We show results and scalability tests for both components separately as well as for the integrated application on different hardware architectures. Due to the high parallelism of the LBM, it is particular suited for using acceleration devices. To this end, we extend ESPResSo's single GPU implementation of the LBM on a regular grid to an adaptive implementation on multiple GPUs. For the GPU implementation, we show upscaling tests on a GPU-rich supercomputer.

Michael Lahnert  
University of Stuttgart  
michael.lahnert@ipvs.uni-stuttgart.de

Takayuki Aoki  
Tokyo Institute of Technology  
taoki@gsic.titech.ac.jp

Malte Brunn  
University of Stuttgart  
malte.brunn@ipvs.uni-stuttgart.de

Carsten Burstedde  
Universität Bonn  
burstedde@ins.uni-bonn.de

Steffen Hirschmann, Dirk Pflüger  
University of Stuttgart  
steffen.hirschmann@ipvs.uni-stuttgart.de,  
Dirk.Pflueger@ipvs.uni-stuttgart.de

Miriam Mehl  
Universität Stuttgart

miriam.mehl@ipvs.uni-stuttgart.de

## MS27

### Acceleration of Turbulent Wind Simulation Using Locally Mesh-Refined Lattice Boltzmann Method

A real-time simulation of the environmental dynamics of radioactive substances is very important from the viewpoint of nuclear security. Since a lot of tall buildings and complex structures make the air flow turbulent in urban cities, large-scale CFD simulations are needed. To this end, a CFD code based on a Lattice Boltzmann Method (LBM) with a block-based Adaptive Mesh Refinement (AMR) method is developed. As the conventional LBM based on a single relaxation time collision operator often becomes numerically unstable at high Reynolds number, we apply a state-of-the-art cumulant collision operator. The code is developed on TSUBAME supercomputer, and high scalability is achieved. By using new functions in CUDA8.0, the GPU kernel functions are tuned to achieve high performance on the latest Pascal GPU architecture. By introducing a temporal blocking technique, the number of the MPI communications is significantly reduced. The code is validated against a wind tunnel test, which was released from the National Institute of Advanced Industrial Science and Technology (AIST). The computational grids are subdivided by the AMR method, and the total number of grid points is reduced to less than 10% compared to the finest meshes. In spite of the fewer grid points, the turbulent statistics are in good agreement with the experiment data. In the presentation, we show a large-scale wind simulation, which takes account of real buildings, and demonstrate the effectiveness of our method.

Naoyuki Onodera  
Center for Computational Science & e-Systems  
Japan Atomic Energy Agency  
onodera.naoyuki@jaea.go.jp

Yasuhiro Idomura  
Japan Atomic Energy Agency  
idomura.yasuhiro@jaea.go.jp

## MS28

### Enlarged GMRES for Reducing Communication

We propose a variant of GMRES method for the solution of linear systems of equations with one right-hand side and multiple right-hand sides. Our method is based on the idea of the enlarged Krylov subspace introduced by L. Grigori et al. to reduce communication. By the nature of this idea, our method inherits a block version of GMRES, so we are interested in the detection of inexact breakdowns focusing on the strategy proposed by H. Calandra et al. with a modification on the test of detection. We also propose an eigenvalues deflation technique aiming to have two benefits, on one hand, to avoid the plateau of convergence after the end of a cycle in the restarted version, and on the other hand to have a very fast convergence when solving the same system with different right-hand sides each one given in a time (useful for CPR preconditioner).

Hussam Al Daas  
INRIA  
hussam.al-daas@inria.fr

Laura Grigori  
INRIA  
France

Laura.Grigori@inria.fr

Pascal Henon  
TOTAL E&P  
pascal.henon@total.com

Philippe Ricoux  
TOTAL SA  
philippe.ricoux@total.com

## MS28

### High Performance Variants of Krylov Subspace Methods

Krylov subspace methods are general and widely-used iterative methods for solving large, sparse linear systems. Computationally, the vector updates in each iteration of a Krylov subspace method require one or more matrix-vector products and inner products, which both require expensive interprocessor communication. The computation of inner products in particular requires a global synchronization, which can be especially costly on large-scale machines. In this introductory talk, we present synchronization-reducing modifications of Krylov subspace methods, focusing on the popular pipelined and s-step variants. We discuss performance characteristics of these methods, as well as their stability and convergence properties in finite precision arithmetic. We emphasize that such synchronization-reducing variants exhibit complicated tradeoffs between the speed of each iteration, the convergence rate, and the attainable accuracy; the optimal point in this tradeoff space is highly application-dependent. The optimization and implementation of high-performance Krylov subspace methods thus requires a holistic approach with respect to a particular application and machine architecture.

Emmanuel Agullo  
INRIA  
emmanuel.agullo@inria.fr

Erin C. Carson  
New York University  
erin.carson@nyu.edu

Siegfried Cools  
University of Antwerp  
siegfried.cools@uantwerp.be

James Demmel  
UC Berkeley  
demmel@eecs.berkeley.edu

Pieter Ghysels  
Lawrence Berkeley National Laboratory  
pghysels@lbl.gov

Luc Giraud  
Inria  
luc.giraud@inria.fr

Miro Rozloznik  
Czech Academy of Sciences  
Prague, Czech Republic  
miro@cs.cas.cz

Zdenek Strakos  
Faculty of Mathematics and Physics,  
Charles University in Prague

strakos@karlin.mff.cuni.cz

Petr Tichy  
Czech Academy of Sciences  
tichy@cs.cas.cz

Miroslav Tuma  
Academy of Sciences of the Czech Republic  
Institute of Computer Science  
tuma@cs.cas.cz

Wim Vanroose  
Antwerp University  
wim.vanroose@uantwerpen.be

Emrullah Fatih Yetkin  
Inria  
emrullah-fatih.yetkin@inria.fr

## MS28

### About Parallel Variants of GMRES Algorithm

Sparse linear systems  $Ax = b$  arise very often in computational science and engineering. Krylov methods are very efficient iterative methods, and restarted GMRES is a reference algorithm for non-symmetric systems. A first issue is to ensure a fast convergence, by preconditioning the system with a matrix  $M$ . Preconditioning must reduce the number of iterations, and be easy to solve. A second issue is to achieve high performance computing. The most time-consuming part in GMRES is to build an orthonormal basis  $V$ . With the Arnoldi process, many scalar products involve global communications. In order to avoid them, s-step methods have been designed to find a tradeoff between parallel performance and stability. Also, solving a system with the matrix  $M$  and for multiplying a vector by the matrix  $A$  should be efficient. A domain decomposition approach involves mainly local communications and is frequently used. A coarse grid correction, based on deflation for example, improves convergence. These techniques can be combined to provide fast convergence and fast parallel algorithms. Numerical results illustrate various issues and achievements.

Jocelyne Erhel  
INRIA-Rennes, France  
Jocelyne.Erhel@inria.fr

## MS28

### Iteration-Fusing Conjugate Gradient

This paper presents the Iteration-Fusing Conjugate Gradient (IFCG) approach which is an evolution of the Conjugate Gradient method that consists in i) letting computations from different iterations to overlap between them and ii) splitting linear algebra kernels into subkernels to increase concurrency and relax data-dependencies. The paper presents two ways of applying the IFCG approach: The IFCG1 algorithm, which aims at hiding the cost of parallel reductions, and the IFCG2 algorithm, which aims at reducing idle time by starting computations as soon as possible. Both IFCG1 and IFCG2 algorithms are two complementary approaches aiming at increasing parallel performance. Extensive numerical experiments are conducted to compare the IFCG1 and IFCG2 numerical stability and performance against four state-of-the-art techniques. By considering a set of representative input matrices, the paper demonstrates that IFCG1 and IFCG2 provide parallel



performance improvements up to 42.9% and 41.5% respectively and average improvements of 11.8% and 7.1% with respect to the best state-of-the-art techniques while keeping similar numerical stability properties. Also, this paper provides an evaluation of the IFCG algorithms' sensitivity to system noise and it demonstrates that they run 18.0% faster on average than the best state-of-the-art technique under realistic degrees of system noise.

Sicong Zhuang, Marc Casas  
Barcelona Supercomputing Center  
sicong.zhuang@bsc.es, marc.casas@bsc.es

## MS29

### If It Doesn't Work, We Learn Something: Instructive Case Studies in Performance Engineering

Performance Engineering is a process that embraces false assumptions and dead ends in working out the solution to a problem. Thus it will never be completely automated: Correcting something that did not work as intended leads to better insight. We present examples from performance modeling and code optimization, showcasing how important it is to get it wrong the first time, and how dangerous it can be to leave some of the work entirely to tools.

Georg Hager  
Erlangen Regional Computing Center (RRZE)  
georg.hager@fau.de

## MS29

### Efficient and Robust Parallel ILU Preconditioners for Quantum Eigenvalue Problems in ppOpen-HPC/ESSEX-II

Targets of the project ppOpen-HPC/ESSEX-II are large-scale generalized eigenvalue problems derived from quantum physics applications. We chose preconditioned Krylov iterative methods for solving the problems on exascale systems. Generally, coefficient matrices are very ill-conditioned, and special preconditioning methods are needed to address them. We focus on ILU preconditioner parallelized with multi-coloring. We proposed regularization methods for the robustness and a hierarchical parallelization of multi-coloring algorithms for the massive parallelism. In this talk, we present overview of the proposed methods, and effective results which are including parallel performances with 4,800 nodes (76,800 cores) of the Oakleaf-FX system.

Masatoshi Kawai  
Supercomputing Research Division  
The University of Tokyo  
kawai@cc.u-tokyo.ac.jp

Akihiro Ida  
The University of Tokyo  
ida@cc.u-tokyo.ac.jp

Kengo Nakajima  
The University of Tokyo  
Information Technology Center  
nakajima@cc.u-tokyo.ac.jp

## MS29

### Fast Matrix-Free High-Order Discontinuous Galerkin Kernels: Performance Optimization and

## Modeling

We present performance optimization of matrix-free discontinuous Galerkin schemes implemented via sum factorization evaluation techniques originating from spectral elements. Performance of the matrix-vector product of the symmetric interior penalty DG operator can come within a factor of 2 of finite difference stencils on Cartesian grids, while naturally extending to complex meshes and arbitrary differential operators. We provide the algorithmic steps to reach more than 500 GFLOP/s and 100 GByte/s measured throughput on two-socket Intel Broadwell systems.

Martin Kronbichler  
Technische Universität München  
kronbichler@lnm.mw.tum.de

Niklas Fehn  
Technical University of Munich  
fehn@lnm.mw.tum.de

Katharina Kormann  
Max Planck Institute for Plasma Physics  
katharina.kormann@ipp.mpg.de

Wolfgang A Wall  
Technical University of Munich  
wall@lnm.mw.tum.de

## MS29

### Roofline: a Throughput Oriented Performance Model

The Roofline performance model is a visually intuitive methodology for extracting key characteristics of scientific applications and upper performance bounds across a broad variety of modern processor architectures. This simple modeling philosophy allows the abstraction of complex nonuniform memory hierarchies and identifies code regions best suited for targeted optimizations. As a result, Roofline has been widely adopted as an effective performance modeling tool within the HPC community. This talk will present a brief overview of the Roofline philosophy and discuss how this model can be used to drive optimization efforts. We will then present several real world examples of successful Roofline utilization to isolate and improve performance of scientific computations. A key challenge of leveraging the Roofline approach has been the lack of automation in generating the necessary detailed code characterization. To address this limitation we have been working collaboratively to integrate Roofline model generation into the Intel Advisor tool suite, and will present highlights of these forthcoming capabilities.

Leonid Oliker, Tuomas Koskela  
Lawrence Berkeley National Laboratory  
loliker@lbl.gov, tkoskela@lbl.gov

Jack Deslippe  
National Energy Research Scientific Computing Center  
jrdeslippe@lbl.gov

Zakhar Matveev  
Intel  
zakhar.a.matveev@intel.com

Samuel Williams  
Lawrence Berkeley National Laboratory

swwilliams@lbl.gov

### MS30

#### Resilience With Asynchronous Many Task (AMT) Programming Models

The resilience of programming systems to deal with frequent hardware faults is expected to be a major challenge towards enabling high-performance and high-productivity programming for future generation computing systems. Asynchronous Many Task (AMT) programming has been proposed as a viable model to address these challenges. The Open Community Runtime (OCR) supports an AMT programming model based on the expression of explicit task dependencies. Its event-driven execution model and runtime knowledge of task dependencies provides a platform to explore AMT resilience as an alternative to traditional checkpoint-restart techniques. In this work, the user program in OCR expresses DAG-based task parallelism as a hybrid mix of resilient and non-resilient tasks. Resilient tasks are intended to be 'replayable' and constitute the outermost tasks in the program. Non-resilient tasks exist under the scope of resilient tasks. The OCR API has been extended to include flags that identify resilient tasks, data and events during their creation. The runtime records appropriate metadata and user data in resilient/persistent storage. To recover from failure, pending replayable tasks are identified by the runtime and re-executed on surviving resources as an online and asynchronous recovery technique. Our experiments have demonstrated recovery from node failures injected while running distributed stencil-based codes.

Sanjay Chatterjee  
General Atomics  
sanjay.chatterjee@gmail.com

Keita Teranishi  
Sandia National Laboratories  
knieran@sandia.gov

Vivek Sarkar  
College of Computing  
Georgia Institute of Technology  
vivek.sarkar@cc.gatech.edu

### MS30

#### Making HPC Systems Resilient with Parallel Objects

Modern supercomputers play a fundamental role in tackling grand challenges in science and engineering. High demanding computer simulations allow researchers to inspect complex phenomena in a wide variety of domains. To keep us with such demand, supercomputers have recently grown in size at an exponential rate. Big parallel machines are incredibly powerful, but also vulnerable to system errors. Sometimes components fail hard, sometimes silently. In either case, applications may execute slowly, or even worse, generate an incorrect result. The parallel-objects paradigm has demonstrated over the years its versatility to protect an HPC system to many types of errors. In this talk, we will present a group of fault tolerance techniques based on parallel objects and implemented in Charm++.

Esteban Meneses  
Costa Rica Institute of Technology  
emeneses@ic-itcr.ac.cr

Laxmikant Kale  
University of Illinois at Urbana-Champaign  
kale@illinois.edu

### MS30

#### Designing and Modelling Selective Replication for Fault-tolerant HPC Applications

Fail-stop errors and Silent Data Corruptions (SDCs) are the most common failure modes for HPC applications. There are studies that address fail-stop errors and studies that address SDCs. However, few studies address both types of errors together. In this talk, we present a software-based selective replication technique for HPC applications for both fail-stop errors and SDCs. Since complete replication of applications can be costly in terms of resources, we develop a runtime-based technique for selective replication. Selective replication provides an opportunity to meet HPC reliability targets while decreasing resource costs. Our technique is low-overhead, automatic and completely transparent to the user. For efficient implementation, we first develop a formal framework where we model the reliability of HPC programs in general and replication in particular via a theoretical model based on Markov Chains. We then design and implement a selective replication heuristic which is based on our Markov model estimating the reliability of the applications at runtime. Our heuristic, called Target Rep, selects the tasks that would increase the reliability of the application the most while obeying an overall system-wide replication percentage threshold - this heuristic is useful in systems having limited spare resources for replication. Our results show that it performs close to the optimum solutions staying within 5% of the optimum solutions with 50% target replication percentage.

Omer Subasi  
Pacific Northwest National Laboratory  
omer.subasi@pnnl.gov

### MS30

#### Fault Tolerance Features of a Multi SPMD Scientific Workflow Programming Model

In the past, we have proposed a multi SPMD scientific workflow programming model to exploit the performance of systems in the exa-scale era, which would consist of a huge number of nodes arranged in a multi-level hierarchy. In order to achieve the fault tolerance, we extend the workflow scheduler and middleware to detect an error and to recover from the error. We present experimental results on K computer and further extension for a large scale system.

Miwako Tsuji  
RIKEN Advanced Institute for Computational Science  
miwako.tsuji@riken.jp

### MS31

#### The Challenges of Formally Proving HPC Systems Software

As HPC systems reach extreme scale the challenges of finding, understanding and fixing bugs becomes increasingly intractable. There has been significant work in the systems security community about creating "Deep Specifications" (DeepSpec) of system behavior that can be used to verify the behavior of systems software in ways that avoid unspecified software behavior as well as providing formal

failure models which allows software to fail and recover without data loss. This talk will discuss potential applications of the DeepSpec methodology to HPC systems and the systems software challenges involved.

Noah Evans

Sandia National Laboratories  
1978  
nevans@sandia.gov

### MS31

#### Recent Advances in the IHK/McKernel Lightweight Multi-Kernel Operating System

RIKEN Advanced Institute for Computation Science is in charge of leading the development of Japan's next generation flagship supercomputer, the successor of the K. Part of this effort is to design and develop a system software stack that suits the needs of future extreme scale computing. In this talk, we primarily focus on operating system (OS) research and discuss IHK/McKernel, a multi-kernel based operating system framework. IHK/McKernel runs Linux with a light-weight kernel (LWK) side-by-side on compute nodes with the primary motivation of providing scalable, consistent performance for large scale HPC simulations, but at the same time to retain a fully Linux compatible execution environment. We provide an overview of the project and discuss some of our recent developments addressing hierarchical memory management and partial device drivers in multi-kernel systems.

Balazs Gerofi

Riken  
bgerofi@riken.jp

### MS31

#### Operating and Runtime Systems Challenges for HPC Systems

System software is shaped by five major factors: the hardware that needs to be supported, the shared services that are available to applications, the overall system usage model, the collection of applications that need to be supported, and history. For the most part, all of these factors have been stable since the early 90s when the community embraced massively parallel systems built from commodity processors. In the past few years, we have seen an end to this stability in each of these factors. What does this mean for system software and how do we move forward without forcing unnecessary constraints?

Arthur McCabe

ORNL  
maccabeab@ornl.gov

### MS31

#### Vanguard: Maturing the Arm Software Ecosystem for U.S. DOE Supercomputing

ARM technologies present unique possibilities for co-design of large-scale HPC systems. The U.S. Department of Energys (DOEs) National Nuclear Security Administration is embarking on an effort to mature the ARM ecosystem for its Advanced Simulation and Computing (ASC) workloads, with a large-scale system deployment planned for 2019. This talk will present the key software stack requirements for this system and our plans for developing an efficient and productive ARMv8 programming environment. Focus areas will include overall integration and robustness

of the stack, addressing known challenges such as compilers and math libraries, and improving support for current HPC workflows and emerging integrated HPC/Data Analytic Computing (DAC) workloads. Working in collaboration with the ARM community and system vendors, our desire is to create an open and integrated software stack that is scalable to the largest supercomputers at DOE and elsewhere.

Kevin Pedretti

Sandia National Laboratories  
ktpedre@sandia.gov

### MS32

#### Developing An Extensible, Portable, Scalable Toolkit for Massively Parallel Incompressible SPH

The stability, accuracy, energy conservation, boundary conditions of the projection based particle method such as incompressible smoothed particle hydrodynamics ISPH have been greatly improved. However, there are still many challenges compared with other particle based methods from the perspective of computation and high performance software implementation when using hundreds of millions of particles above. In this talk, we are particularly concerning the scalable algorithms for the post peta-scale particle method based simulations, these algorithms are low overhead domain decomposition and dynamic load balancing involving irregular particle distributions and complex geometries, flexible parallel communications algorithms to facilitate user scientific software development. Particles ordering for cache-based computing architectures and reducing the sparse matrix bandwidth and efficient sparse linear solvers which is the additional distinct challenge for projection-based particle methods. The implementation details introduced here are intended to form future guidance for the new projection-based particle application development.

Xiaohu Guo, Xiaohu Guo

Hartree Centre  
Science and Technology Facilities Council  
xiaohu.guo@stfc.ac.uk, xiaohu.guo@stfc.ac.uk

### MS32

#### A Lagrangian Inertial Centroidal Voronoi Particle Method for Dynamic Load Balancing in Particle-Based Simulations

In this paper, we develop a Lagrangian Inertial Centroidal Voronoi Particle (LICVP) method to extend the original CVP method [1] to the dynamic load balancing in particle-based simulations. Two new concepts are proposed to address the additional targets encountered in repartitioning the system. A background velocity is introduced to transport Voronoi particle according to the local fluid field, which facilitates the data reuse and lower data redistribution cost during rebalancing. An Inertial CVP method is developed as well to handle problems with skew-aligned computational load. The inertia matrix is utilized to characterize the load distribution, and a splitting operator is introduced to confine Voronoi particle motion. Moreover, by defining an adaptive filter, the partitioning strategy can be varied dynamically. Intensive numerical tests in fluid dynamics simulations reveal that the underlining LICVP method improves the incremental property remarkably without making sacrifices in other objectives, i.e. the inter-processor communication is optimized simul-

taneously, and repartitioning procedure is highly efficient.

Zhe Ji, Lin Fu, Xiangyu Hu, Nikolaus Adams  
 Technical University of Munich  
 zhe.ji@tum.de, lin.fu@tum.de, xiangyu.hu@tum.de, nikolaus.adams@tum.de

## MS32

### OpenFPM: An Open-source Framework for Highly Scalable Hybrid Particle-mesh Simulations

The performance of computing hardware is steadily increasing, enabling simulations of increasingly complex models. However, efficiently implementing scalable simulation codes on heterogeneous, distributed hardware systems becomes the bottleneck. This bottleneck can be alleviated by intermediate software layers that provide higher-level abstractions closer to the problem domain, hence allowing the computational scientist to focus on the simulation. Here, we present OpenFPM, an open and scalable C++ framework that provides an abstraction layer for numerical simulations using particles and/or meshes, handles problems of arbitrary dimension and uses compile-time template meta-programming to support arbitrary user-defined classes and data types. OpenFPM provides transparent and scalable infrastructure for shared-memory and distributed-memory implementations of hybrid particle-mesh simulations of both discrete and continuous models. We present the architecture and design of OpenFPM, detail the underlying abstractions, and benchmark the framework in applications ranging from Smoothed Particle Hydrodynamics (SPH) to Molecular Dynamics (MD) to Reaction-Diffusion solvers, comparing it to the current state of the art and existing software frameworks.

Ivo F. Sbalzarini  
 Center for Systems Biology Dresden  
 TU Dresden, Faculty of Computer Science  
 ivos@mpi-cbg.de

Pietro Incardona  
 TU Dresden, Germany  
 incardon@mpi-cbg.de

## MS32

### Swift: Using Task-Based Parallelism, Asynchronous MPI and Dynamic Workload-Based Domain Decomposition to Achieve Near-Perfect Load-Balancing for Particle-Based Hydrodynamics and Gravity

Traditional large HPC simulation codes rely on MPI or MPI+OpenMP for their parallelisation over clusters of 100k cores and more. This approach is very successful for applications with a large degree of regularity in their data but alternative ways are required for more irregular algorithms such as particle methods like Smoothed Particle Hydrodynamics. In this talk, I will present the task-based parallelism strategy used in the open-source cosmological code SWIFT. This code has been demonstrated to run efficiently on 200k+ cores (20k+ nodes) and makes use of vectorization at the lowest level of parallelization. I will specifically focus on the aspects that make SWIFT unique. We simultaneously

- Use asynchronous point-to-point MPI communications
- Use dynamic scheduling of tasks to accelerators
- Use task-based i/o

The combination of these points means that the FPU/VPUs, inter-connect and i/o sub-systems are all used *at the same time*, implying a full usage of all the components of an HPC cluster without explicit barriers for i/o. The combination of these parallelisation strategies with better algorithms for neighbour finding leads to a speed-up of *25x* when compared to the de-facto standard Gadget code widely used in the astrophysics community.

Matthieu Schaller  
 Department of Physics  
 Durham University  
 matthieu.schaller@durham.ac.uk

## MS33

### GOFMM: A Parallel Geometry-oblivious N-body Algorithm for Sketching SPD Matrices

N-body algorithms are ubiquitous in science and engineering and are a core computational primitive for many production codes in simulation and data analysis. N-body algorithms require the problem to have certain structure, for example we need points and a pairwise function between points that defines the kernel structure. Here we generalize N-body method to arbitrary symmetric positive definite matrices. We present GOFMM (geometry-oblivious fast-multipole methods), a novel method that creates a hierarchical low-rank approximation, or sketching, of an N-by-N arbitrary SPD matrix. For many applications, GOFMM enables an approximate matrix-vector multiplication in  $O(N \log N)$  or even  $O(N)$  time. Sketching requires  $O(N \log N)$  storage and work. In general, our scheme belongs to the hierarchical matrix approximation methods. In particular, it generalizes the fast multipole method (FMM) to a purely algebraic setting by only requiring the ability to sample matrix entries. Neither geometric information (i.e., point coordinates) nor knowledge of how the matrix entries have been generated is required, thus the term geometry oblivious. Also, we introduce a shared-memory parallel scheme for hierarchical matrix computations that reduces synchronization barriers. We present results on the Intel Knights Landing and Haswell architectures, and on the NVIDIA's Pascal architecture for a variety of matrices.

Chenhan Yu  
 University of Texas at Austin  
 chenhan@cs.utexas.edu

James Levitt  
 The University of Texas at Austin  
 jlevitt@ices.utexas.edu

Severin Reiz  
 Technical University of Munich  
 s.reiz@tum.de

George Biros  
 The Institute for Computational Engineering and Sciences  
 The University of Texas at Austin  
 biros@ices.utexas.edu

## MS33

### New Variations of a Fast Hierarchical Low-Rank Solver for Sparse Systems

Hierarchical low-rank methods have proven useful for a wide range of problems, including elliptic PDEs. We explore new variations of a recent hierarchical low-rank solver

algorithm by Pouransari et al.

Erik G. Boman

Center for Computing Research  
Sandia National Labs  
egboman@sandia.gov

Chao Chen

Stanford University  
cchen10@stanford.edu

Eric F. Darve

Stanford University  
Mechanical Engineering Department  
darve@stanford.edu

Siva Rajamanickam

Sandia National Laboratories  
srajama@sandia.gov

Ray S. Tuminaro

Sandia National Laboratories  
Computational Mathematics and Algorithms  
rstumin@sandia.gov

### MS33

#### An Overview of Fast Sparse Linear Solvers Based on Approximate Factorizations

Solving large-scale linear systems remains a challenging problem. For example, in the case of unsymmetric matrices or even indefinite symmetric matrices, few preconditioners with strong convergence guarantees are available. Many general purpose preconditioning techniques, such as the incomplete LU factorization, exist to address this problem. In recent years, a new class of preconditioners has been developed based on approximate factorizations. Contrary to the incomplete LU factorization, instead of dropping small entries, low-rank matrices are used to approximate certain blocks in the LU factors. This typically leads to algorithms that have linear complexity in the matrix size. We will review important developments in this area and open problems. Examples of techniques that will be discussed include HODLR,  $\mathcal{H}$ -matrices, HSS,  $\mathcal{H}^2$  matrices, FMM matrices, and tensor-train decompositions.

Eric Darve

Institute for Computational and Mathematical Engineering  
Stanford University  
eric.darve@stanford.edu

Erik G. Boman

Center for Computing Research  
Sandia National Labs  
egboman@sandia.gov

Chao Chen

Stanford University  
cchen10@stanford.edu

Mathieu Faverge

Bordeaux INP - Labri - Inria  
mathieu.faverge@inria.fr

Siva Rajamanickam

Sandia National Laboratories  
srajama@sandia.gov

Pierre Ramet

Bordeaux University - INRIA  
pierre.ramet@inria.fr

Jean Roman

INRIA  
Jean.Roman@inria.fr

Toru Takahashi

Nagoya University  
ttaka@nuem.nagoya-u.ac.jp

### MS34

#### Scalable Graph Alignment on Modern Architectures

We investigate the graph alignment problem, in which each vertex in a graph is identified with a vertex in another graph. This problem arises in many data analysis applications when data from heterogeneous sources are integrated. Most of the existing approaches to this problem are based on the global topology of underlying graphs. More recent ones exploit vertex and edge attributes for improved alignment. Usually, the global approaches are limited to graphs having less than tens of thousands vertices. We propose techniques in which the global structure of the graphs are used to reduce the problem size while maintaining the alignment accuracy. Further techniques to create easily parallelizable pairwise comparisons are proposed. These techniques are integrated in a data-parallel, architecture (resource)-aware scalable graph alignment framework. The framework takes the deep memory hierarchy into account, and scales from a single node to a distributed memory execution. The talk presents experiments with this framework.

Umit V. Catalyurek, Abdurrahman Yasar

Georgia Institute of Technology  
School of Computational Science and Engineering  
umit@gatech.edu, ayasar@gatech.edu

Bora Ucar

LIP, ENS-Lyon, France.  
Bora.Ucar@ens-lyon.fr

### MS34

#### Processing Massive Scale-Free Graphs on HPC with NVRAM

Many graphs of interests contain irregular structure and high-degree "hub" vertices (e.g., social networks, web graphs, etc.), and such irregular structure creates scaling challenges for parallel and distributed analytics on large-scale datasets. For many applications, the degree distribution of the vertices asymptotically follows a power law, with the "hub" vertices having orders of magnitude more connections than the average vertex in the graph, often referred to as a scale-free graph. Our work develops parallelization and distribution techniques aimed at current and future high performance computing (HPC) systems with tightly coupled interconnects, deep memory hierarchies, and high levels of concurrency. Additionally, we leverage node-local NVRAM storage to persistently store the topology and metadata of datasets too large to fit in main-memory. We present our ongoing research into static, streaming, and temporal graph analytics at scale on HPC systems. Our tools presented are open source, and can be found at: <http://software.llnl.gov/havoqgt/>.

Roger Pearce



Lawrence Livermore National Laboratory  
rpearce@llnl.gov

#### MS34

##### Exploiting High Bandwidth Memory for Graph Algorithms

High bandwidth memory (HBM) is a memory layer between DDR and cache which provides an additional fast memory level for certain high performance compute platforms. The goal of this work was to investigate various methods for utilizing HBM to accelerate graph analysis algorithms. We use label propagation as a prototypical graph algorithm, which allows us to investigate both the impact on solution quality and speedup of our various approaches. By selectively breaking up work into iterative HBM-resident chunks, we demonstrate that significant speedup can result with minimal to no degradation of quality.

George M. Slota  
Rensselaer Polytechnic Institute  
slotag@rpi.edu

Siva Rajamanickam, Cynthia Phillips, Jonathan Berry  
Sandia National Laboratories  
srajama@sandia.gov, caphill@sandia.gov,  
jberry@sandia.gov

#### MS34

##### TriData: High Performance Linear Algebra-Based Data Analytics

TriData is high performance computing software that uses linear algebra algorithms to solve important hypergraph and graph problems in data analytics. It leverages the Trilinos frameworks linear algebra structures and algorithms to solve very large graph and hypergraph problems efficiently in parallel. On the node, TriData relies on Kokkos and KokkosKernels to obtain good parallel performance on different architectures (multi-core, many-core, GPUs). We present results and discuss challenges related to these architectures.

Michael Wolf  
Sandia National Laboratories  
mmwolf@sandia.gov

Daniel M. Dunlavy  
Sandia National Laboratories  
Computer Science and Informatics  
dmdunla@sandia.gov

Rich Lehoucq, Jonathan W. Berry  
Sandia National Laboratories  
rblehou@sandia.gov, jberry@sandia.gov

Daniel Bourgeois  
Rice University  
dcb\_dan@hotmail.com

#### MS35

##### StarPU and KStar, Synergies Between a Task-Based Runtime System and An OpenMP Compiler

Abstract Not Available At Time Of Publication.

Olivier Aumage

INRIA, France  
olivier.aumage@inria.fr

#### MS35

##### Task-Based Parallel Programming with Py-COMPSs

Abstract Not Available At Time Of Publication.

Rosa M. Badia  
Barcelona Supercomputing Center / CSIC  
rosa.m.badia@bsc.es

Javier Conejero  
Barcelona Supercomputing Centre  
Spain  
francisco.conejero@bsc.es

#### MS35

##### The C++ Parallel Programming Model

Abstract Not Available At Time Of Publication.

Bryce Lelbach  
NVidia  
brycellelback@gmail.com

#### MS35

##### Optimizing Recursive Programs for Memory Hierarchy Locality

Automatically optimizing the cache locality of programs across time steps (e.g., time-tiling stencils) is a difficult problem that traditionally requires strong static compiler analysis or re-writing the kernel in a DSL to expose data dependencies. In this talk, we will demonstrate how recursive programs can be augmented with data effect annotations to automatically identify concurrency and data reuse opportunities across function invocations and thereby interleave them to reduce reuse distance. We present algorithms that efficiently track (sub-settable) effects in recursive programs, detect interference and dependencies, and interleave execution with user-level (non-kernel) lightweight threads. To enable multi-core execution, a program is parallelized using a nested fork/join programming model. The data effect annotations allow the runtime to override the strictness of the fork/join parallelism, exposing more concurrency and improving data locality, while retaining the benefits of fork/join parallelism.

Jonathan J. Lifflander  
Sandia National Labs  
jliffa@sandia.gov

#### MS36

##### Overview of Kokkos

Abstract Not Available At Time Of Publication.

H. Carter Edwards, Christian Trott, Daniel Sunderland,  
Daniel Ibanez, Nathan Ellingwood  
Sandia National Laboratories  
hcedwar@sandia.gov, crtrott@sandia.gov,  
dsunder@sandia.gov, daibane@sandia.gov,

mailto:ndellin@sandia.gov

### MS36

#### Implementing the Lattice Quantum Chromodynamics Wilson-Dslash Kernel Using Kokkos

The Wilson-Dslash kernel is a covariant finite difference operator describing the coupling of quarks and gluons in simulations of lattice Quantum Chromodynamics (LQCD). However, high performance implementations are usually not performance portable or even portable and frequently rely on low level code generation frameworks which generate intrinsics, or assembly. Further, Dslash is typically used as a component of linear operators in the context of linear solvers, which also need to be rewritten for each new implementation, a cumbersome task for more complicated solvers such as Multi-Grid approaches. In this talk, we report on implementing the Wilson-Dslash Kernel using Kokkos, for Intel Xeon Phi, Knights Landing (KNL) and Pascal (P100) and Volta (V100) GPUs. We carry out a performance analysis of the Kokkos implementations and also compare the performance with known high performance implementations such as the QPhiX the QUDA libraries for KNL and GPUs respectively. We find that how the operator is implemented can have an important effect on performance portability. A secondary consideration is the implementation of vectorization in a performance portable way. Our naive implementation on GPUs can reach 80% and 98% of the performance of the highly optimized QUDA Library on Pascal (P100) and Volta (V100) GPUs respectively. To get good performance on KNL we had to code forms of the operator with explicit vectorization (including using a few intrinsics).

Balint Joo

Jefferson Laboratory  
bjoo@jlab.org

Thorsten Kurth  
NERSC  
tkurth@lbl.gov

Jack Deslippe  
National Energy Research Scientific Computing Center  
jrdeslippe@lbl.gov

Kate Clark  
NVIDIA  
mclark@nvidia.com

Daniel Ibanez, Daniel Sunderland  
Sandia National Laboratories  
daibane@sandia.gov, dsunder@sandia.gov

### MS36

#### Performance Portability in SPARC Using Kokkos in a Compressible CFD Code

SPARC is a compressible computational fluid dynamics (CFD) code being developed to support aerodynamic analyses at Sandia National Laboratories and uses Kokkos to enable execution on a variety of advanced architectures. The software design considerations we encountered and addressed to attain performance portability will be discussed, along with challenges that remain. Performance analyses will be presented for Intel Xeon (HSW/BDW), Xeon Phi (KNL) and Nvidia Pascal (P100) systems. A candid discussion of the high and low points of large-scale ap-

plication code development in the context of performance portability will also be given.

James Overfelt, Micah Howard, Steve Bova  
Sandia National Laboratories  
jroverf@sandia.gov, mhoward@sandia.gov,  
swbova@sandia.gov

Travis Fisher  
Sandia National Labs  
tcfishe@sandia.gov

Mark Hoemmen, Derek Dinzl, Andrew Bradley, Ross Wagnild  
Sandia National Laboratories  
mhoemme@sandia.gov, djdinzl@sandia.gov, ambradl@sandia.gov, rmwagni@sandia.gov

### MS36

#### KokkosKernels: Math Building Blocks for Performance Portable Applications

KokkosKernels is a library providing fundamental math capabilities for dense and sparse linear algebra as well as graph algorithms based on the Kokkos C++ programming model. Its interface makes full use of Kokkos' data abstractions, allowing for inputs with different data layouts as well as memory spaces. While KokkosKernels aims at providing high performance implementations in itself, it also serves as an interface to third party libraries taking into account data layouts and memory spaces. Furthermore the library starts to address math needs on all levels of the hardware hierarchy, from full socket utilization, via thread groups to vectorization. In this talk an overview of current capabilities of KokkosKernels will be given and data supporting its performance portability claim is shown.

Christian Trott, Siva Rajamanickam, Kyungjoo Kim, Mehmet Deveci  
Sandia National Laboratories  
crtrott@sandia.gov, srajama@sandia.gov,  
kyukim@sandia.gov, mnndevci@sandia.gov

### MS37

#### Quantum Chemistry Methods with Gaussian Basis

This talk will describe some recent methodological and algorithmic development as part of the NWChem and NWChemEx projects. NWChem is an open source software package that combines several different computational chemistry approaches. The software development of NWChem has always been driven by the importance of exploiting the aggregate resource of parallel computers. Therefore NWChem can efficiently use the computing power of HPC resources (e.g. Supercomputer, departmental clusters, accelerators, etc ...) The NWChemEx project is currently redesigning the architecture of NWChem to work with the pre-exascale and exascale computers. The talk will also discuss the scientific output enabled by this combination of software and hardware. For example, we will report of the use of embedded cluster approach to model the ground state and excited state properties of crystalline compounds. The theoretical methods analyzed in the talk will range from Density-Functional based to wavefunction based (e.g. Coupled Cluster).

Edoardo Apra  
Pacific Northwest National Laboratory

edoardo.apra@pnnl.gov

### MS37

#### Applications of Matrix Functions to Electronic Structure Calculations

In this talk, we will review the applications of matrix functions to electronic structure calculations. In particular, we will focus on the use of matrix functions as diagonalization free methods for linear scaling calculations. We will also present our newly developed library for computing matrix functions, and show how it can be integrated into a variety of programs to enable large scale calculations across a wide range of fields.

William Dawson

Riken Advanced Institute for Computational Science  
william.dawson@riken.jp

### MS37

#### 100-Nano-Meter-Scale Electronic Structure Calculation for Organic Device Materials

Large-scale electronic state calculations were realized by novel linear-algebraic algorithm that is based on generalized shifted linear equations, instead of conventional generalized eigenvalue equations. The benchmark shows an extreme strong scaling on the full system of the K computer. The method was applied to material research for ultra-flexible (organic) devices or the next-generation Internet-of-Things (IoT) products like display, sensor, and battery. The related data scientific approach was also used.

Takeo Hoshi

Tottori University  
hoshi@damp.tottori-u.ac.jp

### MS37

#### Large Scale Discontinuous Galerkin Density Functional Theory (DGDFT) Method on GPU Supercomputer

In this talk, I will present our recent work on large scale discontinuous Galerkin density functional theory (DGDFT) method on GPU supercomputer. The parallelization and overall speedup will be presented.

Weile Jia

University of California at Berkeley  
wjia@lbl.gov

### MS38

#### Scalable Hyperparameter Search for Deep Learning in Scientific Applications

Many deep learning (DL) algorithms require user-specified values for hyper parameters that strongly influence the performance of these algorithms such as training time, prediction accuracy, and even scalability. These parameters include numerical as well as categorical parameters. The former include number of hidden layers, number of units per layer, sparsity/overfitting regularization parameters, batch size, and learning parameters such as rate, schedule, etc. The latter are often used to determine which of a number of interchangeable components are to be used, such as type of type of initialization, training method, activation functions, loss functions. Nevertheless, finding performance-optimizing parameter settings is a difficult task. In the past

few years, new algorithmic methods including Bayesian and multi arm bandit approaches have begun to emerge to address this problem. In this talk, we will present a scalable model-based search approach that is suitable for large-scale hyperparameter search involving 100s of nodes on leadership-class systems. We will illustrate the effectiveness of the search algorithm for deep learning applications in high energy physics, climate modeling, and environment science. We will provide advantages and limitations of the search algorithm with respect to the accuracy, scalability, ease of use, effectiveness with respect to number of parameters, and support for nested hyperparameters.

Prasanna Balaprakash

Argonne National Laboratory  
pbalapra@mcs.anl.gov

### MS38

#### Trends in Parallel and Distributed Deep Learning

Deep Neural Networks (DNNs) are becoming an important tool in modern computing applications. The main success of DNNs bases on the available compute power in modern devices. Speeding up their core computation is a major challenge in computer science and approaches range from decentralized distributed algorithms to low-level circuit design. In this talk, we briefly outline the problem of DNN inference and training from a theoretical perspective, followed by the various approaches for their parallelization. In particular, we present recent trends in DNN architectures, and the resulting implications on parallelism strategies and memory considerations. We then discuss the different types of parallelism exhibited in deep and recurrent neural networks; synchronous and asynchronous stochastic gradient descent, as well as variants thereof; distributed system architectures, communication schemes and implementations; network topology and I/O concerns; performance modeling; and custom hardware architectures for inference and training. Based on the discussed approaches, we extrapolate potential future directions for parallelism in the field of neural networks and its applications in machine learning and artificial intelligence.

Tal Ben-Nun

ETH Zurich  
tal.bennun@inf.ethz.ch

### MS38

#### Be Fast & Productive: Accelerating Deep Learning Without Sacrificing Productivity

Deep learning algorithms have recorded significant results in a wide range of machine learning tasks, including image and audio recognition, anomaly detection, and image generation. These algorithms require a vast amount of computing resources and optimizing performance is critical. Learning processes typically take days or weeks even on multiple GPUs. Also, to obtain a good model, one needs to explore a large parameter space, such as the structure of the neural network models and meta-parameters like the learning rate. Given the research is changing rapidly and the competition is severe, accelerating the cycle of algorithm design, code development, and experiments is critical. In this talk, we introduce Chainer and ChainerMN, a flexible deep learning framework written in Python. Chainer is designed to be flexible and intuitive to minimize code development time, and ChainerMN can scale Chainer applications to tens or hundreds of GPUs to reduce the runtime of experiments. We describe how we

apply performance optimization and HPC techniques to the dynamic and flexible framework to achieve both speed and high productivity.

Keisuke Fukuda  
Preferred Networks  
kfukuda@preferred.jp

Shuji Suzuki, Takuya Akiba  
Preferred Networks, Inc.  
ssuzuki@preferred.jp, akiba@preferred.jp

### MS38

#### **Faster, Smaller, and More Energy-Efficient Inference Using Codebook-based Quantization and FPGAs**

We are investigating novel methods to reduce the execution time, space, and energy of inference in deep neural networks by combining algorithm-level innovations, such as codebook-based quantization for convolutional neural networks compression, with reconfigurable hardware, via field-programmable gate array (FPGA) processors, achieving accuracy close to uncompressed networks. We plan to evaluate this work in the Caffe2 deep learning framework and test on real FPGAs, building on recent advances in automatic OpenCL-to-FPGA translation.

Mikhail Isaev, Jeffrey Young, Rich Vuduc  
Georgia Institute of Technology  
michael.v.isaev@gatech.edu, jyoung9@gatech.edu,  
richie@cc.gatech.edu

### MS39

#### **Optimising Massively Parallel Lattice-Boltzmann Codes for Use in Time-Critical Clinical Blood Flow Modelling**

Cardiovascular disease is a major cause of death in the developed world, particularly among ageing populations. Haemorrhagic strokes, for example, occur when a congenital weakness in the blood vessel wall known as an aneurysm - bulges out and ruptures, causing bleeding into the surrounding brain tissue. Such events are difficult to predict, but appear to be associated with unusually high or low wall shear stress exerted by the blood fluid, and as such Computational Fluid Dynamics (CFD) has been proposed as a tool for informing clinical treatment decisions. This allows patient specific pre-interventional prediction of the short and long range effects of the introduction of flow diverters to treat aneurysms. However, the high resolution required for clinical accuracy, along with the comparative largeness of the intracranial vasculature coupled with the requirement for fast results present a uniquely challenging system. Extreme-scaling, efficient CFD codes are therefore essential for the transition to effective clinical decision-making. We present recent scaling and performance work with HemeLB, an advanced lattice-Boltzmann code, to hundreds of thousands of cores. We also present a performance analysis of a large human brain simulation containing magnetically steered colloidal particles for use in drug targeting and visualisation.

Peter Coveney  
University College London, United Kingdom

p.v.coveney@ucl.ac.uk

### MS39

#### **Development of LBM Code Toward Large-Scale Aerodynamics/Aeroacoustics Simulations**

In this paper, we propose a computational approach based on Lattice Boltzmann Method (LBM) and apply the method to the large-scale aerodynamic simulations around complex geometries. In our approach, we couple LBM and Building-Cube Method (BCM), which is the framework of a block-structured Cartesian grid approach proposed by Nakahashi, to enhance the parallel efficiency. To overcome the stability problem due to the treatment of collision operator in standard LBGK model, cascaded-LBM approach is used. The proposed method is applied to several 2D/3D problems for the validations and verifications.

Takashi Ishida  
Japan Aerospace Exploration Agency  
tishida@chofu.jaxa.jp

### MS39

#### **OpenLB: Parallel Open Source Lattice Boltzmann Software**

Many real-world problems in chemical engineering and biotechnology cannot be solved with the help of today's measurement and numerical simulation technology, despite significant progress in both areas. This holds in particular for the detection of time-dependent flows in complex flow domains, in which the geometry is coupled with the flow behavior, through complex biological, chemical or mechanical processes at the flow-solid interface. In this area there are a number of unanswered fundamental questions, which can only be answered by a combined method of measurement and simulation techniques. For it, inverse 3D fluid flow problems needs to be solved. However, to this day, such methods are only partially explored and software is hardly available, both, most likely, because of the enormous associated computational cost. In this talk, a holistic simulation and optimization approach based on Lattice Boltzmann Methods (LBM) is presented and the corresponding extensive parallel software is introduced. A demonstration of the open source library OpenLB ([www.openlb.net](http://www.openlb.net)) relies on a number of process engineering applications, such as a particle separator with a turbulent particle flow, magnetic separation with adsorption, growth of algae in photobioreactors or the turbulent simulation of a safety valve.

Mathias J. Krause  
Karlsruhe Institute of Technology (KIT)  
mathias.krause@kit.edu

### MS39

#### **Large-Scale Lattice Boltzmann Simulations of Particulate Flows**

Particulate flows are encountered in various fields, examples being sediment transport in riverbeds relevant in environmental engineering and fluidized beds in chemical engineering. The simulation of such flows is based on fluid-structure coupling mechanisms that transfer momentum from the particulate to the fluid phase and vice versa. In this talk, we will focus on geometrically fully resolved particles and discuss suitable fluid-solid coupling algorithms to enable accurate predictions from first principles [Ret-

tinger & R  de, A comparative study of fluid-particle coupling methods for fully resolved lattice Boltzmann simulations, CaF, 2017]. The lattice Boltzmann method with a two relaxation time collision model is used to represent the fluid dynamics. We will highlight parallelization aspects essential for massively parallel execution and discuss the performance of our approach. In combination with grid refinement techniques, large scale simulations of particulate flows can be realized efficiently featuring several thousand particles interacting with the fluid flow [Rettinger et al., Fully Resolved Simulations of Dune Formation in Riverbeds, Springer, ISC, 2017]. Illustrating examples from the aforementioned application fields will be presented. The generality and flexibility of the approach can be used to investigate the influence of particle shapes and size distributions on such systems.

Christoph Rettinger

Friedrich-Alexander-Universit  t Erlangen-N  rnberg (FAU)  
christoph.rettinger@fau.de

Sebastian Eibl

Friedrich-Alexander-University Erlangen-Nuremberg  
sebastian.eibl@fau.de

Christian Godenschwager

University of Erlangen-Nuremberg  
christian.godenschwager@fau.de

Ulrich R  de

Friedrich-Alexander-University Erlangen-Nuremberg, Germany  
ulrich.ruede@fau.de

#### MS40

##### Scalable Krylov Methods for Spectral Graph Partitioning

We present a scalable and robust implementation of multi-level Lanczos to compute the Fiedler vector, the eigenvector for the smallest non-zero eigenvalue of the graph Laplacian. Scalability is improved by aggressively hiding communication by overlapping it with computation and other communication. Reorthogonalization and implicit restart are included for stability. We present a distributed memory nested dissection code based on this spectral partitioning algorithm, with comparisons of ordering quality and performance against ParMetis and PTScotch.

Pieter Ghysels

Lawrence Berkeley National Laboratory  
pghysels@lbl.gov

Siegfried Cools

University of Antwerp  
siegfried.cools@uantwerp.be

Roel Van Beeumen

KU Leuven  
rvanbeeumen@lbl.gov

Mathias Jacquelin

Lawrence Berkeley National Lab  
mjacquelin@lbl.gov

Esmond G. Ng

Lawrence Berkeley National Laboratory

engng@lbl.gov

#### MS40

##### Using Non-Blocking Communication to Achieve Scalability for Preconditioned Conjugate Gradient Methods

The preconditioned conjugate gradient method (PCG) can be optimized to perform better at scale by rearranging the method to overlap communication and computation with non-blocking allreduces and decrease the number of allreduces. We investigate standard PCG as well as four scalable variations. Experiments on up to 128k cores show that non-blocking PCG methods can outperform standard PCG at scale. We observe that the fastest method varies depending on work per core, suggesting that a suite of scalable solvers is needed to obtain the best performance. In particular non-blocking PCG methods are able to obtain improved performance at scale by hiding noise that can cause performance variation across cores. Weak scaling experiments show that non-blocking PCG methods are able to obtain more consistent performance than blocking PCG methods, suggesting that non-blocking methods are more resistant to the many sources of noise found on supercomputers. We use performance modeling and more detailed experimentation to better understand the impact of different types of network noise on communication algorithms commonly used by parallel numeric algorithms and on PCG solvers. These results help show how we can develop more efficient linear solvers and use them effectively in the performance irregular environments found on many supercomputers.

William D. Gropp, Paul Eller

University of Illinois at Urbana-Champaign  
Dept of Computer Science  
wgropp@illinois.edu, eller3@illinois.edu

#### MS40

##### Performance of S-Step and Pipelined Krylov Methods

We compare the performance of pipelined and s-step variants of a Krylov solver. Our implementations of both s-step and pipelined methods focus on reducing the cost of global all-reduce operations needed for the orthogonalization. We present performance results on a distributed-memory multicore CPUs to demonstrate the effects of these two methods.

Piotr Luszczyk

University of Tennessee  
luszczyk@icl.utk.edu

Ichitaro Yamazaki

University of Tennessee, Knoxville  
iyamazak@icl.utk.edu

Mark Hoemmen

Sandia National Laboratories  
mhoemme@sandia.gov

Jack J. Dongarra

University of Tennessee, Oak Ridge National Laboratory, USA



dongarra@icl.utk.edu

#### MS40

##### Impact of Noise Models on Pipelined Krylov Methods

In this work we present a performance model for Krylov and pipelined Krylov methods in a probabilistic setting. We let runtimes be stochastic to account for noise present in computer simulations and introduce the complication of uneven workload distributions, a common occurrence in evolutionary problems. We verify our assumptions through experimentation and perform some theoretical experiments to study the impact of noise on Krylov methods.

Hannah Morgan  
University of Chicago  
hannah.mairs@gmail.com

Matthew G. Knepley  
Rice University  
knepley@gmail.com

Patrick Sanan  
USI - Università della Svizzera italiana  
ICS  
patrick.sanan@gmail.com

Ridgway Scott  
Department of Computer Science  
University of Chicago  
ridg@cs.uchicago.edu

#### MS41

##### Isoefficiency in Practice: Configuring and Understanding the Performance of Task-Based Applications

Task-based programming offers an elegant way to express units of computation and the dependencies among them, making it easier to distribute the computational load evenly across multiple cores. Unfortunately, finding a good match between input size and core count usually requires significant manual experimentation. We present an automated empirical method for finding the isoefficiency function of a task-based program, which binds efficiency, core count, and the input size in one analytical expression.

Sergei Shudler  
Technische Universität Darmstadt  
shudler@cs.tu-darmstadt.de

Alexandru Calotoiu  
Laboratory for Parallel Programming  
TU Darmstadt  
calotoiu@cs.tu-darmstadt.de

Torsten Hoefer  
ETH Zürich  
htor@inf.ethz.ch

Felix Wolf  
Technische Universität Darmstadt  
wolf@cs.tu-darmstadt.de

#### MS41

##### Extending the Roofline Model: Bottleneck Analy-

##### sis with Microarchitectural Constraints

We present a generalized roofline analysis that identifies, for a given numerical program, hardware bottlenecks such as latency/throughput of caches in the memory hierarchy or the size of out-of-order buffers. The analysis uses DAG-based modelling and visualizes bottlenecks in a single plot with multiple rooflines. The bottleneck information can be used to guide the search for balance, i.e., a processor adapted to the program. The analysis is implemented in the tool ERM available at <http://www.spiral.net/software/bottleneck.html>.

Victoria Caparros Cabezas  
Department of Computer Science  
ETH Zurich  
caparrov@inf.ethz.ch

Markus Pueschel  
Department of Computer Science  
ETH Zürich  
pueschel@inf.ethz.ch

#### MS41

##### Performance Engineering of Emerging Memory Systems

Abstract Not Available At Time Of Publication.

Jeffrey S. Vetter  
Oak Ridge National Laboratory  
Georgia Institute of Technology  
vetter@computer.org

#### MS41

##### Designing an Algorithm with a Tuning Knob that Controls its Power Consumption

Consider a parallel algorithm running on a node with an explicit limit on the amount of physical power (e.g., Watts) that it may use. Can you design the algorithm to be fast while still meeting this cap? For the problem of computing single-source shortest paths, we show one such algorithm. We then add a model-driven controller that makes it automatically “power-tunable.” Lastly, we validate the method experimentally on real systems.

Sara Karamati, Jeffrey Young, Richard Vuduc  
Georgia Institute of Technology  
s.karamati@gatech.edu, jyoung9@gatech.edu,  
richie@cc.gatech.edu

#### MS42

##### The Case for Resilience Support in MPI

Over the last years, the topic of MPI resilience evolved mainly around its interaction with the HPC community. The increase in the number of future computational resources will certainly impact raise the need for resilience in large-scale platforms, but meanwhile the potential of resilience in MPI have gained traction in other communities. Outside HPC the usage of MPI remains marginal, because many communities cherish more the current capability of their programming models to seamlessly cope with faults than the potential gain in performance. Thus, if MPI is to grow outside the HPC world and become a mainstream programming support for other parallel programming environments it needs to acquire the capability to detect and deal with faults not by imposing a unique, HPC driven

model, but by empowering the application domain to design its own resilience model. This talk will present an addition to the MPI API that allows users to detect and deal with faults, as well as application-driven techniques to survive faults, and their expected costs at scale, as well as the necessary support from the programming paradigms and their runtimes.

George Bosilca  
University of Tennessee - Knoxville  
bosilca@icl.utk.edu

Aurélien Bouteiller  
University of Tennessee, Knoxville, US  
bouteill@icl.utk.edu

Thomas Herault  
Univ. of Tennessee - Knoxville  
herault@eecs.utk.edu

#### MS42

##### **Evaluating the Fenix MPI Fault Tolerance Programming Framework**

The current implementation of the Fenix specification leverages the User Level Fault Mitigation (ULFM) proposal for extension of the MPI standard. A prototype implementation of the framework proved successful for an important production-level simulation code, but concerns have been raised about overheads incurred by Fenix and ULFM in the absence of failures, as well as of their applicability for bulk-synchronous HPC applications. In this presentation, we address both concerns through the use of the Intel Parallel Research Kernels, and firmly conclude that ULFM adds no measurable overhead to applications, not even at very fine granularity. Moreover, we find that overheads of Fenix are negligible. Finally, we will present some empirical evaluations that Fenix+ULFM are perfectly suitable for tightly coupled HPC applications.

Marc Gamell  
Intel  
marc.gamell.balmana@intel.com

Keita Teranishi  
Sandia National Laboratories  
knteran@sandia.gov

Rob Van Der Wijngaart  
Intel Corporation  
rob.f.van.der.wijngaart@intel.com

Manish Parashar  
Department of Computer Science  
Rutgers University  
parashar@rutgers.edu

#### MS42

##### **Local Rollback of MPI Applications by Combining Application-Level Checkpointing and Message Logging**

Despite being one of the most popular parallel programming models, the MPI standard lacks FT support. The ULFM interface, under discussion in the MPI Forum, proposes resilience features to extend the MPI to adequate it to fault-prone environments, such as future exascale systems. Resilient MPI programs are able to detect and react

to failures without stopping their execution, thus, avoiding re-spawning the entire application. Failures usually affect only to a subset of the resources being used by parallel applications. A global rollback introduces unnecessary overheads and energy consumption: all processes not affected by the failure discard their state and rollback to the last checkpoint to repeat a computation already done. To avoid a global rollback, techniques exploiting the mathematical characteristics of algorithms can improve the recovery, however, they can not be generally applied. We propose a local rollback for generic SPMD programs, in which only the failed processes are recovered from the last checkpoint, while consistency and further progress of the computation is enabled using message logging capabilities. Our proposal combines ULFM, the CPPC application-level checkpointing tool, and the OpenMPI Vprotocol pessimist message logging. Collectives are logged at the application-level, potentially reducing the log size and enabling its use with architecture aware collectives. Additionally, the spatially coordinated protocol used by CPPC helps reducing the log size.

Nuria Losada  
University of A Coruña  
nuria.losada@udc.es

George Bosilca  
University of Tennessee - Knoxville  
bosilca@icl.utk.edu

Aurélien Bouteiller  
University of Tennessee, Knoxville, US  
bouteill@icl.utk.edu

Patricia González  
University of A Coruña, Spain  
pglez@udc.es

Thomas Héroult  
University of Tennessee, Knoxville, US  
herault@icl.utk.edu

María J. Martín  
University of A Coruña, Spain  
mariam@udc.es

#### MS42

##### **GASPI – Scaling in Dynamic Environments**

Owned by the open GASPI forum, GASPI is a small API specification for a programming model that focuses on concurrent, asynchronous and single sided communication in a global address space. GASPI aims for maximal overlap, cheap and fast synchronization by remote notifications and provides basic mechanisms for resilience. GASPI's remote notifications have proven to be an excellent basis to build efficient task based models on top. Challenges like resource consumption, heterogeneity, interoperability with other programming models, interfaces scalability or minimal CPU impact of the communication have been addressed during the specification process. We will sketch the solutions that are proposed by GASPI and which are provided by the high performance reference implementation GPI for many systems and give more details about the resilience approach that includes a healthy state vector for observation of remote ranks, a timeout mechanism to allow to always return to valid system states or a mechanism for a dynamic node set to allow to replace faulty machines at run time. Also we will emphasize the importance of single-

sided and asynchronous communication combined with fine grained notifications for resilience against (system) latency and noise in the presence of dynamically routed multi hop networks. We will conclude that GASPI is a vehicle that enables applications for efficient use of HPC systems today, addresses resilience on different levels and has proven its hardness already.

Mirko Rhan, Franz-Josef Pfreundt  
Fraunhofer Institute for Industrial Mathematics ITWM  
mirko.rahn@itwm.fraunhofer.de, pfreundt@itwm.fraunhofer.de

#### MS43

##### Dense Linear Systems for Extreme Scale

Abstract Not Available At Time Of Publication.

Jack J. Dongarra  
University of Tennessee, Oak Ridge National Laboratory,  
USA  
dongarra@icl.utk.edu

#### MS43

##### The Batched Blas

In this talk we shall motivate and describe the proposed Batched Basic Linear Algebra Subprograms (BBLAS). The BBLAS are intended to independently perform a large number of a specific BLAS operation, such as matrix multiplication, on small matrices. As with the existing BLAS, the aim is to agree a specification for the BBLAS so that code which requires the BBLAS can be portable, but at the same time utilise efficient versions of the BBLAS produced by vendors and other developers. It is hoped that by the time of the SIAM conference, the proposed specification should be nearing completion.

Sven J. Hammarling  
Numerical Algorithms Group Ltd  
Wilkinson House  
sven.hammarling@gmail.com

Mawussi Zounon  
Univeristy of Manchester  
mawussi.zounon@manchester.ac.uk

#### MS43

##### Task Based Robust Computation of Eigenvectors

Dense eigenvalue problems are solved using orthogonal similarity transformations. In the case of the standard eigenvalue problem, one first reduces the matrix to upper Hessenberg form. Then the QR algorithm is applied until the real Schur form emerges. Eigenvectors of the Schur matrix for the user's selection of eigenvalues can now be computed and transformed to the original basis. In this talk, we focus on the problem of computing eigenvectors from matrices in real Schur form. The problem is interesting for the following reasons. If the eigenvalues are clustered, then the calculation is likely to overflow. In LAPACK, overflow is prevented by xLATRS and its descendants. We have recently shown how to protect against overflow in a parallel setting. Moreover, while the calculation of a single eigenvector is memory-bound, the computation of several eigenvectors can be interleaved to increase the overall arithmetic intensity.

Carl Christian Kjelgaard Mikkelsen, Lars Karlsson

Umeå University, Dept. of Computing Science  
spock@cs.umu.se, larsk@cs.umu.se

Mirko Myllykoski  
Umeå University  
mirkom@cs.umu.se

Angelika Schwarz  
Umeå University  
angies@cs.umu.se

#### MS43

##### Parallel Algorithms for Dense Eigenvalue Problems - Progress on Schur Form Computations

In this presentation we discuss recent progress in parallel algorithms for dense nonsymmetric eigenvalue problems with focus on computing standard as well as generalized Schur forms associated with  $Ax = \lambda x$  and  $Ax = \lambda Bx$ , respectively. The main steps include: (i) reduction to condensed form (upper Hessenberg and Hessenberg-triangular forms), (ii) multishift QR and QZ algorithms using aggressive early deflation (AED). We compare our MPI-based parallel algorithms and implementations with new task-based algorithms using the StarPU runtime system, including computational results on state-of-the-art HPC systems.

Bo T. Kågström  
Umeå University, Sweden  
bokg@cs.umu.se

Björn Adlerborn  
Umeå University  
Computing Science and HPC2N  
adler@cs.umu.se

Mahmoud Eljammaly  
Umeå University  
Dept. of Computing Science  
mjamaly@cs.umu.se

Lars Karlsson  
Umeå University, Dept. of Computing Science  
larsk@cs.umu.se

Carl Christian K. Mikkelsen  
Computing Science  
Umeå University  
spock@cs.umu.se

Mirko Myllykoski  
Umeå University  
mirkom@cs.umu.se

Angelika Schwarz  
Umeå University  
angies@cs.umu.se

#### MS44

##### A Landau Collision Integral Solver with Adaptivity on Emerging Architectures with Coupling to PIC Vlasov Methods

The Landau collision integral is an accurate model for the small-angle dominated Coulomb collisions in fusion plasmas. We investigate a high order accurate, fully conserva-

tive, finite element discretization of the nonlinear multi-species Landau integral with adaptive mesh refinement using the PETSc library ([www.mcs.anl.gov/petsc](http://www.mcs.anl.gov/petsc)). We present algorithms and techniques that efficiently utilize emerging architectures by minimizing memory movement and being amenable to vector processing. We present performance results on the Second Generation Intel Xeon Phi, Knights Landing, processor. We discuss issues of coupling this phase space solver with PIC Vlasov methods and the future deployment of this solver in a publicly available library.

Mark Adams  
Lawrence Berkeley Laboratory  
mfadams@lbl.gov

#### MS44

##### Masively Parallel Flow Simulations Using Particles and Grids

Grids and particles are two of the most prominent discretizations schemes for flow simulations. In this talk I will discuss the major advantages and limitations of each in terms of numerical methods and the related software that may exploit massively parallel computer architectures. I will present examples from simulations in fluid dynamics at multiple scales including blood flows and cloud cavitation collapse.

Petros Koumoutsakos  
CSE-lab, Institute for Computational Science, D-MAVT  
ETH Z  
ETH Zurich, Switzerland  
petros@ethz.ch

#### MS44

##### Molecular Dynamics Towards ExaScale Age

Abstract Not Available At Time Of Publication.

Alin Marin Elena  
Scientific Computing Department  
STFC  
alin-marin.elena@stfc.ac.uk

#### MS44

##### Parallel Particle-in-Cell Method for Solving Non-Relativistic Vlasov-Maxwell System

Abstract Not Available At Time Of Publication.

Cui Tao  
LSEC  
CAS  
tcui@lsec.cc.ac.uk

#### MS45

##### Performance Evaluations and Optimizations of H-Matrices for Many-Core Processors

Hierarchical matrices ( $\mathcal{H}$ -matrices) is one of the approximation techniques for dense matrices such as a coefficient matrix arise from the convolution integral term in the fundamental equation of the Boundary Element Method (BEM). We evaluate and optimize a  $\mathcal{H}$ -matrices library  $\mathcal{HACapK}$  for GPUs and Xeon Phi. We achieve 3.2x and 4.4x performance improvement by the series of optimiza-

tions on P100 GPU and Xeon Phi KNL, respectively.

Tetsuya Hoshino, Akihiro Ida  
The University of Tokyo  
hoshino@cc.u-tokyo.ac.jp, ida@cc.u-tokyo.ac.jp

Toshihiro Hanawa  
Information Technology Center  
The University of Tokyo  
hanawa@cc.u-tokyo.ac.jp

#### MS45

##### Accumulator Based Task-Parallel H-Factorization

Arithmetic for hierarchical matrices is often performed on a block-wise level where updates to subblocks of the H-matrix are applied as soon as they are computed and independent on the position of the subblock in the hierarchy. Recently, a modified formulation of the H-matrix arithmetic was introduced, which collects updates for each subblock in an accumulator and where the application of updates strictly follows the hierarchy of the H-matrix. By this, the number of low-rank truncations is significantly reduced, which therefore also reduces the runtime of H-arithmetic. On a parallel system with a task-based formulation of H-arithmetic, this approach permits further variations, e.g., eager and lazy evaluation of updates or the sorting of individual updates, e.g., based on their norms. We will discuss the properties of these forms of H-arithmetic for the algorithmical formulation and the influence on the approximation.

Ronald Kriemann  
Scientific Computing, Max Planck Institute for  
Mathematics i  
rok@mis.mpg.de

#### MS45

##### Low-Rank Matrix Computations on Manycore Systems

We present fast low-rank matrix computations on many-core architectures. By exploiting the low-rank off-diagonal block structure, we design and implement fast linear algebra operations on massively parallel hardware architectures. The main idea is to refactor the numerical algorithms and the corresponding implementations by aggregating similar numerical operations in terms of highly optimized batched kernels. Applications in weather prediction, seismic imaging and material science are employed to assess the trade-off between numerical accuracy and parallel performance of these fast matrix computations compared to more traditional approaches.

Hatem Ltaief  
King Abdullah University of Science & Technology  
(KAUST)  
hatem.Ltaief@kaust.edu.sa

Kadir Akbudak  
Bilkent University  
kadir.cs@gmail.com

Ali M. Charara  
KAUST  
ali.charara@kaust.edu.sa

Aleksandr Mikhalev

King Abdullah University of Science & Technology  
(KAUST)  
aleksandr.mikhalev@kaust.edu.sa

Sameh Abdullah, David E. Keyes  
KAUST  
sameh.abdulah@kaust.edu.sa, david.keyes@kaust.edu.sa

#### MS45

##### Parallel Task Frameworks for FMM

Abstract Not Available At Time Of Publication.

Simon McIntosh-Smith  
University of Bristol  
cssnmis@bristol.ac.uk

#### MS46

##### Underappreciated Bottlenecks for Graph Algorithm Communication

Graph algorithms, like many other data-intensive workloads, are commonly bottlenecked by communication. In this talk, I'll describe communication bottlenecks present in today's hardware platforms as well as a few techniques to mitigate them. To reason about memory communication at the algorithmic and implementation levels, I'll provide a taxonomy of communication inefficiencies and a simple performance model to evaluate tradeoffs. I'll conclude by surveying how blocking has been successfully applied to PageRank.

Scott Beamer  
Lawrence Berkeley National Laboratory  
sbeamer@lbl.gov

#### MS46

##### SlimSell: A Vectorized Graph Representation for Breadth-First Search

We propose SlimSell: a vectorizable graph representation to accelerate Breadth-First Search (BFS) based on sparse-matrix dense-vector (SpMV) products. SlimSell extends and combines the state-of-the-art SIMD-friendly Sell-C<sub>if</sub> matrix storage format with tropical, real, boolean, and sel-max semiring operations. We evaluate SlimSell on Intel Haswell multicore CPUs, the state-of-the-art Intel Xeon Phi KNL manycore CPUs, and NVIDIA Tesla GPUs. Our experiments illustrate that SlimSell can accelerate a tuned Graph500 BFS code by up to 33%.

Torsten Hoefer  
ETH Zürich  
htor@inf.ethz.ch

#### MS46

##### Dynamic Load Balancing of Massively Parallel Graphs for Scientific Computing on Many Core and Gpu Based Systems

Simulating systems with evolving relational structures on massively parallel computers requires maintaining even work distribution across the processing resources. Adaptive, unstructured, mesh-based finite element and finite volume workflows best exemplify this need. To support balancing these workflows across many-core and GPU based systems we present EnGPar's multiple criteria diffusive partition improvement methods. EnGPar's inter-node

performance is compared against its predecessor, ParMA. The performance of intra-node data parallel procedures written in Kokkos are compared to other leading implementations of similar algorithms.

Cameron Smith  
Scientific Computation Research Center  
Rensselaer Polytechnic Institute  
smithc11@rpi.edu

Gerrett Diamond, George M. Slota  
Rensselaer Polytechnic Institute  
diamog@rpi.edu, slotag@rpi.edu

Mark Shepard  
Scientific Computing Research Center  
Rensselaer Polytechnic Institute  
shephard@scorec.rpi.edu

#### MS46

##### Triangle Counting and Truss Decomposition on Modern Parallel Architectures

The scale of data used in graph analytics grows at an unprecedented rate. More than ever, domain experts require scalable parallel algorithms for tasks in graph analytics. Two such related tasks are triangle counting and truss decomposition, which are used extensively to analyze graph structures and discover valuable insights. In this talk, we present recent advances in high-performance algorithms for these tasks on modern parallel architectures. We address challenges including spatial and temporal cache locality, synchronization overheads, and load imbalance.

Ancy Tom, Shaden Smith  
University of Minnesota  
tomxx030@umn.edu, shaden@cs.umn.edu

George Karypis  
University of Minnesota / AHPCRC  
karypis@cs.umn.edu

#### MS47

##### Accelerated HPC Considerations for Modern Fortran

Organizing large codebases with divergent targets is no easy task in Fortran. Code diverges both to allow for different numerical approximations for different scientific applications and to allow for optimizations for a given architecture that harm performance on other architectures. The traditional module-based organization is not a viable option once complexity goes beyond two to three different targets. We need something more sophisticated to provide sufficient flexibility and clarity when a codebase branches significantly. Modern Fortran has a lot to offer with class inheritance and type-bound procedures, but there has always been a strong tension between flexible software engineering practices and fast performance. In this talk we discuss the practical reasons for this tension as well as some real-world examples of how to effectively provide performance while maintaining scientific and architectural flexibility. Topics will include pushing sufficient looping down the call-stack to allow for vectorization for the architectural target at hand, respecting cache bounds for more traditional CPU-based architectures, and understanding how class-based data and calling structures interact with modern directives-based porting techniques like OpenMP and

OpenACC.

Matthew R. Norman

Oak Ridge National Laboratory  
Oak Ridge, TN  
normanmr@ornl.gov

Robert Pincus  
University of Colorado at Boulder  
robert.pincus@colorado.edu

#### MS47

##### **ForTrilinos: Fortran Interface for Trilinos Library**

In this talk, we will describe ForTrilinos, an object-oriented Fortran interface to the Trilinos project. We will present a way to automatically generate such interface using a newly developed Fortran module for the SWIG project. We will describe the available functionality and the path forward for the project including the interface for the next generation architectures.

Andrey Prokopenko, Katherine J. Evans  
Oak Ridge National Laboratory  
prokopenkoav@ornl.gov, evanskj@ornl.gov

Michael Heroux  
Sandia National Laboratories  
maherou@sandia.gov

Seth Johnson  
Oak Ridge National Laboratory  
johnsonsr@ornl.gov

#### MS47

##### **Unstructured-Grid CFD Algorithms on Many-Core Architectures**

In this work, core kernels of an existing large-scale unstructured-grid computational fluid dynamics solver are ported to two nascent many-core architectures. Data layout strategies and disparate programming models are detailed, including explicit domain decomposition with message passing and shared-memory approaches. A gamut of optimization techniques are explored, from compiler directives to assembly-level programming. Their impacts on all levels of performance are presented. Performance comparisons with conventional multi-core architectures are also shown.

Aaron Walden

NASA  
aaron.walden@nasa.gov

Eric Nielsen  
NASA Langley  
eric.j.nielsen@nasa.gov

M Zubair  
Department of Computer Science  
Old Dominion University  
zubair@cs.odu.edu

John Linford  
ParaTools, Inc.  
jlinford@paratools.com

John Wohlbier

HPCMP PETTT, Engility Corp.  
john.wohlbier@engilitycorp.com

Justin Luitjens  
NVIDIA Corp.  
jluitjens@nvidia.com

Izaak Beekman  
ParaTools, Inc.  
ibeekman@paratools.com

Samuel Khuvis  
Paratools, Inc.  
skhuvis@paratools.com

Sameer Shende  
ParaTools, Inc.  
sameer@paratools.com

#### MS47

##### **Enamr: XRage's Code Modernization Strategy**

ENAMR is the methodology LANL is developing to address separation of concerns in the xRage project. We will discuss the abstractions we have developed for a large Fortran code and the implementation of those abstractions, focusing on our experiences and those of the physics developers we have worked with. We will discuss preliminary work we have done behind the abstraction layer to improve performance and present timing results for this code as it exists today. Finally, we will talk about how such abstractions can, in the future, allow us to target multiple computer platforms, and maintain correctness while increasing performance.

Geoff Womeldorff, Joshua Payne, Zach Medin, Sriram Swaminarayan, Galen Shipman  
Los Alamos National Laboratory  
womeld@lanl.gov, payne@lanl.gov, zmedin@lanl.gov, sriram@lanl.gov, gshipman@lanl.gov

#### MS48

##### **Performant Inline NLTE Calculations for Thermal Radiation Transport**

Relaxing the assumption of local thermodynamic equilibrium in thermal radiation transport applications requires inline opacity computations. A system of collisional-radiative equations must be solved explicitly in order to obtain the atomic level populations. This is done for every spatial cell and every timestep of a multi-physics simulation and can take > 90% of the calculation time. Due to high cost, performant solutions were sought, resulting in the use of **Kokkos**.

Daniel A. Holladay  
Los Alamos National Laboratory  
dholladay00@lanl.gov

#### MS48

##### **Kokkos Implementation of Shock Hydrodynamics with Complex Topology Changes and Electric Potential in Alexa**

Alexa is a next-generation shock hydrodynamics application that combines performance portability in the form of an implementation based on Kokkos loops and mesh



adaptation through an interface to the Omega\_h adaptation library. The goals of Alexa are to simulate complex multi-material cases with extreme deformation and topology changes, while at the same time providing as much of this simulation capability as possible when running on one or more GPUs. This talk presents recent developments in Alexa including the implementation of a Kokkos-parallel technique for changing mesh material topology, the integration of a GPU-accelerated linear solver to compute electric potential, and new performance studies on platforms with multiple GPUs with specific consideration of the costs of communication across GPUs using different network technologies.

Daniel Ibanez, Glen Hansen  
Sandia National Laboratories  
daibane@sandia.gov, gahanse@sandia.gov

Thomas Voth, Edward Love  
Sandia National Laboratory  
tevoth@sandia.gov, elove@sandia.gov

James Overfelt  
Sandia National Laboratories  
jroverf@sandia.gov

#### MS48

##### **Intrepid2: a Performance-Portable Package for Compatible High-Order Finite Element Discretizations**

We present Intrepid2, a performance portable C++ toolkit for the local assembling of high-order compatible finite elements (FEs). Intrepid2 is part of the Trilinos suite and it is at the core of several FE libraries. High-order FEs are ideally suited for next generation architectures (high flop to byte ratio) and Intrepid2 achieves performance portability by relying on Kokkos package. Performance results are presented for different manycore architectures when assembling high-order FEs.

Mauro Perego  
CSRI Sandia National Laboratories  
mperego@sandia.gov

Kyungjoo Kim, Nathan Ellingwood  
Sandia National Laboratories  
kyukim@sandia.gov, mailto:ndellin@sandia.gov

Kara Peterson  
Sandia Natl. Labs  
kjpeter@sandia.gov

#### MS48

##### **Performance Portability of Climate Applications Within the Albany Finite Element Code Using Kokkos**

High resolution simulations play a crucial role in the ongoing effort to develop more accurate and reliable regional and global integrated Earth system models. Higher resolutions require more computational power, and, as large high performance computing clusters become more and more heterogeneous, performance portability becomes increasingly more important, since it is needed to take full advantage of next generation computer architectures. By utilizing Kokkos, performance portability is achieved in the Aeras atmospheric dynamical core and the FELIX land-

ice solver within the Albany multi-physics finite element code. This talk focuses on key performance developments in the finite element assembly process and future performance goals in strong and weak scalability on a variety of different architectures including NVIDIA GPUs and Intel Xeon Phis. A perspective towards performance portability of the solver for land-ice simulations will also be provided.

Jerry Watkins, Irina K. Tezaur  
Sandia National Laboratories  
jwatkin@sandia.gov, ikalash@sandia.gov

#### MS49

##### **Implementation and Evaluation of BLAS on PEZY-SC/SC2 Processor**

We have been developing the power-efficient super computer “ZettaScaler” series. We adopted three novel technologies, MIMD ultra-manycore processor called “PEZY-SC” series, high density server board “Brick,” and direct liquid immersion cooling system. They enable us to realize high performance, low power consumption, and space miniaturization at a same time. A 1st generation processor “PEZY-SC” has 1024 MIMD core, and a 2nd generation processor “PEZY-SC2” has 2048 MIMD core. For a high performance numerical simulation, we implement the BLAS on the PEZY-SC/SC2 processor that called “PZBLAS.” We provide an OpenCL-Like programming environment that named “PZCL” in order to use the PEZY-SC/SC2 processor easily. The “PZBLAS” has two interfaces. A 1st interface is “PZC interface” that is same interface as CBLAS interface. A user does not need rewrite program using the CBLAS. However, it needs to communicate the PEZY-SC/SC2 processor from the Intel Xeon processor in each calling function. A 2nd interface is “PZCL interface” that is OpenCL-like one. It can avoid extra-communications in exchange for rewriting program. In this talk, we introduce our systems and performance tuning techniques. And we evaluate these performances, the difference of each interface, and the power-efficiency.

Toshiaki Hishinuma, Ryo Sakamoto, Hitoshi Ishikawa  
PEZY Computing  
hishinuma@pezy.co.jp, sakamoto@pezy.co.jp,  
ishikawa@pezy.co.jp

#### MS49

##### **Performance Analysis of 2.5D-PDGEMM on the K Computer**

In recent years, performance improvements of supercomputers mainly rely on increasing the number of cores and nodes, and it is predicted that those will increase more and more for the Exa-scale era. On highly parallel environments, the performance of a computation could be limited by the cost of collective communications, and therefore, communication avoiding techniques are required. Toward the next generation supercomputers, we have been developing new parallel matrix multiplication (PDGEMM) with the 2.5D algorithm which is a communication avoiding algorithm. In this talk, we will show the performance evaluation result on the K computer with up to 16384 nodes and the performance estimation using a performance model.

Daichi Mukunoki  
RIKEN, Japan  
daichi.mukunoki@riken.jp

Toshiyuki Imamura

RIKEN Advance Institute for Computational Science  
imamura.toshiyuki@riken.jp

#### MS49

##### Performance Evaluation of Application Kernel using ARM SVE

Modern high-performance processors are equipped with a wide SIMD instruction set. SVE (Scalable Vector Extension) is a new ARM SIMD technology that supports vector lengths from 128-bits to 2048-bits. One of its promising features is to offer vector-length agnostic programming to allow the same SVE code to run on hardware of any vector length without any modification of the code. This feature would be useful to explore the best vector length with appropriate hardware resources in the space of various combinations of hardware parameters in order to make more efficient use of hardware resources, since we can use the same vectorized SIMD code. In this presentation, we report the performance of application kernels and numerical libraries using ARM SVE with multiple vector lengths while keeping the hardware resource the same. We have confirmed that when the performance of the program is limited by a bottleneck of a long chain of arithmetic operations or instruction issues, the performance can be improved by increasing the vector length. However, it was necessary to prepare a sufficient number of physical registers for performance improvement, and when the number of physical registers was too small, it was found that with such a program, the performance might be reduced.

Tetsuya Odajima, Yuetsu Kodama, Motohiko Matsuda, Miwako Tsuji, Jinpil Lee, Mitsuhisa Sato  
RIKEN Advanced Institute for Computational Science  
tetsuya.odajima@riken.jp, yuetsu.kodama@riken.jp,  
m-matsuda@riken.jp, miwako.tsuji@riken.jp, jinpil.lee@riken.jp, msato@riken.jp

#### MS49

##### OpenArray: An Auto-parallel Numerical Library on Sunway TaihuLight

This talk presents a numerical library named OpenArray developed for large-scale ocean model simulations on Sunway TaihuLight through providing a set of user friendly API functions similar to Matlab. In order for more programming flexibilities and higher performance, this library is build based on lazy evaluation technique, in which expression trees are constructed dynamically using operator overload mechanism in Fortran and C++, therefore we are able to perform optimizations on the expression tree before evaluation. Further more, we introduce fusion kernel generation mechanism for maximum performance potential through eliminating temporary matrix variables and improvement of operational intensity (FLOPS/byte). Finally, this talk introduce optimization techniques on Sunway TaihuLight used in OpenArray.

Shixun Zhang, Jiuzhou Tang  
National Supercomputer Center in Wuxi  
zhangshixun@mail.nscw.cn,  
zhangshixun@mail.nscw.cn

Xing Huang, Xiaomeng Huang, Qi Wu  
Tsinghua University  
zhangshixun@mail.nscw.cn,

zhang-

shixun@mail.nscw.cn, wuqi9395@126.com

#### MS50

##### Meshfree Simulations of Complex Flows Using General Finite Differences

We develop a moving grid meshfree solver for incompressible flows based on General Finite Differences (GFD). A semi-implicit approximate projection method is directly discretized at grid points with stencil coefficients modified to handle complex boundary conditions. An explicit shift is added to the position of grid points in order to maintain a regular distribution of neighbors such that local GFD moment matrices may be inverted. Through several benchmarks we demonstrate that the grid regularization and associated velocity field update maintains the second order spatial convergence rate, the projection splitting error reduces the convergence rate to first order when time marching under the Courant condition. We also demonstrate the use of this method for problems with moving boundaries and variable grid resolution.

Yaroslav Vasylyv, Alexander Alexeev  
George W. Woodruff School of Mechanical Engineering  
Georgia Institute of Technology  
yvasyliv3@gatech.edu, alexander.alexeev@me.gatech.edu

#### MS50

##### Numerical Simulations of Respiratory Aerodynamics in the Lung of Japanese Quail

The present study implements simulations of flow in an avian lung. Respiratory flow in a Japanese quail was simulated by the Lattice Boltzmann method with the CT-data based Cartesian mesh. To save memories, a coarse graining lattice method and an adaptive meshing method, in which airways are covered by isotropic subdomains were used. GP-GPU was employed to achieve the large scale flow simulation. The proposed method certainly helps better an understanding an avian respiratory system.

Masanori Nakamura  
Nagoya Institute of Technology  
Japan  
nakamura.masanori@nitech.ac.jp

Masahiro Yano  
Saitama University  
m.yano.072@ms.saitama-u.ac.jp

Naoki Tateishi  
Osaka University  
ntakeishi@me.es.osaka-u.ac.jp

Shunichi Ishida, Yohsuke Imai  
Tohoku University  
ishida@pfs.mech.tohoku.ac.jp, yimai@pfs.tohoku.ac.jp

#### MS50

##### Coarse-Grained Simulations of Biomembranes Using Meshless Membrane Model

Various types of coarse-grained models have been developed to simulate lipid membranes. In a micro meter scale, bilayer membranes can be considered as a continuum membrane surface with bending rigidity. For numerical simulations, triangulated mesh is widely used to discretize the surface. In my talk, I will focus on another type of the

method, a particle-based meshless method, that doesn't employ any meshes. In the meshless method, the topological changes and dynamics are more easily treated than the mesh methods. As an example of the application of the meshless methods, I will present membrane shape deformation by banana-shaped protein rods. Membrane-mediated interactions between the rods induce rod assembly coupled with membrane shape changes such as protrusion of membrane tubules and formation of polygonal tubes and polyhedral vesicles.

Hiroshi Noguchi

Institute for Solid State Physics

University of Tokyo, Kashiwa, Chiba 277-8581, Japan

noguchi@issp.u-tokyo.ac.jp

## MS50

### Coarse-Grained Simulations of Eukaryotic Cell in Flow Using Dissipative Particle Dynamics

We will present a new experimentally validated Dissipative Particle Dynamics (DPD) model for simulations of eukaryotic cells in flow. The model was calibrated using micropipette aspiration data, and validated using data from microfluidics experiments. Impact of cell properties as well as different cell components such as nucleus and cytoskeleton on cell deformation in microfluidic device used in experiment will be discussed.

Igor V. Pivkin

USI Lugano

Switzerland

igor.pivkin@usi.ch

## MS51

### Data Processing in Seismic Imaging

Typical set of seismic data are sets of common source records, acquired by the thousands of receivers, so that each receiver record signals from multiple sources. For the geophysical purposes it is convenient to deal with the data in space of common-offset, common azimuth spatial coordinate, thus requires processing of all possible source-receiver pairs independently, which makes the algorithms extremely inefficient. We suggest an algorithm for seismic imaging based on the optimal parallel data flow which significantly reduces computational intensity of the algorithm.

Alexander Pleshkevich

Central Geophysical Expedition

psdm3d@yandex.ru

Vadim Lisitsa

Institute of Petroleum Geology & Geophysics of SB RAS  
Russia

lisitsavv@ipgg.sbras.ru

Dmitry Vishnevsky

Trofimuk Institute of Petroleum Geology and Geophysics

SB RAS, Novosibirsk, Russia

vishnevsky@ipgg.sbras.ru

Vadim Levchenko

Institute of Applied Math RAS

vadimlevchenko@mail.ru

## MS51

### Direct Cluster Solver for 3D Helmholtz Equation

### with Hss-Based Data Compression

We describe a frequency-domain MPI-based solver for numerical simulation of acoustic waves in 3D heterogeneous media. The system of linear algebraic equations, arising as a finite difference approximation of a boundary value problem for the Helmholtz equation, is solved by using  $LDL^t$  factorization with optimization by Nested Dissection re-ordering algorithm, low-rank approximation technique and Hierarchically Semi Separable (HSS) format. The numerical experiments have demonstrated advantages (in memory and performance terms) of the proposed approach in comparison with direct solvers and possibility to solve in acceptable time a system of more than  $10^8$  equations corresponding to a realistic geophysical model.

Sergey Solov'yev, Viktor Kostin

Institute of Petroleum Geology and Geophysics SB RAS

511ssa@mail.ru, kostinvi@ipgg.sbras.ru

Andrey Bakulin, Maxim Dmitriev

EXPEC ARC Saudi Aramco

andrey.bakulin@aramco.com,

maxim.dmitriev@aramco.com

## MS51

### Parallel Multi-Scale Multi-Physics Time-Domain Solver for Seismic Wave Propagation

This paper presents a problem-oriented approach, designed for the numerical simulation of seismic wave propagation in models containing geological formations with complex properties such as anisotropy, attenuation, and small-scale heterogeneities. Each of the named property requires a special treatment that increases the computational complexity of an algorithm in comparison with ideally elastic isotropic media. At the same time, such formations are typically relatively small, filling about 25% of the model, thus the local use of computationally expensive approaches can speed-up the simulation essentially. In this paper we discuss both mathematical and numerical aspects of the hybrid algorithm paying most attention to its parallel implementation. At the same time essential efforts are spent to couple different equations and, hence, different finite-difference stencils to describe properly the different nature of seismic wave propagation in different areas. The main issue in the coupling is to suppress numerical artifacts down to the acceptable level, usually a few tenths of the percent.

Galina Reshetova

Institute of Comp. Math. & Mathematical Geophysics of  
SB RAS

kgv@sscc.nsc.ru

Vladimir A. Tcheverda

Inst. of Petroleum Geology and Geophysics SB RAS

Russia

vova\_chev@mail.ru

Vadim Lisitsa

Institute of Petroleum Geology & Geophysics of SB RAS  
Russia

lisitsavv@ipgg.sbras.ru

Viktor Kostin

Institute of Petroleum Geology and Geophysics SB RAS

kostinvi@ipgg.sbras.ru

## MS52

### Open Hardware: How Open Source Designs Will Drive the Next Generation of HPC Systems

As we approach the end of Moores law modern, complex, HPC systems are increasingly relying upon specialized accelerators in order to deliver continued performance increases for specific computational workloads. Developers of these accelerators, especially in many low volume scientific applications, face a stark choice: spend millions on a commercial license for processors and other IP, or face the significant risk and of developing custom hardware. Rapid prototyping methods need to be explored in order to make the design, verification and programming tools for these new accelerators more accessible to the broader scientific community. To increase access and innovation while reducing cost there has been a consistent march towards open source solutions for each of these components including Facebooks Open Compute Project and Intels OpenHPC effort, as well as a burgeoning community surrounding RISC-V based processors. Looking beyond accelerators that may be tightly integrated with HPC systems we see opportunities for open source hardware to include programmable logic embedded within high performance sensors and detectors for aggressive data reduction or being used in conjunction with FPGA and other reconfigurable computing based platforms. This talk will explore the emerging open source hardware effort as well as showcase new platforms for the rapid generation of future HPC accelerators.

David Donofrio

Lawrence Berkeley National Laboratory  
ddonofrio@lbl.gov

## MS52

### A Partitioned Approach for Highly Integrated Exascale Processors

The challenges to push computing to exascale levels are difficult given desired targets for memory capacity, memory bandwidth, energy efficiency, reliability, and cost. Rethinking in every part of the system architecture is required to realize exascale computing. This talk presents a vision for an architecture that can be used to construct exascale systems. I describe a conceptual Exascale Node Architecture (ENA), which is the computational building block for an exascale supercomputer. The ENA consists of an Exascale Heterogeneous Processor (EHP) coupled with an advanced memory system. The EHP achieves high compute density and high memory density through aggressive use of die-stacking technologies. The processors are partitioned into smaller components for process-technology optimizations and cost reduction. To reduce data movement and enhance energy efficiency, these partitioned components are tightly integrated with one another and with in-package high-bandwidth 3D memory via 2.5D and 3D stacking.

Yasuko Eckert

AMD  
Yasuko.Eckert@amd.com

## MS52

### Architectural Innovations for Applications of the Next Decades

High Performance Computing applications have typically

been used for prediction, for example, of the weather, financial derivative behavior or how a mechanical structure responds to loads. Their approach has traditionally been deductive. Say, starting with the Black-Scholes or Navier-Stokes equations, run compute-intensive calculations on input initial conditions, to then make the needed predictions in finance or fluid dynamics respectively. However, with the reemergence of Machine Learning, fueled by decades of accumulated records and output data from the above applications, scientists, engineers and analysts are starting to also employ the data-intensive inductive approach.

Eng-Lim Goh

HPE  
englim.goh@hpe.com

## MS52

### Disruption in Memory

Memory technology is changing at a rate not seen in four decades. This talk will discuss emerging trends in memory technology, and how they will disrupt the development of scientific and engineering applications. New systems, programming abstractions, and algorithms will all be needed to fully exploit the new memories. The end of Moores Law is looming, and the volume of high-speed memory will continue to decrease, on a per core basis. Memory will likely become explicitly visible with the integration of both high-speed DRAM variants (e.g., HMC and HBM) and slower, but persistent devices (i.e., 3DX). 3DX is already available as high-speed SSDs, and is making out-of-core computation more attractive. But its wear issues, when exposed to the transaction rates expected in memory, will make "write avoiding" algorithms critical for ensuring the successful completion of long duration jobs, and continued usefulness of the systems they run on. Overall, significant improvements will be possible in the performance and efficiency of applications, but achieving them will require innovation and investment in new software.

Bob Lucas

ISI  
Univ. of Southern California  
rflucas@isi.edu

Richard Murphy  
Micron Technology, Inc.  
rcmurphy@micron.com

## MS53

### Communication Throughput in HPC Applications

In order to hide latency and bandwidth limitations, runtimes that support high productivity programming paradigms for HPC and data analytics expose a high level of logical concurrency to programmers. On the other hand, when mapping logical concurrency to physical concurrency (cores, pthreads, processes) the performance and productivity promise of tasking runtimes gets hampered. To address intra-node concurrency challenges, we have developed novel active communication concurrency control techniques which combine throttling and transparent injection parallelization. To address scale-out challenges we have developed novel communication reordering algorithms which combine node local with system wide strategies.

Costin Iancu

Lawrence Berkeley National Laboratory

cciancu@lbl.gov

### MS53

#### Chapel: Task-Based Communication in a Productive Language

Chapel is a modern parallel programming language that's designed to make programming at scale significantly more productive. Chapel makes scalable computing accessible while still supporting performance competitive with MPI+OpenMP.

For context, this talk will start with a brief overview of key Chapel features. The bulk of the talk will describe Chapel's runtime support for communication and tasking, along with related optimizations that permit Chapel to compete with, and occasionally outperform, conventional techniques.

Elliot Ronaghan, Bradford L. Chamberlain  
Cray Inc.  
eronagha@cray.com, bradc@cray.com

### MS53

#### From OpenSHMEM to UCX - Co-design Programming Models and Communication Libraries with Network Interconnects

OpenSHMEM is a communication library and runtime implementing PGAS programming model. The library provides a simple but yet high-performing abstraction for development application using one-sided communication semantics. UCX framework emerged out of a desire to consolidate the efforts in development high-performance communication substrates for research and development PGAS and MPI based programming models. In this talk, we will the history of the development of OpenSHMEM and UCX and their impact on co-design hardware and software features for the next generation of interconnects.

Pavel Shamis  
ARM Research  
Pavel.Shamis@arm.com

### MS53

#### A Communication Layer for Fine Grain Communication and Task Parallelism

We will describe our ongoing efforts on a global (inter-node) task parallelism and a novel communication layer designed for handling fine grain communication arising from task parallelism. We devised a flexible mechanism to offload communication to multiple communication threads and their efficient sleep/wakeup primitive. They attain a high message rate when the communication demand is high while ensuring no wastage of CPU resources when it is low.

Kenjiro Taura, Wataru Endo  
University of Tokyo  
tau@eidoss.ic.i.u-tokyo.ac.jp, wendo@eidoss.ic.i.u-tokyo.ac.jp

### MS54

#### Iterative Refinement in Three Precisions for Fast and Accurate Solution of Ill-Conditioned Sparse Linear Systems

We explain how iterative refinement in three precisions, rather than the traditional two, can be used to produce

accurate solutions to potentially ill conditioned and sparse linear systems of equations. One algorithm uses LU factorization both in the initial solve and in solving for the corrections in the refinement step. A second GMRES-based iterative refinement algorithm (GMRES-IR) makes use of the computed LU factors as preconditioners. Both algorithms are supported by rounding error analysis that provides sufficient conditions for convergence and bounds for the attainable normwise forward error and normwise and componentwise backward errors. Since the initial LU factorization can be carried out at less than the working precision, our algorithms offer a potential factor 2 speedup over solving  $Ax = b$  using LU factorization at the working precision.

Nicholas J. Higham  
School of Mathematics  
The University of Manchester  
nick.higham@manchester.ac.uk

Erin C. Carson  
New York University  
erin.carson@nyu.edu

### MS54

#### Accurate and Verified Numerical Computations with HPC

We discuss problems on the current state of HPC (high-performance computing) in terms of verified numerical computations. First, we pose some problems with HPC concretely, and mention how to improve such problems. Next, we introduce our research project on the Post-K computer. We show several recent results in the project including verified numerical computations for large linear systems, accurate solutions of ill-conditioned linear systems, application to problems in physics, and so forth.

Takeshi Ogita  
Tokyo Woman's Christian University  
ogita@lab.twcu.ac.jp

### MS54

#### Faithfully Rounded Floating-Point Computations

We present a pair arithmetic for the four basic operations and square root. It can be regarded as a simplified, more efficient double-double arithmetic. We prove rigorous error bounds for the computed result depending on the relative rounding error unit  $u$  according to base  $\beta$ , the size of the arithmetic expression, and possibly the condition. Under precisely specified assumptions, the result is proved to be faithfully rounded for up to  $1/\sqrt{\beta u} - 2$  operations. The assumptions are weak enough to apply to many algorithms. For example, our findings cover a number of previously published algorithms to compute faithfully rounded results, among them Horner's scheme, products, sums and dot products, or Euclidean norm. Beyond that, several other problems can be analyzed such as polynomial interpolation, orientation problems, Householder transformations, or the smallest singular value of Hilbert matrices of large size.

Marko Lange  
Waseda University  
m.lange@aoni.waseda.jp

Siegfried M. Rump  
Hamburg University of Technology

rump@tuhh.de

#### MS54

##### Open-Source Software Development That Bridges the Gap Between Numerical Analysis and Material Simulation on Supercomputers

In condensed matter physics, quantum lattice model is widely used to study physical properties. Analysis of the model requires high precision large-scale numerical calculation and improvement of accuracy, stability and calculation speed are desired. In this presentation, we introduce a project through software development and enhancement of its usability on supercomputer, concerning construction of environment to try various mathematical algorithms. As an example of this trial, we also introduce the development of  $K\omega + H\Phi$ .

Kazuyoshi Yoshimi, Mitsuaki Kawamura  
The Institute for Solid State Physics  
The University of Tokyo  
k-yoshimi@issp.u-tokyo.ac.jp,  
mkawamura@issp.u-tokyo.ac.jp

Youhei Yamaji  
Department of Applied Physics  
The University of Tokyo  
yamaji@solis.t.u-tokyo.ac.jp

Tomohiro Sogabe  
Nagoya University  
sogabe@na.nuap.nagoya-u.ac.jp

Takeo Hoshi  
Tottori University  
hoshi@damp.tottori-u.ac.jp

#### MS55

##### Sparse Solvers on Extreme-Scale Systems

We introduce and develop algorithms for the solution of sparse linear systems providing background and an overview for the other talks in this minisymposium. We also consider some aspects of the topic not considered by the other talks. For the direct solution of symmetrically structured systems, we express the computation as a DAG and use dense matrix kernels. We consider both multifrontal and supernodal methods and show how numerical pivoting can complicate the exploitation of parallelism providing the background for the second talk in the minisymposium that addresses this issue. We then discuss in some depth algorithms based on a Markowitz threshold approach for highly unsymmetric systems. The basic algorithms are easy to describe but are very complicated to code efficiently even in serial mode. We show that in spite of the complication they can be designed to enable good levels of parallelism. We discuss the issues involved in obtaining high parallelism for iterative solvers noting that preconditioners that are effective in the serial case may not work well in parallel. This part of the talk serves as an introduction to the third and fourth talks in the minisymposium. Finally we consider hybrid techniques that combine the best features of direct and iterative methods and give added scope for the exploitation of parallelism. This minisymposium presents research done in NLAFET, a Horizon 2020 FET-HPC project funded by the European Union.

Iain Duff  
Science & Technology Facilities Council, UK

and CERFACS, Toulouse, France  
iain.duff@stfc.ac.uk

#### MS55

##### Robust Algebraic Preconditioners for Large Scale Applications

In this talk we discuss a robust algebraic preconditioner for solving sparse linear systems of equations involving symmetric and positive definite matrices. The preconditioner relies on a decomposition into subdomains and the approximation of a Schur complement that involves the factorization of a separator and a low rank correction obtained by solving a generalized eigenvalue problem. The preconditioner can be built and applied in parallel. Numerical results on a set of matrices arising from the discretization by the finite element method of linear elasticity models illustrate the robustness and the efficiency of our preconditioner.

Laura Grigori  
INRIA  
France  
Laura.Grigori@inria.fr

Simplic Donfack  
Inria  
simplic.donfack@inria.fr

Olivier Tissot  
INRIA  
olivier.tissot@inria.fr

#### MS55

##### Task-Based Sparse Direct Solver for Symmetric Indefinite Systems

Many applications in science and engineering require the solution of large sparse linear systems of equations. For solving such problems, direct methods are frequently employed because of their robustness, accuracy and usability as black-box solvers. As modern architectures become more and more complex, with an increasing number of cores per chip, a deeper memory hierarchy and the integration of accelerators such as GPUs, it becomes all the more challenging to exploit the potential performance of such machines for sparse matrix factorization algorithms especially in the context of symmetric indefinite systems. Although significant efforts have gone into positive-definite systems, little progress has been reported in the much harder indefinite case. One major advance for tackling these problems is the design of the APTP (a posteriori threshold pivoting) strategy that has been implemented in the SSIDS solver and proven to be efficient on both multicores and GPUs while giving accurate solutions. In this talk, we present the DAG-based solver SpLDLT that relies on a APTP strategy and uses the StarPU runtime system for implementing its parallel version. We demonstrate the benefits of this approach by showing our ability to exploit heterogeneity in the context of GPU-accelerated multicore systems.

Florent Lopez  
Rutherford Appleton Laboratory  
florent.lopez@enseeiht.fr

Iain Duff  
Science & Technology Facilities Council, UK  
and CERFACS, Toulouse, France



iain.duff@stfc.ac.uk

### MS55

#### Enlarged Conjugate Gradient Method for Reducing Communication

In this talk, we present enlarged Krylov Conjugate Gradient (ECG) method for reducing communications when solving symmetric positive definite linear systems. This method relies on searching the solution in  $t$  directions at each iteration instead of one as in classic Conjugate Gradient. It leads to a faster convergence and thus an important reduction in the number of global communications while increasing arithmetic intensity. Then we propose a method to reduce dynamically the number of search directions constructed at each iteration which is both effective and stable. We present experimental and parallel performance results that show the efficiency of the method.

Olivier Tissot

INRIA

olivier.tissot@inria.fr

Laura Grigori

INRIA

France

Laura.Grigori@inria.fr

### MS56

#### A Set of Building Blocks for Tensor Operations: Transposition, Summation, and Contraction

Tensors naturally appear in a variety of disciplines and applications, including computational chemistry, computational physics, mathematics, and even machine learning. While a range of high-performance software tools exist for computations involving one- and two-dimensional arrays, i.e. vectors and matrices, the availability for tensors is much more limited. Moreover, until recently the contrast between the efficiency attained by matrix and tensor operations was staggering. With this talk we give an overview of a set of high-performance kernels for three common tensor operations: transposition, summation, and contraction. Specifically, we present 1) TTC and HPTT, respectively a compiler and a library for the efficient transposition of tensors of arbitrary size, 2) a code generator for tensor summations, and 3) TCCG, a compiler for tensor transpositions. In all cases, the tools exhibit significant speedups over state-of-the-art solutions. All tools are available for download and use.

Paolo Bientinesi

AICES, RWTH Aachen

pauldj@aices.rwth-aachen.de

Paul Springer

AICES

RWTH Aachen University

paul.springer@rwth-aachen.de

### MS56

#### Analysis of Performance and Portability of Sparse Tensor Decompositions on CPU/MIC/GPU Architectures

Tensors have found utility in a wide range of applications, such as chemometrics, network traffic analysis, neuroscience, and signal processing. Many of these applications

have increasingly large amounts of data to process and require high-performance methods to provide a reasonable turnaround time for analysts. In this work, we consider decomposition of sparse count data using CANDECOMP-PARAFAC alternating Poisson regression (CP-APR) with both multiplicative update and quasi-Newton methods. For these methods to remain effective on modern large core count CPU, Many Integrated Core (MIC), and Graphics Processing Unit (GPU) architectures, it is essential to expose thread- and vector-level parallelism and take into account the memory hierarchy and access patterns for each device to obtain the best possible performance. In this presentation, we will highlight the effect of the sparsity patterns in the input tensor on each architecture and how that relates to the access patterns and effective cache and/or shared memory utilization of each solver method and how to potentially exploit the sparsity structure in performance optimizations.

Christopher J. Forster, Keita Teranishi

Sandia National Laboratories

cjforst@sandia.gov, knteran@sandia.gov

Daniel M. Dunlavy

Sandia National Laboratories

Computer Science and Informatics

dmdunla@sandia.gov

Tamara G. Kolda

Sandia National Laboratories

tgkolda@sandia.gov

### MS56

#### Sparse Tensor Computations on Graphic Processing Units

Sparse tensors appear in many large-scale applications with multidimensional and sparse data. While multidimensional sparse data often need to be processed on manycore processors, attempts to develop highly-optimized GPU-based implementations of sparse tensor operations are rare. The irregular computation patterns and sparsity structures as well as the large memory footprints of sparse tensor operations make such implementations challenging. We leverage the fact that sparse tensor operations share similar computation patterns to propose a unified tensor representation called F-COO. Combined with GPU-specific optimizations, F-COO provides highly-optimized implementations of sparse tensor computations on GPUs. The performance of the proposed unified approach is demonstrated for tensor-based kernels such as the Sparse Matricized Tensor-Times-Khatri-Rao Product (SpMTTKRP) and the Sparse Tensor-Times-Matrix Multiply (SpTTM) that are used in tensor decomposition algorithms.

Maryam M. Mehri Dehnavi, Bangtian Liu

Rutgers University

mmehride@gmail.com, bl512@scarletmail.rutgers.edu

### MS56

#### Tensor Decompositions on Emerging Manycore Hardware with Genten and Kokkos

We describe Genten, a new software implementing canonical polyadic (CP) tensor decompositions of sparse tensors on emerging manycore hardware. The software uses Kokkos to provide a performant implementation of the numerical kernels arising in the alternating least-squares (ALS) approach for computing CP decompositions on a

variety of multicore and manycore hardware, such as multi-core CPUs, Intel Xeon Phi, and Nvidia GPUs. We also describe a simple variation of the traditional coordinate storage format that reorders the traversal of tensor nonzeros in the matricized-tensor-times-Khatri-Rao product (MTTKRP) to reduce atomic-write contention and can lead to significantly improved performance with minimal increase in memory footprint. Performance of the CP-ALS decomposition is measured on several hardware platforms, including Xeon Phi and Nvidia GPUs, and performance of the MTTKRP calculation is compared to other state-of-the-art approaches available in the literature.

Eric Phipps

Sandia National Laboratories  
Optimization and Uncertainty Quantification Department  
etphipp@sandia.gov

Tamara G. Kolda

Sandia National Laboratories  
tgkolda@sandia.gov

## MS57

### Hierarchical Matrix Preconditioner on a Distributed Memory Environment

We present a robust and scalable preconditioner for the solution of large-scale linear systems that arise from the discretization of elliptic PDEs amenable to rank compression, in a distributed memory environment. The preconditioner is based on hierarchical low-rank approximations and the cyclic reduction method. The setup and the apply phases of the preconditioner achieve log-linear complexity in memory footprint and number of operations, and it exhibits good weak and strong scalability at large processor counts. Numerical experiments with linear systems that feature symmetry and nonsymmetry, definiteness and indefiniteness, constant and variable coefficients demonstrate the preconditioner applicability and robustness. Furthermore, it is possible to control the number of iterations via the accuracy threshold of the hierarchical matrix approximations and their arithmetic operations, and the tuning of the admissibility condition parameter. Together, the preconditioner allows for optimization of memory requirements and performance.

Gustavo Chavez

Computational Research Division, Lawrence Berkeley National  
gichavez@lbl.gov

George M Turkiyyah

American University of Beirut and KAUST Technology  
gt02@aub.edu.lb

David E. Keyes

KAUST  
david.keyes@kaust.edu.sa

## MS57

### Dynamic Load Balancing for Construction and Arithmetic of Hierarchical Matrices

Hierarchical matrix (H-matrix) is an approximated form to represent  $N \times N$  correlations of  $N$  objects, which usually requires an  $N \times N$  huge dense matrix. We presents our hybrid MPI/OpenMP implementations of H-matrix generation and H-matrix-vector multiplication with dynamic load balancing. H-matrix generation is done by partitioning a

matrix into submatrices, and then calculating element values of the submatrices. We can apply task parallelism to the latter operation by treating each submatrix as a parallelization unit. However, we cannot achieve a good speedup when assigning a set of tasks to threads statically, because (1) we cannot predict the computational amount of each task precisely and (2) there exist tasks whose relative computational costs to the total are too large. We solved these problems by (1) performing dynamic task assignment based on the hierarchical master-worker method in both of MPI and OpenMP parallelizations, and (2) dividing a task of large estimated cost into subtasks to execute them in parallel using all threads in an MPI process. As for H-matrix-vector multiplication, we keep MPI-level task assignment of the generation, while OpenMP-level task assignment is performed dynamically. We can still expect a good load balance because the assignment is based on the exact size of each submatrix. We will also discuss parallelization of matrix partitioning with dynamic load balancing using task parallel languages.

Tasuku Hiraishi

Kyoto University  
tasuku@media.kyoto-u.ac.jp

## MS57

### Distributed-Memory Block Low-Rank Factorization for Large-Scale Systems and Applications

We consider the use of the Block Low-Rank format (BLR) within the MUMPS solver to solve large sparse real-life problems with multifrontal direct methods. In this talk, we focus on distributed-memory environments and present numerical experiments where MUMPS is used in a parallel, hybrid (MPI+OpenMP) setting on a variety of applications. Our strong scalability analysis of the distributed-memory BLR factorization shows a number of issues due to the use of low-rank compression, among which memory efficiency, high weight of communications, and load unbalance. We analyze these issues and propose some ways to overcome them.

Patrick Amestoy

INP(ENSEEIH)-IRIT  
amestoy@enseeiht.fr

Alfredo Buttari

CNRS-IRIT-Université de Toulouse, France  
alfredo.buttari@enseeiht.fr

Jean-Yves L'Excellent

INRIA-LIP-ENS Lyon  
jean-yves.l.excellent@ens-lyon.fr

Theo Mary

Université de Toulouse, UPS-IRIT  
theo.mary@enseeiht.fr

## MS57

### Highly Optimized $H^2$ -Matrix Implementation on Multiple GPUs

We present high performance implementations of matrix vector multiplication and compression operations for the  $H^2$  variant of hierarchical matrices on GPUs. The  $H^2$  variant exploits, in addition to the usual hierarchical block partitioning, hierarchical bases for the block representations and results in a scheme that requires only  $O(n)$  storage and  $O(n)$  complexity for the mat-vec and compression

kernels. These two operations are at the core of algebraic operations for hierarchical matrices, the mat-vec being a ubiquitous operation in numerical algorithms, and compression/recompression representing a key building block for other algebraic operations, which require periodic recompression during execution. The difficulties in developing efficient GPU algorithms come primarily from the irregular tree data structures that underlie the hierarchical representations, and the key to performance is to recast the computations on flattened trees in ways that allow batched linear algebra operations to be performed. This requires marshaling the irregularly laid out data in a way that allows them to be used by the batched routines. Marshaling operations only involve pointer arithmetic with no data movement and as a result have minimal overhead. Inter-GPU communications is performed with the NCCL library. The high efficiency achieved by our algorithms will be shown for the bandwidth-limited mat-vec and the compute bound compression.

George Turkiyyah  
American University of Beirut  
gt02@aub.edu.lb

Wajih Boukaram, David E. Keyes  
KAUST  
wajihhalim.boukaram@kaust.edu.sa,  
david.keyes@kaust.edu.sa

#### MS58

##### Dynamic Algorithms for Network Analysis at Scale

The efficient computation of the top- $k$  most central nodes of a graph has been widely studied, especially for well-known centrality metrics such as Closeness, Harmonic or Betweenness. However, these studies were mostly focused on static graphs but many real-world networks evolve over time. Algorithms that keep the most central nodes updated in dynamic networks faster than static recomputation did not receive as much attention yet. We present several techniques that allow us to efficiently recompute the top- $k$  nodes with highest Closeness Centrality after the insertion or removal of an edge in a network without requiring asymptotically more memory than the static algorithms. In instances up to millions nodes our dynamic approach outperforms a state-of-the-art static algorithm by a factor of  $10^3$ . We also apply similar strategies to the problem of finding a group of  $k$  nodes with high closeness. Our greedy algorithm with an approximation guarantee is an order of magnitude faster than the previously best algorithm depending on the group size.

Eugenio Angriman  
Karlsruhe Institute of Technology  
eangrima@uni-koeln.de

#### MS58

##### Massive-Scale Streaming Analytics

The continued growth of data sizes and streaming rates presents us with numerous challenges and opportunities. In this talk I will discuss several recent developments in analyzing streaming data. We reason about the validity of streaming analytics on dynamic graphs using a new algorithmic model. As an example, we present Hornet, our new scalable data structure that can easily grow to graphs with hundreds of millions of vertices and can support billions of updates per second. We give several case studies of new streaming analytics for real-world applications based

on dynamic graphs.

David A. Bader  
Georgia Institute of Technology  
bader@cc.gatech.edu

#### MS58

##### Title Not Available At Time Of Publication

Abstract Not Available At Time Of Publication.

Robert S. Schreiber  
HP  
r\_schreiber\_98@yahoo.com

#### MS58

##### An Adaptive Graph Processing Library on GPUs

Graph processing is essential for a wide range of applications such as social networking, machine learning, and computational sciences. Recently, GPUs have been used to accelerate graph processing, yet the irregular data structure and the unpredictable data/control flow make it hard to Deploy graph algorithms on GPUs. By far, lots of runtimes have been developed to provide users with both the simple abstraction and the high performance. However, they often deliver unsatisfactory overall performance compared with hardcoded GPU-specific graph algorithms. We analyze the behavior of these runtimes and find that they are inclined to use the same strategy to explore different graphs in different iterations which may make their solutions suboptimal. In this work, we show that there is no best strategy for all the algorithms and inputs. Our adaptive graph processing runtime system targeting the GPU, which is generated in an offline stage by a machine learning method with a training set of more than 1500 graph datasets, to automatically choose the best strategy in different iterations for different algorithm patterns. We evaluate PlanB on five popular graph algorithms (BFS, CC, BC, SSSP, and PageRank) and show that PlanB is about 2-4x faster than Gunrock and CuSha. and sometimes even faster than the fastest GPU hardcoded algorithms.

Guangming Tan  
Institute of Computing Technology  
Chinese Academy of Sciences  
tgm@ncic.ac.cn

#### MS59

##### From Serial to Parallel: Visual Programming for High Performance Numerical Simulation Development

The scientific computing community is suffering from a lack of good development tool that can handle well the unique problems of coding for high performance computing. It is much more difficult for domain experts to parallelize inherited serial codes written in FORTRAN which are very common in CSE research field. An interactive visual programming approach is proposed for rapid development of large-scale numerical simulations. GUI is designed to facilitate variable and parallel component definition according to the specific programming pattern. The serial numerical subprograms coded in FORTRAN are linked in parallel numerical component to fulfil the computation in a sub-domain. A flowchart editor is designed to facilitate the algorithm design and parallel program construction with assembling parallel computing components. Then the application code are generated automatically. A programming

standard of parallel numerical simulations is implemented in the system to standardize the code which ensures the quality of application codes. The approach has been implemented in HiPro, a programming tool supporting structured and unstructured parallel applications development basing on domain specific framework JASMIN and JAU-MIN. HiPro is developed to ease parallel programming for domain experts. Real applications demonstrate that it is helpful on developing complex numerical applications and enabling increased software productivity greatly.

Li Liao

Institute of Applied Physics and Computational Mathematics  
liliao@iapcm.ac.cn

**MS59**

**Sustainability and Efficiency for Simulation Software in the Exascale Era**

Abstract Not Available At Time Of Publication.

Ulrich J. Ruede

University of Erlangen-Nuremberg  
Department of Computer Science (Simulation)  
ulrich.ruede@fau.de

**MS59**

**Wasatch: A Use Case of High Level Abstractions for Computational Multiphysics**

In an effort to improve parallel computing performance, many emerging supercomputers are built using heterogeneous architectures comprised of both CPUs and GPUs. However, this move towards heterogeneous computing comes at a price, which often falls upon application developers who must learn dedicated programming languages and models for the various processing units. This proves to be an impediment for legacy code as well as scientists and engineers who are not trained in computer programming. In this talk, we present our experience in creating Wasatch - a programming environment for computational multiphysics that shelters application developers from the need to learn new programming languages (e.g. CUDA) and models (e.g. threading, MPI). We discuss Wasatch's multiphysics capabilities which range from simple transport-type partial differential equations to low-Mach, turbulent reacting flows and particle-laden flows. Finally, we demonstrate Wasatch's performance on Mira and Titan for both CPUs and GPUs.

Tony Saad

Institute for Clean and Secure Energy  
Department of Chemical Engineering, University of Utah  
tony.saad@chemeng.utah.edu

James C. Sutherland

Department of Chemical Engineering  
The University of Utah  
james.sutherland@chemeng.utah.edu

**MS59**

**A Pattern-Driven Implicit Parallel Computing Abstraction for Mesh-Based CSE Applications**

Parallel computing is hard, and productive high performance parallel computing is even harder, especially when it comes to CSE domain experts. High level parallel com-

puting abstractions are essential to support wide range of applications, ease programming, and enable high performance. However, current parallel computing abstractions, including general purpose abstractions and domain-specific ones, often fall short in either application range, programmability, or performance optimization opportunities. In this paper, we propose a domain specific high level parallel programming abstraction for mesh-based CSE applications, based on the "algorithmic skeleton" idea. Our abstraction tries to cover a full range of mesh-based CSE applications, yet remains high level enough to enable performance optimization opportunities. Based on our abstraction and two reference implementations, over 50 different real-world CSE applications are developed. These applications scale to thousands of cores and some scale beyond 500,000 cores on Tianhe-2 super computer.

Zeyao Mo

CAEP Software Center  
for High Performance Numerical Simulations  
zeyao\_mo@iapcm.ac.cn

Zhang Yang, Xu Liu

Institute of Applied Physics and Computational Mathematics  
yang\_zhang@iapcm.ac.cn, liu\_xu@iapcm.ac.cn

**MS60**

**High Performance Computation of Radiative Transfer Equation Using the Finite Element Method**

Energy transport via radiative transfer is significant for many fields of science and technology, especially when high temperatures are considered. The modeling of such radiative transfer phenomena has been gaining attention these last decades within diverse scientific fields, such as heat transfer, neutron transport, optical imaging, biomedical optics, astrophysics, radiative transport, etc. In this talk, I will present two different strategies for discretizing the steady-state monochromatic radiative transfer equation. These strategies yield similar linear systems, but with very different sparsity patterns. I will show how this may affect the performance of the underlying linear algebra backend (in our case, PETSc), and how to design efficient preconditioners for these systems. Indeed, most of the solvers available in the literature (direct methods, domain decomposition, multigrid) fail to solve such systems.

Pierre Jolivet

CNRS  
pierre.jolivet@enseeiht.fr

**MS60**

**Domain Decomposition Approach for the Iterative Solution of Strongly Coupled Vibro-Acoustic Problems**

Strongly coupled problems lead to ill conditioned systems due to the very high heterogeneity between the blocks associated to the various physics. Furthermore, when code coupling strategy is implemented, the global matrix is not explicitly formed. In this talk, we compare different strategies based on either block approximate factorization or FETI-like domain decomposition approaches for solving time-harmonic coupled fluid-structure problems in vibro-acoustics. All approaches use exact solution of the local

problem associated with each physics.

François-Xavier Roux  
ONERA

Francois-Xavier.Roux@onera.fr

## MS60

### A Parallel Direct Factorization with Symmetric Pivoting for Nonsymmetric and Indefinite Matrices in a Semi-Conductor Problem

A parallel direct solver: Dissection, whose factorization strategy is based on multi-frontal ordering and block factorization with symmetric pivoting and postponing computation with pivot threshold, is developed. This solver is suitable to solve a mixed finite element equation for the drift-diffusion model that describes interaction among electrostatic potential, electron and hole concentrations, though the stiffness matrix for current and carrier density in a Newton iteration has large condition number due to exponentially nonlinear diffusion coefficient.

Atsushi Suzuki

Osaka University

atsushi.suzuki@cas.cmc.osaka-u.ac.jp

## MS60

### A Parallelized Fill-in Controlled ILU Preconditioner for Multiphysics Fem Applications

In dealing with multiphysics applications with implicit scheme, we often meet with complicated linear systems that contain different types of unknowns and equations inside the coefficient matrix. To construct an efficient preconditioner for such a problem, it is important to exploit physical knowledge of underlying PDEs. Here, we propose a parallel ILU preconditioner, where the fill-ins and the overlaps are controlled based on the knowledge of individual problems.

Takumi Washio

University of Tokyo, Japan

washio@sml.k.u-tokyo.ac.jp

## MS61

### A Geometric Partitioning Method for Distributed Tomographic Reconstruction

With the growing sizes of typical tomographic datasets, the computational requirements for tomographic reconstruction algorithms are rapidly increasing. Parallel and distributed methods for tomographic reconstruction are therefore becoming increasingly common. Here, we present a geometric partitioning method that aims to minimize the necessary inter-processor communication when performing distributed iterative reconstruction. The efficiency and flexibility of the method are demonstrated by parallelizing the SIRT method for various acquisition geometries, including parallel-beam, circular and helical cone-beam geometries.

Jan-Willem Buurlage

Utrecht University

janwillembuurlage@gmail.com

Rob H. Bisseling

Utrecht University

Mathematical Institute

r.h.bisseling@uu.nl

Joost Batenburg

Centrum Wiskunde en Informatica

joost.batenburg@cwi.nl

## MS61

### S-Step Methods in Machine Learning

We build on existing work on communication-avoiding and  $s$ -step Krylov subspace methods by extending those results to machine learning where scalable algorithms are especially important given the enormous amount of data. Iterative methods (like block coordinate descent) are routinely used in machine learning to solve optimization problems. However, these methods require communication at every iteration which becomes a bottleneck when scaling on distributed-memory, parallel machines. We show how to derive an  $s$ -step variant of the block coordinate descent (BCD) method for several linear and nonlinear problems (e.g. least-squares, LASSO, SVM) frequently solved in machine learning. Similar to communication-avoiding Krylov subspace methods, we show that  $s$ -step variants of BCD can reduce the latency cost by a factor of  $s$  at the expense of more flops and bandwidth. We show that, in practice, performing this trade-off can result in large running time speedups. In contrast to other results such as CoCoA and HogWILD! , we reduce the latency cost in BCD methods by a factor of  $s$  without changing the convergence behavior, in exact arithmetic.

Aditya Devarakonda, Kimon Fountoulakis

UC Berkeley

aditya@eecs.berkeley.edu, kfount@berkeley.edu

James W. Demmel

University of California

Division of Computer Science

demmel@berkeley.edu

Michael Mahoney

UC Berkeley

mmahoney@stat.berkeley.edu

## MS61

### Communication-Optimal Loop Nests

Communication (data movement) often dominates runtime and energy costs. We consider the problem of reorganizing loop nests to minimize communication. For loop nests that access array variables subscripted by linear functions of the loop iteration vector, the algebraic relationship between variables and operations leads to communication lower bounds applicable to any reordering of the loop iterations.

Nicholas Knight

New York University

nknight@nyu.edu

## MS61

### A Bridging Model for High Performance Cloud Computing

HPC has been extremely successful by focusing on how best to support software developers that have the technical skills required to design highly parallel algorithms, to optimize data partitioning and load balancing, to minimize communications and synchronization, and to do all of the other things necessary to achieve the highest possible

performance on large-scale dedicated clusters and supercomputers. In HPC, where choices have to be made between high performance and fault tolerance, the tendency has almost always been to focus on the former. Big Data, AI and Cloud Computing have also been extremely successful in recent years by focusing on exactly the opposite community, those software developers who do not have the technical skills, or motivation, required for HPC. In Big Data, AI and Cloud Computing, most of the key design decisions regarding parallelization, data partitioning, load balancing, communications, synchronization, redundancy and fault tolerance are automated, enabling developers to produce applications with a minimum of effort, applications that will typically be run on low-cost pools of commodity cloud resources, that will often be virtualized or use containers. In this talk, I will describe some of the challenges in bringing high performance to big data, AI and cloud computing, and describe some new approaches to this problem. I will describe a new bridging model for general purpose parallel computing with fault tolerance and tail tolerance.

Bill McColl

Huawei Technologies France  
bill.mccoll@huawei.com

## MS62

### UPC++: a PGAS Library for Exascale Computing

UPC++ is a library that implements the Asynchronous PGAS model. We are revising the library under the auspices of the United States Department of Energy's Exascale Computing Project, to meet the needs of exascale applications requiring PGAS support. UPC++ is intended for implementing elaborate distributed data structures where communication is irregular or fine-grained. The UPC++ interfaces for moving non-contiguous data and handling memories with different optimal access methods are composable and closely resemble those used in modern C++. The key abstractions in UPC++ are global pointers, that enable the programmer to express ownership information for improving locality, and asynchronous programming via RPC, also known as function shipping, and futures. Futures enable the programmer to capture data readiness state, which is useful in making scheduling decisions, or to chain together operations to execute asynchronously as high-latency dependencies become satisfied. The UPC++ programmer can expect communication to run at close to hardware speeds. To this end, UPC++ runs atop the GAS-Net communication library and takes advantage of GAS-Net's low-overhead communication as well as access to any special hardware support, e.g. RDMA. I will present the latest results with UPC++ obtained from scalable systems.

Scott B. Baden

Lawrence Berkeley National Laboratory  
baden@lbl.gov

## MS62

### Opportunities for Data-Centric Programming in the PGAS Model: Experience with DASH

In recent years, focus has started to shift from compute-centric to data-centric models of execution. Among the reasons for this trend are the ever-increasing relative costs for accessing data in terms of time and energy, and the resulting need to exploit data access locality. One architectural answer for these requirements lies in heterogeneous

memory architectures that integrate multiple kinds of RAM, such as high-bandwidth and non-volatile memory, in addition to regular DRAM. In this talk we explore options for supporting heterogeneous memory architectures and a data-centric execution model in the context of the PGAS (partitioned global address space) programming style. We provide an overview of our C++ PGAS library DASH, which offers distributed data structures and parallel algorithms, and describe early experiences for supporting heterogeneous memory architectures by means of a memory space abstraction and by providing support for mirror objects.

Karl Fuerlinger

Ludwig-Maximilian University of Munich  
karl.fuerlinger@nm.ifi.lmu.de

## MS62

### The OmpSs Programming Model and Its Runtime Support

The talk will present a vision how multicore architectures are impacting parallel programming practices in the high performance context. Using the OmpSs programming model developed at BSC as a conductor, we will show how we think the challenges new architectures are posing should be addressed. Overall, key ideas of our vision are to provide a clean interface to the programmer that decouples her from the architecture itself and lets her focus on the algorithmic issues, data accesses and dependences. We advocate for intelligent runtimes to take the responsibility of mapping the computations expressed by the programmer to the available resources in a potentially very dynamic environment. We aim at OmpSs being a forerunner for OpenMP, proposing and experimenting with features that we believe should be included in such model aiming not only at HPC but also at all kinds of general purpose computing. Example features we will briefly describe include the asynchronous data flow execution of tasks, the support for heterogeneous devices such as GPUs or systems with big and little cores, the benefits of automatic locality aware scheduling policies and hybrid MPI+OmpSs programming.

Jesus Labarta

Barcelona Supercomputing Center  
jesus@cepba.upc.edu

## MS62

### Trends and Challenges on Programming Models for Extreme-Scale Computing

US, Europe, China and Japan are racing to develop the next generation of supercomputer exascale machines - capable of a million trillion calculations a second by around 2020. To realize exascale systems, there are many challenges and issues including architectures and programming models to exploit billions of parallelism. There are two major different approaches for exascale, manycore-based and accelerator-based. OpenMP is a key to make use of manycore efficiently, and is to be evolved for exploiting large parallelisms and integration with communication layers. While MPI is commonly used as a programming model between nodes for large-scale distributed memory system, new problems arise for many core and new communication models such as PGAS is emerging. In Japan, FLAGSHIP2020 project is underway to develop and deploy the post-K computer as the next flagship system of the K computer around 2020. In this talk, an overview of



projects for post-petascale and exascale computing including FLAGSHIP2020 will be described, and challenges for parallel programming models and languages for exascale computing will be addressed.

Mitsuhisa Sato

RIKEN Advanced Institute for Computational Science  
msato@riken.jp

### MS63

#### **Semi-Analytical Highly-Scalable Preconditioner for Helmholtz Equation**

Abstract Not Available At Time Of Publication.

Mikhail Belonosov

Aramco Overseas  
mikhail.belonosov@aramcooverseas.com

### MS63

#### **High-Performance DTgeo4 Software for 3D Wave Modeling of Seismic Fields**

We present the software for the 3D numerical solution of the system of elastic equations to analyze wave propagation in seismic media with the use of GPGPU or GPGPU cluster. The novelty is the use of LRnLA (Locally Recursive non-Locally Asynchronous) DiamondTorre algorithm, which increases data locality, decreases the number of synchronization events, and brings the problem closer to the compute-bound domain. The dependency graph of the problem is analyzed in 4D time-space, and the optimal traversal rule for the cross stencil for GPGPU implementation is found. It enables to increase computation rate by few orders of magnitude, so that batch acquisition of synthetic seismograms is possible. The algorithm is implemented in the software, that was used in many seismic exploration projects. It features

- non-reflective boundary conditions;
- Gaussian seismic source on the surface, or a point source with a defined moment tensor;
- interface for defining the seismic media model as a set of layers. Elliptic inclusions, parameter gradients, anisotropy and coarse layer boundary are possible;
- various signal acquisition systems.

We present the results of computation for various complex geophysical models along with the performance and scaling efficiency measurements. The  $1500 \times 256 \times 600$  mesh is updated with a  $1.1 \cdot 10^9$  cell update per second rate on one node with one GPGPU.

Anastasia Perepelkina

Keldysh Institute of Applied Mathematics RAS  
mogmi@ya.ru

Anastasia Perepelkina, Vadim Levchenko  
Keldysh Institute of Applied Mathematics  
mogmi@ya.ru, lev@keldysh.ru

Andrey Zakirov  
Kintech Lab Ltd.  
zakirov@kintechlab.com

Anton Ivanov  
Keldysh Institute of Applied Mathematics  
aiv.racs@gmail.com

Tatiana Levchenko, Vladimir Rok

VNIGNI

tatlevchenko@mail.ru, vladimir.rok@gmail.com

### MS63

#### **Highly-Scalable Algorithm for Numerical Simulation of Sonic Logs**

Modern high performance computers allow to study complete wave fields which of acoustic logging for 3D heterogeneous media with realistic properties such as anisotropy and attenuation. Anisotropy is rather widespread property of reservoirs as it is caused by thinlayering or systems of oriented fractures, while attenuation may indicate fluid saturation. We present a new approach to finite difference simulation of sonic log for anisotropic viscoelastic media on the base of Lebedev and staggered grid schemes in cylindrical coordinates.

Galina Reshetova

Institute of Comp. Math. & Mathematical Geophysics of SB RAS  
kgv@sscc.nsc.ru

Vladimir A. Tcheverda

Inst. of Petroleum Geology and Geophysics SB RAS  
Russia  
vova\_chev@mail.ru

Vadim Lisitsa

Institute of Petroleum Geology & Geophysics of SB RAS  
Russia  
lisitsavv@ipgg.sbras.ru

### MS64

#### **Quantum Machine Learning**

Abstract: Quantum computing promises an unprecedented ability to solve intractable problems by harnessing quantum mechanical effects such as tunneling, superposition, and entanglement. The Quantum Artificial Intelligence Laboratory (QuAIL) at NASA Ames Research Center is the space agency's primary facility for conducting research and development in quantum information sciences. QuAIL conducts fundamental research in quantum physics but also explores how best to exploit and apply this disruptive technology to enable NASA missions in aeronautics, Earth and space sciences, and space exploration. At the same time, machine learning has become a major focus in computer science and captured the imagination of the public as a panacea to myriad big data problems. In this talk, we will discuss how classical machine learning can take advantage of quantum computing to significantly improve its effectiveness. Although we illustrate this concept on a quantum annealer, other quantum platforms could be used as well. If explored fully and implemented efficiently, quantum machine learning could greatly accelerate a wide range of tasks leading to new technologies and discoveries that will significantly change the way we solve real-world problems.

Rupak Biswas

NASA

Rupak.Biswas@nasa.gov

### MS64

#### **Quantum Supremacy in Near-Term Devices**

A critical question for the field of quantum computing in

the near future is whether quantum devices without error correction can perform a well-defined computational task beyond the capabilities of state-of-the-art classical computers, achieving so-called quantum supremacy. We argue that quantum supremacy can be achieved in the near-term with approximately fifty superconducting qubits for a modest depth quantum computation: sampling the output of a universal quantum circuit selected at random. We show how to use the cross-entropy with supercomputer simulations to compare the experimental error (fidelity) with predictions, consistent with the exponential growth of the computational (Hilbert) space. We will outline the implications for future experiments demonstrating a practical quantum advantage.

Sergio Boixo  
Google  
boixo@google.com

#### MS64

##### Computing Molecular Energies on a Quantum Processor

Harnessing the full power of nascent quantum processors requires the efficient management of a limited number of quantum bits with finite lifetime. Hybrid algorithms leveraging classical resources have demonstrated promising initial results in the efficient calculation of Hamiltonian ground states—an important eigenvalue problem in the physical sciences that is often classically intractable. In these protocols, a Hamiltonian is parsed and evaluated term-wise with a shallow quantum circuit, and the resulting energy minimized using classical resources. This reduces the number of consecutive logical operations that must be performed on the quantum hardware before the onset of decoherence. We demonstrate a complete implementation of the Variational Quantum Eigensolver (VQE), augmented with a novel Quantum Subspace Expansion, to calculate the complete energy spectrum of the  $H_2$  molecule with near chemical accuracy. The QSE also enables the mitigation of incoherent errors, potentially allowing the implementation of larger-scale algorithms without complex quantum error correction techniques.

Irfan Siddiqi  
UC Berkeley  
irfan\_siddiqi@berkeley.edu

Jonathan Carter  
Lawrence Berkeley National Laboratory  
jtcarter@lbl.gov

#### MS64

##### Quantum Computing by Quantum Annealing

Quantum annealing is a generic platform to solve combinatorial optimization and sampling problems using quantum mechanical effects. An enhanced form of quantum annealing is known to be theoretically equivalent to the conventional circuit model of quantum computing. D-Wave Systems in Canada has built hardware to realize quantum annealing with more than 2000 qubits and has sold several of their machines to Google, NASA, Los Alamos National Laboratory, and Lockheed-Martin, which raised a good amount of interest not just from scientific communities but also from a wider public. In this talk, I will explain the basic theoretical framework of quantum annealing, overview the current status of hardware and theoretical developments, and discuss its impact on the society

in general. In particular, I will describe our recent contribution for an exponential speedup of quantum annealing.

Hidetoshi Nishimori  
Tokyo Institute of Technology  
nishimori@phys.titech.ac.jp

#### MS65

##### A Proposal of Monitoring and Forecasting Method for Crustal Activity in and Around Japan with 3-Dimensional Heterogeneous Medium Using a Large-Scale High-Fidelity Finite Element Simulation

We propose a system for monitoring and forecasting of crustal activity, such as spatiotemporal variation in slip velocity on the plate interface including earthquakes, seismic wave propagation, and crustal deformation. It is necessary to develop a physics-based data analysis system including (1) a structural model with the 3D geometry of the plate interface and the material property such as elasticity and viscosity, (2) calculation code for crustal deformation and seismic wave propagation using (1), (3) inverse analysis or data assimilation code both for structure and fault slip using (1) and (2). To accomplish this, Ichimura et al. (2015) has developed unstructured FE non-linear seismic wave simulation code, and Ichimura et al. (2013) has developed high fidelity FEM simulation code with mesh generator to calculate crustal deformation in and around Japan with complicated surface topography and subducting plate geometry. Fujita et al. (2016) has improved the code for crustal deformation and achieved the high-resolution analysis which enables computation of change of stress acting on the plate interface. Further, for inverse analyses, Agata et al. (2015) has improved FEM code with an adjoint method for estimating fault slip and asthenosphere viscosity. Furthermore, Hori et al. (2014) developed a method for forecasting the slip velocity variation on the plate interface, introducing ensemble based data assimilation method. We are integrating them for monitoring and forecasting.

Takane Hori  
Japan Agency for Marine-Earth Science and Technology  
horit@jamstec.go.jp

Tsuyoshi Ichimura, Kohei Fujita  
Earthquake Research Institute  
The University of Tokyo  
ichimura@eri.u-tokyo.ac.jp, fujita@eri.u-tokyo.ac.jp

Ryoichiro Agata  
Japan Agency for Marine-Earth Science and Technology  
agatar@jamstec.go.jp

Takuma Yamaguchi  
Earthquake Research Institute  
The University of Tokyo  
yamaguchi@eri.u-tokyo.ac.jp

#### MS65

##### Scalable Solvers and Preconditioners for Nonlinear Stokes Equations with Yielding

Nonlinear Stokes equations involving a yield stress play an important role in earth science simulations, e.g., in plate tectonics. Application of a Newton method to the resulting saddle point problem typically requires a large number of iterations due to the non-linearity of the constitutive

relation. We present a non-linearly preconditioned Newton method, which involves an independent stress variable. Our approach is matrix-free and only requires minor modifications to the ordinary Newton method. It uses a BFBT-type Schur complement preconditioner, which builds on matrix-free geometric multigrid solvers. We show efficiency and extreme scalability of the method for earth science problems on adaptively refined meshes and for higher-order finite element approximations.

Johann Rudi  
Institute for Computational Engineering and Sciences  
The University of Texas at Austin  
johann@ices.utexas.edu

Georg Stadler  
Courant Institute for Mathematical Sciences  
New York University  
stadler@cims.nyu.edu

Omar Ghattas  
The University of Texas at Austin  
omar@ices.utexas.edu

#### MS65

##### **Towards Exascale Seismology: Challenges and Opportunities**

Advances in numerical methods combined with developments in high-performance computing have enabled unprecedented simulations of seismic wave propagation. In a complementary development, adjoint-state methods efficiently incorporate the full nonlinearity of 3D wave propagation in iterative seismic inversions. In earthquake seismology, seismic imaging based on spectral-element and adjoint methods has enabled assimilation of this information for the construction of 3D (an)elastic Earth models. These methods account for the physics of wave excitation and propagation by numerically solving the anelastic, anisotropic equations of motion, and require the execution of complex computational procedures that challenge the most advanced high-performance computing systems. Current research is petascale; future research will require exascale capabilities. There are successful applications of Full-Waveform Inversion (FWI) both on regional and continental scales, and recently the first global model based on FWI has been produced. In this presentation I will discuss our latest results.

Jeroen Tromp  
Princeton University  
Department of Geosciences  
jtromp@princeton.edu

#### MS65

##### **Simulation Software for the Simulation of the Sumatra 2004 Earthquake and Tsunami**

The 2004 Sumatra-Andaman earthquake, with a moment magnitude of 9.1, triggered devastating tsunamis in the Indian Ocean. The conditions under which earthquakes may trigger large tsunamis, however, are not well understood. A potential way to gain insight into the underlying physics is to couple state-of-the-art earthquake and tsunami simulation software. To this end, we present advances from the project ASCETE. Using SeisSol we can simulate rupture dynamics and complex fault networks at extreme scales, using thousands of compute nodes. The surface displacement computed by SeisSol may be used as

input for tsunami simulation in the adaptive mesh refinement framework sam(oa)<sup>2</sup>.

Carsten Uphoff  
Technische Universität München  
uphoff@in.tum.de

Leonhard Rannabauer, Michael Bader  
Technical University of Munich  
rannabau@in.tum.de, bader@in.tum.de

Alice A. Gabriel  
Department of Earth and Environmental Sciences  
Ludwig-Maximilians-University Munich  
gabriel@geophysik.uni-muenchen.de

#### MS66

##### **Parallel Accurate Algorithms for Polynomial and Rational Function Evaluation**

The Compensated Horner Scheme is an accurate and fast algorithm to evaluate univariate polynomials in floating-point arithmetic. The accuracy of the computed result is similar to the one given by the Horner scheme computed with twice the working precision. We will use this algorithm to accurately evaluate rational functions. We will also present a parallel version of the Compensated Horner Scheme. Some experiments on multicore and Graphics Processor Units (GPU) architectures will be presented to show the efficiency of this algorithm.

Stef Graillat  
University Pierre and Marie Curie (UPMC)  
LIP6  
stef.graillat@lip6.fr

#### MS66

##### **A Reproducible Solution of Linear Systems**

Solving a linear system is an important underlying part of numerous scientific applications. In this talk, we address the issue of non-reproducible solution of linear systems. Our approach is based on the hierarchical and modular structure of linear algebra algorithms. Thus, we divide computations into smaller logical blocks – such as a blocked LU factorization and triangular system solve – and ensure their reproducible results. Finally, these reproducible building blocks lead, as we prove, to the reproducible solution of linear systems. This is a joint work with Erwin Laure (KTH), Stef Graillat (Paris 6), and Enrique S. Quintana-Orti (UJI).

Roman Iakymchuk  
KTH Royal Institute of Technology  
CSC, CST/PDC  
riakymch@kth.se

#### MS66

##### **Verified Numerical Computations for Large-Scale Linear Systems**

This talk concerns verified numerical computations for large-scale linear systems. If a problem size is large, then numerical results may not be accurate due to the accumulation of rounding errors. Our aim is to produce an accurate approximation and its tight error bound for a linear system, where the order of the coefficient matrix is over one million. We will introduce verification theory, implemen-

tation, and numerical examples using the K computer at Riken AICS.

Katsuhisa Ozaki  
Shibaura Institute of Technology / JST  
ozaki@sic.shibaura-it.ac.jp

Takeshi Ogita  
Tokyo Woman's Christian University  
ogita@lab.twcu.ac.jp

#### MS66

##### **Eddy Current Analysis by Iterative Domain Decomposition Method Using Pseudo-Quadruple Precision**

This study purpose is an improvement of convergence properties of an interface problem in the eddy current analysis based on the iterative domain decomposition method. A governing equation based on the A-f method for the time-harmonic eddy current analysis is solved taking the magnetic vector potential  $A$  and the electric scalar potential  $f$  as unknown functions. However, when a numerical model become huge, the convergence properties of the interface problem get worse. To overcome this difficulty, we introduce a pseudo-quadruple precision into iterative solvers in the interface problem and the subdomain problem. Convergence properties in the interface problem are improved by applying the pseudo-quadruple precision to both solvers for the interface problem and the subdomain problem.

Amane Takei  
Faculty of Engineering  
University of Miyazaki  
takei@cc.miyazaki-u.ac.jp

#### MS67

##### **A Stencil Penalty Approach For Improving Accuracy of Constraint Immersed Boundary Method**

The constraint based immersed boundary (cIB) method has been shown to be accurate between low and moderate Reynolds number (Re) flows. In our studies of flow around immersed bodies with very thin or zero-thickness walls, we found that cIB is not able to produce accurate results. The thin IB interfaces typically result in a large pressure gradient across the interface. As a consequence of the pressure jump, the pressure gradient is incorrectly evaluated near the fluid-IB interface. This leads to inaccuracies in the boundary layer around the IB and, in severe cases, leakage of flow across the fluid-IB interface. A numerical formulation that avoids the jump in pressure through a modified pressure gradient operator will be presented. The pressure gradient operator is modified through a WENO based stencil penalization. Application of the proposed method to unsteady vehicle aerodynamics simulations will be discussed.

Rahul Bale  
RIKEN AICS  
rahul.bale@riken.jp

Niclas Jansson, Niclas Jansson  
KTH Royal Institute of Technology  
njansson@kth.se, njansson@kth.se

Keiji Onishi  
RIKEN AICS  
keiji.onishi@riken.jp

Makoto Tsubokura  
RIKEN AICS  
Kobe University  
tsubo@tiger.kobe-u.ac.jp

Neelesh Patankar  
Northwestern University  
neeshpatankar@gmail.com

#### MS67

##### **Efficient Gather Scatter Operations in Nek5000 Using PGAS**

Gather-scatter operations are one of the most important communication kernels in Nek5000 for fetching data dependencies (gather), and spreading back results (scatter). The current implementation in Nek5000 is based on the Gather-Scatter library, GS, which utilizes three different strategies: pairwise nearest neighbour communication; the Crystal router, which aggregates smaller messages and route them around the network in a hypercube like pattern; and finally collective all-to-all communication. Implemented using non-blocking two-sided message passing, the library has proven to scale well to hundreds of thousands of cores. However, the necessity to match sending and receiving messages in the two sided communication abstraction can quickly increase latency and synchronisation costs for very fine grained parallelism, in particular for the unstructured communication patterns created by unstructured problems. We present our work on developing ExaGS, a reimplement of the Gather-Scatter library, with the intent to use the best suitable programming model for a given architecture. A first implementation of ExaGS using the one-sided programming model provided by the Partitioned Global Address Space (PGAS) is presented, together with a performance analysis when used together with Nek5000 or its co-design application Nekbone.

Niclas Jansson, Niclas Jansson  
KTH Royal Institute of Technology  
njansson@kth.se, njansson@kth.se

#### MS67

##### **A Parallel Algorithm of Pre and Post Processing for Industrial CFD Applications**

Pre- and post- processing is still a major problem in industrial Computational Fluid Dynamics (CFD) analysis. With the rapid development of computers, physical solvers are getting faster, while pre- still remains slow because it is mainly a serial process and it is less likely of benefit from HPC technology. Then a methodology using MPI+OpenMP hybrid parallelization has been adopted to eliminate the manual work which required for pre- processing, correcting the surface imperfections of CAD data. And speed-up of visualization is insufficient compared to the rapidly increasing amount of data in recent years. Then the in-situ visualization has been adopted to parallelize the post-processing using libsim (VisIt) library on K-computer. The performance of pre-/post- processing is investigated in this research and showed that the pre-processing time has been reduced to minutes order, during the current process requires several days. The post-processing time has been reduced seconds order per frame, which typically requires several hours of human labor for a data size of 2GB/frame on a vehicle aerodynamics case. This results in an approx. 30% increase in computational time. We believe these are indispensable technologies towards the EXA FLOPS CFD.

Keiji Onishi  
RIKEN AICS  
keiji.onishi@riken.jp

Niclas Jansson, Niclas Jansson  
KTH Royal Institute of Technology  
njansson@kth.se, njansson@kth.se

Rahul Bale  
RIKEN AICS  
rahul.bale@riken.jp

Makoto Tsubokura  
RIKEN AICS  
Kobe University  
tsubo@tiger.kobe-u.ac.jp

### MS67

#### Adaptive Meshes and Error Control in Nek5000

When performing computational fluid dynamics (CFD) simulations, the expected flow physics is often unknown. For this reason, the generation of a high-quality mesh is a complex task and is usually considered as a bottle-neck for any user of a CFD code. Ideally, the mesh should avoid unnecessary over-resolution in unimportant regions and require as few computational resources as possible. Adaptive mesh refinement (AMR) techniques have been developed to tackle this issue by, first, identifying poorly resolved regions with effective error estimators and then, applying a mesh refinement strategy. In the present work, we propose to implement non-conforming mesh refinement within the code Nek5000: an open-source, highly scalable and portable code based on the spectral element method (SEM). We apply our techniques first to 2D flow configurations together with spectral and adjoint error estimators. In addition, we present an effective preconditioning using the Hypr library.

Philipp Schlatter  
KTH Stockholm  
pschlatt@mech.kth.se

### MS68

#### Parallel Unstructured Mesh (Re)partitioning Based on Sfc

Mesh partitioning is traditionally addressed by means of graph partitioning, which is a well studied NP-complete problem generally treated by means of multilevel heuristics composed of three phases: coarsening, partitioning, and un-coarsening. Different variants of them have been implemented in publicly available libraries including parallel versions. However limitations on the parallel performance and the quality of the solution at growing the number of parallel processes are present. This lack of scalability makes graph-based partitioning a potential bottleneck for large scale simulations. Motivated by these circumstances, we have developed a fully parallel geometric approach, based on the utilization of distributed and anisotropic space filling curves, applicable to any kind of mesh. On its implementation we have avoided any computing or memory bottleneck that could limit the scalability of the algorithm, imposing that the solution achieved is independent (discounting rounding off errors) of the number of parallel processes used to compute it. In this talk, we will show both the quality of the resulting partitions and the parallel performance of

the partitioning process for different applications. In particular, we will focus on the utilization of this approach for runtime load balancing purposes.

Ricard Borrell  
Barcelona Supercomputing Center, Spain  
ricard.borrell@bsc.es

Guillaume Houzeaux  
BSC  
guillaume.houzeaux@bsc.es

Juan Carlos Cajas  
Barcelona Supercomputing Centre  
juan.cajas@bsc.es

### MS68

#### Nonlocal Algorithms on Adaptive Meshes

Most algorithms for adaptive mesh refinement (AMR) have been developed to support the numerical solution of partial differential equations (PDEs). Apart from refining and coarsening selected mesh elements, the parallel distribution of the elements and a load-balanced repartitioning are key capabilities required of any AMR implementation. Over the past decade, several approaches have been developed and demonstrated that offer such functionality in-core, at large scale, and with fast absolute run times. In this presentation, after a brief introduction of basic AMR operations, their technical intricacies, and the highly scalable software library `p4est`, we consider algorithms that address problems outside of the minimal set needed for solving PDEs. In particular, we focus on nonlocal relations in the mesh, that is, relations that reach beyond the usual one-element layer of ghost elements. Our goal is to solve global problems with the performance of local algorithms. One example is an efficient generic search that can be used in various ways, for example to locate the sources and receivers for earthquake simulations in the mesh or to compute the departure points of a semi-Lagrangian method. Other examples relate to efficient parallel I/O and post-processing. We close with illustrations from practical applications.

Carsten Burstedde  
Universität Bonn  
burstedde@ins.uni-bonn.de

### MS68

#### Scalable Algorithms for Tree-Based AMR with General Element Types

When working with adaptive meshes that are refined and coarsened frequently during a single simulation, partitioning according to space-filling curves (SFC) has been established as a fast and reliable method. So far, SFC have been used predominantly for hexahedral meshes. In our work, we have developed a SFC for adaptive simplicial meshes that can be computed using bitwise interleaving operations similar to the Morton curve for cubes. Using this SFC we have established constant-time algorithms to compute the parent, children and face-neighbors of a given simplex, as well as predecessor and successor elements with respect to the SFC. Presently we are working on the AMR library `t8code` using a tree-based SFC approach, implementing parallel algorithms for the optimized handling of the coarse mesh of trees, adaptation, partitioning, and construction of a ghost layer. In `t8code` we allow to change and mix the underlying SFC and element implementations, thus sup-

porting hybrid meshes in 2D (triangles and quadrilaterals) and 3D (tetrahedra, hexahedra, and prisms). We present our latest results, demonstrating parallel scalability with several hundred thousand MPI ranks.

Carsten Burstedde  
Universität Bonn  
burstedde@ins.uni-bonn.de

Johannes Holke  
University of Bonn  
holke@ins.uni-bonn.de

## MS68

### Structured and Unstructured Adaptivity in PETSc

The PETSc library is developing an interface for fast, scalable solution methods integrating adaptive mesh refinement (AMR), structure-preserving discretizations, and parallel, multilevel algebraic solvers. The goal of this work is to entirely separate the concerns of mesh adaptivity from the numerical methods of simulations, such as residual/Jacobian construction and solution methods. The first implementation of a dynamic hierarchical AMR interface in PETSc is a quadtree/octree-based method as implemented by the p4est library. An implementation of unstructured AMR is provided by Pragmatic, a parallel library for unstructured anisotropic mesh adaptivity based upon metric tensor fields. A unified interface specifies adaptation for all approaches, while multilevel solvers can access mesh data directly.

Tobin Isaac  
Georgia Tech  
tisaac@cc.gatech.edu

Matthew G. Knepley  
Rice University  
knepley@gmail.com

## MS69

### $O(N)$ Method for Spatiotemporal Boundary Integral Equation Method

A large issue is the numerical cost in the boundary integral equation method (BIEM). The total-binding property of the boundaries yields the onerous dense interaction matrix of BIEM, which is the cost of its accuracy to resolve the stress singularities of fractures and complicated geometries of boundaries and cracks. By hierarchical low-rank approximations, as the Fast Multipole Method [Rokhlin, V., 1985] and the Hierarchical Matrix Method (H-matrix method) [Hackbusch, W., 1999], space-domain BIEM achieved the linear complexity: the best numeric cost  $O(N)$  in  $N$ -element problems. However, the cost generally has been restricted to  $O(N^2)$  or  $O(NM)$  in the spatiotemporal domain BIEM (ST-BIEM), where  $M$  denotes the number of time steps. The inhibitors are the singular wavefronts, which cannot be approximated by the above expansions around singularities. We use ST-BIEM to discuss the wave propagations as earthquakes, so that we need further improvement in the methodology. In this talk, we propose and demonstrate an algorithm to overcome such difficulty: FDP=H-matrix method. The method is based on H-matrix method and the Fast Domain Partitioning Method (FDPM), which regularizes the singular wavefronts along the ray [Ando, R., et al. 2007, Ando, R., 2016]. Further, introducing an asymptotic singularity relocation technique and an additional quantization method to com-

press the sparse tensors, we finally derive the algorithm of the desired cost  $O(N)$  for ST-BIEM.

Daisuke Sato  
The University of Tokyo  
dice@eps.s.u-tokyo.ac.jp

## MS69

### Fast Numerical Methods for Space-Time Fractional Partial Differential Equations

Abstract Not Available At Time Of Publication.

Hong Wang  
University of South Carolina  
Department of Mathematics  
hwang@math.sc.edu

## MS69

### Fast Iterative Solvers for Fractional Partial Differential Equations

We study a multigrid method for linear systems arising from two-dimensional space-fractional diffusion equations. In the discretization matrix, the information of fractional derivative along one spatial variable is distributed in the block diagonal part, while that along another spatial variable is evenly distributed in the whole matrix. To capture the information of fractional derivative along both directions, we propose to combine two alternative-direction smoothers for the multigrid method. One smoother can be block Jacobi iteration, banded splitting iteration, etc. The other smoother is the same smoothing iteration but acting on a permuted linear system in which the previously evenly distributed fractional-derivative information is gathered in block diagonal part. The extension of the alternative-direction smoothers onto high-dimensional cases, non-rectangular domains, are discussed. Numerical experiments are tested to show the performance of the proposed multigrid method and the corresponding multilevel preconditioner.

Michael K. Ng  
Hong Kong Baptist University  
Department of Mathematics  
mng@math.hkbu.edu.hk

Lin Xuelei  
Hong Kong Baptist University  
17481228@life.hkbu.edu.hk

## MS69

### Application of Hierarchical Matrices to Adaptive Finite Element Method for the Multi-Term Fractional Differential Equations

We develop a fast solver for the fractional differential equations (FDEs) involving multi-term Riesz/Riemann-Liouville fractional derivatives. It is based on the use of hierarchical matrices for the representation of the stiffness matrix resulting from the finite element discretization of the multi-term FDE and employs a geometric multigrid method for the solution of the algebraic system of equations. The combination of hierarchical matrices and multigrid leads to linear computational complexity. We also propose an adaptive algorithm based on a posteriori error estimation for multi-term fractional derivatives to deal with general-type singularities arising in the solution of



the FDE. The proposed method resolves boundary singularities at linear complexity while maintaining full second-order accuracy. Several numerical examples are considered aiming to demonstrate the validity and applicability of the proposed techniques.

Xuan Zhao

Department of Mathematics, Southeast University  
xuanzhao11@seu.edu.cn

## MS70

### Role of Framework Development in Multiphysics High Performance Computing Software

Improvements in computational capabilities has led to rising complexity in models, and therefore software implementing them. Frameworks play a crucial role in Multiphysics high performance computing software by providing a substrate for capabilities to grow. Therefore, the Framework lifecycle has different needs than software as a whole. We will present a lifecycle model for Multiphysics software with focus on the role of framework development within the lifecycle.

Anshu Dubey

Argonne National Laboratory  
adubey@anl.gov

Lois Curfman McInnes

Mathematics and Computer Science Division  
Argonne National Laboratory  
curfman@mcs.anl.gov

## MS70

### An Overview of Progress and Challenges for Scalable Manycore and Accelerator Systems

In this presentation we give an overview of some of the challenges and approaches currently facing the computational science community as it conducts research and development of algorithms and software for current and next-generation scalable computing systems. We focus in particular on architecture changes that provide the biggest challenges and opportunities, general strategies the community is using to address these challenges. We also focus on the important role that improve software practices can play as we produce software for these systems.

Michael Heroux

Sandia National Laboratories  
maherou@sandia.gov

## MS70

### Towards Performance Portable Assembly Tools for Multi-Fluid Plasma Simulations

Supporting scalable and performant finite element assembly on next generation architectures can add significant code complexity. This presentation will discuss the design of general finite element assembly tools applied to multi-fluid plasma and magnetohydrodynamics simulation. Performance portability is achieved via the Kokkos programming model. The assembly library uses a directed acyclic graph (DAG) for composable physics kernels in a multiphysics setting. Embedded automatic differentiation, applied via templates and operator overloading, is used for generating machine precision sensitivities for implicit and IMEX solvers. Performance results will be shown for Intel

Phi, NVIDIA GPU and Intel Haswell architectures.

Roger Pawlowski

Multiphysics Simulation Technologies Dept.  
Sandia National Laboratories  
rppawlo@sandia.gov

Eric C. Cyr

Computational Mathematics Department  
Sandia National Laboratories  
eccyr@sandia.gov

Sean Miller, Matthew Bettencourt

Sandia National Laboratories  
seamill@sandia.gov, mbetten@sandia.gov

Eric Phipps

Sandia National Laboratories  
Optimization and Uncertainty Quantification Department  
etphipp@sandia.gov

Christian Trott

Sandia National Laboratories  
crtrott@sandia.gov

John Shadid

Sandia National Laboratories  
Albuquerque, NM  
jnshadi@sandia.gov

## MS70

### Preparing Sundials for Exascale Computing Platforms

The SUNDIALS Suite of Nonlinear Differential-Algebraic Solvers and Integrators includes highly robust and adaptive time integration methods and software for ODEs and DAEs. In this talk, we will overview SUNDIALS and discuss ongoing and future developments to prepare the suite for use in exascale environments. These developments will include implementation optimizations as well as algorithmic innovations.

Carol S. Woodward

Lawrence Livermore National Laboratory  
woodward6@llnl.gov

Daniel R. Reynolds

Southern Methodist University  
Mathematics  
reynolds@smu.edu

David J. Gardner

Lawrence Livermore National Laboratory  
gardner48@llnl.gov

Slaven Peles

Lawrence Livermore National Laboratory  
peles2@llnl.gov

John Loffeld

Center for Applied Scientific Computing  
Lawrence Livermore National Laboratory

loffeld1@llnl.gov

## MS71

### Routing on the Channel Dependency Graph

Lossless interconnection networks become more popular in high performance computing systems, data centers and network-on-chip architectures. Such networks require efficient and deadlock-free routing functions to utilize the available hardware. Unfortunately, topology-aware routing functions become increasingly inapplicable, due to irregular topologies, which either are irregular by design or as a result of hardware failures. Existing topology-agnostic routing methods either suffer from poor load balancing or are not bounded in the number of virtual channels needed to resolve deadlocks in the routing tables. We propose a novel topology-agnostic routing approach which implicitly avoids deadlocks during the path calculation, by applying this calculation directly to the complete channel dependency graph, instead of solving both problems separately. We present a model implementation, called Nue, of a destination-based and oblivious routing function. Nue routing heuristically optimizes the load balancing while enforcing deadlock-freedom without exceeding a given number of virtual channels, which we demonstrate based on the InfiniBand architecture.

Jens Domke

Satoshi MATSUOKA Laboratory  
GSIC, Tokyo Institute of Technology  
domke.j.aa@m.titech.ac.jp

## MS71

### Circuit-Switched Interconnects Using Limited Number of Slots

In electrical circuit-switched (ECS) networks, the time slot utilization is usually low because communication traffic load is not balanced and causes resource waste and inefficiency over the network. We develop a network topology generator for ECS to minimize the number of time slots optimized to target applications whose traffic patterns are predictable. The proposed network topology generator uses a recursive bisection technique to systematically determine what topology with minimal set of resources are used to efficiently support a target communication pattern. The methodology makes use of the static analysis of the given communication pattern to reduce path conflicts between communications while using as few resources as possible. Evaluation results show that the minimum necessary number of slots (MNNS) can be reduced to a small number in a generated topology while keeping resource amount less than that in a standard mesh network.

Yao Hu

KOIBUCHI Laboratory  
National Institute of Informatics  
huyao@nii.ac.jp

## MS71

### Optimizing TSUBAME2.5 Network with Sub-Optimal Infrastructure

Data centers or data warehouses as well as supercomputers grow to unprecedented scales. Tens of thousands of computers are integrated in large-scale facilities to process floods of queries or large compute jobs. The network connecting these machines plays a crucial role for successful

operation. As these networks grow, failures of switches and connectors or cables become common, which is something we regrettably had to observe during more than six years of operating our TSUBAME2 system. Especially, many of the InfiniBand cables of TSUBAME2's dual-rail QDR fabric, consisting of over 500 switches and over 7000 links, became broken or half-broken, as well as other network components however with lower annual failure rates. Unfortunately, replacing those malfunctioning interconnect components or having spares already in place in a realistic computing center may be constraint by other factors, such as difficulties to remove specific cables from cable pit or limited space within racks and cable tunnels. These constraints can make it impossible to deploy theoretically optimal topologies and network routing strategies. In this talk, we report our network-related failure analysis and all the real-world challenges we were facing during TSUBAME2's lifetime. Furthermore, we will outline the architectural and operational compromises, with respect to our dual-rail fabric, we had to make to circumvent these challenges.

Akihiro Nomura

Satoshi MATSUOKA Laboratory  
GSIC, Tokyo Institute of Technology  
nomura.a.ac@m.titech.ac.jp

## MS71

### Cable-Geometric Error-Prone Approach for Low-Latency Interconnection Networks

Interconnection network is a main concern in the architecture design of highly parallel systems such as high-density data centers and supercomputers that reach millions of endpoints, e.g., 10M cores for Sunway TaihuLight system. As the number of endpoints of such systems has gradually increased to meet the higher computing and storage demand, the interconnection network is required to provide a low-latency and high communication bandwidth, i.e., less than 1-microsecond latency across systems with a link bandwidth is greater than 100GB/s. In the low-latency design context, our primary aim is to provide a novel solution based on the technology-driven approaches: (1) cable-geometric small-world network topology with custom routing, and (2) the use of an low-latency FEC(forward error correction) error-prone mechanism for a high-speed low-reliable link. Both can present a logarithmic diameter low-radix network with low-latency and high-bandwidth switches, although special computing or ABFT (algorithm-based fault tolerance) design is required for parallel applications.

Nguyen Truong

Graduate University for Advanced Studies (SOKENDAI)  
nguyentt@nii.ac.jp

## MS72

### The Flexible Computational Science Infrastructure (FLECSI) Programming System: Backend Interfaces and Applications

Abstract Not Available At Time Of Publication.

Marc Charest

Los Alamos National Laboratory

charest@lanl.gov

cjbattagl@gmail.com

**MS72****Title Not Available At Time Of Publication**

Abstract Not Available At Time Of Publication

Rob Hoekstra  
Sandia National Labs  
rjhoeks@sandia.gov

**MS73****Title Not Available At Time Of Publication**

Abstract Not Available At Time Of Publication.

Gianfranco Bilardi  
University of Padova  
bilardi@dei.unipd.it

**MS72****Enhancing Runtime and Applications with Non-Blocking Collective Operations**

Scientific codes are perpetually modernized to better handle portability on emerging architectures. To help developers achieve that goal, languages and programming models perpetually evolve, with new features being included in their specification. These new features either provide a new functionality or a new abstraction layer. However, the usefulness of a feature greatly depends on how it is handled in the underneath runtime or compiler. If the implementation of a promising functionality cannot achieve the expected improvement, the feature will be soon abandoned by the developers. And this can start a vicious circle: if a feature is not performant, it will not be used, but if the feature is not used, very little effort will be made to improve the performance. One example of the failing of promising features is MPI Non-blocking collective (NBC) communications. NBC has been released as a part of MPI standard version 3.0 in 2012. Five years later, MPI implementations still struggle to achieve good performance with these communications. The consequence is the rare use of NBC in scientific codes, and moreover, the very rare implementations of algorithms designed for asynchronous collective communications. In this minisymposium, I will present the work done both on one Conjugate Gradient implementation and in the MPC runtime to efficiently support MPI NBC communications.

Julien Jaeger  
CEA  
julien.jaeger@cea.fr

**MS73****Parallel 3D CholeskyQR2 for Rectangular Matrices**

The increasing complexity of modern computer architectures has greatly influenced algorithm design. Algorithm performance on these architectures is now determined by the movement of data. Therefore, modern algorithms should prioritize minimizing communication. In this work, we present a new parallel QR factorization algorithm solved over a tunable processor grid in a distributed memory environment. The processor grid can be tuned between one and three dimensions, resulting in tradeoffs in the asymptotic costs of synchronization, horizontal bandwidth, flop count, and memory footprint. This parallel algorithm is the first to efficiently extend the Cholesky-QR2 algorithm to matrices with an arbitrary number of rows and columns. Along its critical path of execution on  $P$  processors, our tunable algorithm improves upon the horizontal bandwidth cost of the existing CholeskyQR2 algorithm by up to a factor of  $c^2$  when solved over a  $c \times d \times c$  processor grid subject to  $P = c^2 d$  and  $c \in [1, P^{\frac{1}{3}}]$ . The costs attained by our algorithm are asymptotically equivalent to state of the art QR factorization algorithms that have yet to be implemented. We argue that ours achieves better practicality and flexibility while still attaining minimal communication.

Edward Hutter  
Illinois  
hutter2@illinois.edu

**MS72****Integrating Existing Fortran Codes with Modern Runtime Systems**

Abstract Not Available At Time Of Publication.

Joshua Payne  
Los Alamos National Laboratory  
payne@lanl.gov

**MS73****Matrix Multiplication, A Little Faster**

Strassen's matrix multiplication algorithm (1969) was followed by Winograd's (1971) who improved it's complexity by a constant factor. Many asymptotic improvements followed. Yet, they have done so at the cost of large, often gigantic, hidden constants. Consequently, for almost all feasible matrix dimensions, Strassen-Winograd's remains the fastest. The leading coefficient of Strassen-Winograd's algorithm was believed to be optimal for matrix multiplication algorithms with  $2 \times 2$  base case, due to a lower bound of Probert (1976). Surprisingly, we obtain a faster matrix multiplication algorithm, with the same base case size, the same asymptotic complexity, but with a smaller leading coefficient. To this end, we transform the input matrices to a different basis, which can be done fast. We discuss improvements in the communication costs, its effects on parallelization and the extension of this method to other Strassen-like algorithms. We also prove a generalization of Probert's lower bound that holds under change of basis. This shows that for matrix multiplication algorithms with a  $2 \times 2$  base case, the leading coefficient of our algorithm cannot be further reduced, hence optimal.

**MS73****Faster Parallel Tensor Compression using Randomization**

Abstract Not Available At Time Of Publication.

Casey Battaglin  
Georgia Tech

Elaye E. Karstadt, Oded Schwartz  
Hebrew University of Jerusalem

elaye.karstadt@mail.huji.ac.il, odedsc@cs.huji.ac.il

#### MS74

##### **Xcalableacc: Highly Productive Accelerated Programming Language for Extreme-Scale Computing**

In the next generation of ultra scale high performance computing platform, one of the most important hardware issues is how to improve the performance/power efficiency as well as its absolute performance, and accelerators such as advanced GPUs are considered as the most promised solution. There are several programming models on GPU coding for HPC within a node such as CUDA, OpenACC or OpenCL. However, these programming environments should be combined with an orthogonal programming paradigm for message passing such as MPI to be executed on distributed memory architecture systems. It causes very high cost on programming and code maintenance. We have been developing a programming environment named XcalableACC to allow OpenACC description combined with directive base parallel language named XcalableMP in PGAS model for distributed memory systems. Both languages support two memory models named global-view and local-view where the local-view model allows programmers to describe a remote memory access between nodes by coarray access while the global-view model enables programmers to describe array partitioning and computing assignments on distributed arrays with quite simple directive notations. In this talk, we present the overview and implementation of XcalableACC compiler for commodity GPU clusters and low power cluster by Exascale's many core accelerators.

Taisuke Boku

Center for Computational Sciences, University of Tsukuba  
taisuke@cs.tsukuba.ac.jp

Hitoshi Murai

RIKEN

h-murai@riken.jp

Masahiro Nakao

RIKEN Advanced Institute for Computational Science

masahiro.nakao@riken.jp

Akihiro Tabuchi

Graduate School of Systems and Information Engineering

University of Tsukuba

tabuchi@hpcs.cs.tsukuba.ac.jp

Mitsuhisa Sato

RIKEN Advanced Institute for Computational Science

msato@riken.jp

#### MS74

##### **MPC: A Runtime Suited for Hybrid Programming**

With the advent of multicore and manycore processors as building blocks of HPC supercomputers, many applications shift from relying solely on a distributed programming model (e.g., MPI) to mixing distributed and shared-memory models (e.g., MPI+OpenMP). It allows better exploitation of shared-memory communications and reduction of the overall memory footprint. One side effect of this programming approach is runtime stacking: mixing multiple models involves various runtime libraries to be alive at the same time and to share the underlying computing resources. Indeed runtime libraries implementing those models are not usually designed to be interoperable

with each other, leading to potential cache interference, synchronization overhead or unnecessary time loss in communications. MPC offers a solution to the runtime stacking limitations. In MPC, several programming models (including MPI, OpenMP and Pthread) are implemented on top of a unified runtime. This unified runtime uses all information about the supported programming models to optimize the placement of worker units (processes and threads), and ensure fairness in the competition for processing resources. Having one runtime to handle both distributed and shared memory parallel programming paradigms allows dodging most of the issues raised by runtime stacking. In this minisymposium, I will present the main features of MPC framework regarding mixing multiple programming models.

Patrick Carribault

CEA

patrick.carribault@cea.fr

#### MS74

##### **Sunway Openacc: Extended Programming Support for the Sunway Many-Core Architecture**

The Sunway TaihuLight supercomputer is the world's first system with a peak performance greater than 100 PFlops, and a parallel scale of over 10 million cores. Different from other existing heterogeneous supercomputers, the system adopts its unique design strategies in both the architecture of its 260-core Shenwei CPU and its way of integrating 40,960 such CPUs as 40 powerful cabinets. This talk reports Sunway OpenACC, an extended design from the OpenACC 2.0 standard that supports more convenient multi-threading programming on the Sunway system. Performance results of typical HPC applications using the OpenACC approach would also be discussed.

Haohuan Fu

Tsinghua University

National Supercomputing Center in Wuxi

haohuan@tsinghua.edu.cn

#### MS74

##### **Graph of Component Languages for Extreme Computational and Data Science; the YML Example**

Future supercomputers are expected to have highly hierarchical architectures with nodes composed by lot-of-core processors and accelerators. The different programming levels will generate new difficult algorithm issues. New intelligent applications would mix computational and data sciences. Moreover, methods have to be redesigned and new ones introduced or rehabilitated, in particular in terms of communication optimizations and data distribution. Then, new languages and frameworks should be defined and evaluated with respect to modern state-of-the-art of scientific methods. In this talk, we first present a solution based on graph of components and/or container-programming experimented on several machines. Hence, we focus on YML with its high level language allowing to automate and delegate the managements of dependencies between loosely coupled clusters of processors to a specialized tool controlling the execution of the application using parallel tasks. Each components may be developed using the PGAS language XMP and container would be associated to YARN or other data science tools. We present some methods for scientific computation well-adapted for such programming paradigm and we discuss new researches mixing components and containers associated with compu-

tational science and data science applications.

Serge Petiton  
University of Lille  
CNRS  
serge.petiton@univ-lille1.fr

Nahid Emad  
PRiSM laboratory and Maison de la Simulation,  
University of  
Nahid.Emad@uvsq.fr

Laurent Bobelin  
LI, universit  Fran ois Rabelais  
Tours, France  
laurent.bobelin@univ-tours.fr

## MS75

### Use of Synthetic Benchmarks for Machine Learning-Based Performance Auto-Tuning

Abstract Not Available At Time Of Publication.

Tarek S. Abdelrahman  
Department of Computer Science  
The University of Toronto  
tsa@eecg.utoronto.ca

## MS75

### Compiler-based Software Autotuning

Abstract Not Available At Time Of Publication.

Protonu Basu  
Lawrence Berkeley National Laboratory  
pbasu@lbl.gov

## MS75

### High Precision Computing of Matrix-Matrix Multiplications and a New Approach of Auto-Tuning to Numerical Libraries by Deep Learning

In this presentation, topics are two-folded. First topic is related to accuracy assurance for fundamental numerical computations. Matrix-Matrix Multiplications (MMM) is one of fundamental numerical computations, but results of accuracy are not well-discussed, while that of execution speeds is carefully evaluated. With respect to this background, our team is now studying high performance implementations of high precision algorithm of MMM. In the first topic of the talk, we parallelize a high-precision MMM using thread-level parallelism. In addition, we conduct a performance evaluation from the viewpoints of execution speed and accuracy with a recent supercomputer environment in Japan, which is the Fujitsu FX100 installed in Information Technology Center, Nagoya University. Second topic is related to auto-tuning (AT) methodology by using AI technologies. Recently, deep learning (DL) is focused on many fields. In this presentation, we propose an AT method by utilizing DL for numerical libraries. Key idea is using pictures of input sparse matrices, and its associated information. Results with SuiteSparse Matrix Collection indicates that: Proposed method is enough high accuracy to select better parameter of performance on a numerical library.

Takahiro Katagiri  
Information Technology Center

Nagoya University  
katagiri@cc.nagoya-u.ac.jp

Shuntaro Ichimura  
Graduate School of Information Science, Nagoya  
University  
ichimura@hpc.itc.nagoya-u.ac.jp

Kenya Yamada  
Graduate School of Informatics, Nagoya University  
yamada@hpc.itc.nagoya-u.ac.jp

## MS75

### Leveraging Formal Methods for Algorithmic Design

Selecting an appropriate algorithm from a set of multiple algorithms for a single operation often depends on the desired performance characteristic (amount of parallelism, memory footprint etc). The solution space becomes significantly larger when we want to find compatible algorithms for multiple operations that can be co-optimized. In this talk, we discuss how the search space of algorithms can be reduced through the raising of the optimization framework to a higher level of abstraction. Specifically, we leverage the loop invariant, a formal method construct traditionally used for verification of loop-based programs, to identify algorithms with specific performance characteristics. Using examples from the dense linear algebra domain, we demonstrate how the loop invariant allows us to quickly eliminate algorithms without the desired performance characteristics.

Tze Meng Low  
Carnegie Mellon University  
lowt@andrew.cmu.edu

## MS76

### Programmable Neuromorphic Learning Machines

Abstract: Intel has embarked on a revolutionary new program to innovate in the area of computer architectures by drawing inspiration from biology. The goal of this effort is to address the rapidly growing demands for chips that can support various artificial intelligence (AI) applications. This bottoms-up rethink of the computer architecture has several ways by which it differs from the traditional Intel architecture, including fine grained parallelism, massive fanout, fully integrated memory and computation and support for various forms of neuronal and synaptic plasticity mechanisms on-chip. Intel has developed a first of its kind hardware design that supports all these features with an on-chip programmable learning engine using Intel's 14 nm CMOS technology. I will outline the current state of this technology including development and benchmarking of spiking models with plasticity to enable unsupervised, supervised and reinforcement modes of learning. Finally, I will highlight challenges from both a modeling and hardware perspective to enable intelligent machines that can benefit from this technology.

Mike Davies, Narayan Srinivasa  
Intel  
"mike.davies@intel.com" jmike.davies@intel.com,  
narayan.srinivasa@intel.com

## MS76

### Neuromorphic Computing at Lawrence Livermore

## National Laboratory

As a world leader in scientific computing, Lawrence Livermore National Laboratory (LLNL) is deeply invested in advances in supercomputing technologies for science and engineering applications. The Sierra system will move LLNL closer to exascale, with 120 to 150 petaflops peak performance<sup>1</sup>. The increases in performance result in increases in energy usage. Neuromorphic hardware, specifically IBMs TrueNorth chip, provides a potential for increased, energy-efficient compute power, with 46 billion synaptic operations per second consuming 70 milliwatts<sup>2</sup>. To evaluate this potential, LLNL installed an NS16e system, which combines 16 TrueNorth chips in a 4x4 array. This research evaluates various applications on neuromorphic and neuro-inspired simulators, emulators, and hardware for performance, accuracy, scalability, and energy usage. These applications include graph-based optimization problems and modern artificial neural networks, with the goals of quantifying which algorithms are best suited for these architectures and influencing hardware vendors when trade-offs between performance, accuracy, scalability, and energy usage are required. The initial set of applications and the evaluation of the aforementioned metrics for these applications on advanced, neuro-inspired architectures will be presented. <sup>1</sup><https://asc.llnl.gov/coral-info> <sup>2</sup><http://www.research.ibm.com/articles/brain-chip.shtml>

Katie Lewis

Lawrence Livermore National Laboratory  
lewis66@llnl.gov

## MS76

### Neuromorphic Computing for Scientific Applications

Neuromorphic computers promise the ability to do intelligent computing through neural network-style operation, but they do so faster, on lower power and with smaller footprint than traditional computer architectures, making them ideal for deployment into real-world environments and for applications requiring real-time performance. In this talk, an overview of the current state of neuromorphic computing will be presented, including a brief background on neuromorphic models, algorithms, hardware, and applications in the literature. Then, two neuromorphic models will be presented in detail: a fully digital architecture with both custom chip and FPGA-based implementations and a mixed analog-digital architecture with memristor-based synapses. Several applications of these two architectures will be presented to demonstrate the capabilities of neuromorphic systems in solving complex problems, including those in the scientific domain. These applications include a high-energy physics data classification application, a real-world anomaly detection application, and an autonomous robot control application. The promise of neuromorphic computing as applied to future, more complex, spatiotemporal scientific data analysis and control tasks will also be discussed.

Catherine Schuman

Oak Ridge National Laboratory  
schumancd@ornl.gov

## MS76

### Neural-Inspired Technologies for Data Processing and Scientific Computing

As HPC has faced power-budget challenges using commer-

cial hardware, researchers have turned towards alternative computing approaches to reduce power-per-computation metrics and to develop more efficient algorithms. Neural-inspired computing has matured into a field that has produced promising methodologies for solving difficult problems in diverse areas such as pattern recognition and scientific computing. Yet the challenges with neural-inspired computing are well-known such as the need for large amounts of training data and the relative brittleness of algorithms to data variability. We have addressed these concerns by strengthening the connection between machine learning and neuroscience concepts, termed neural machine learning, in order to improve the resiliency and adaptability of such algorithms. We present a suite of software and hardware technologies for implementing neural machine learning algorithms for pattern recognition (acoustic and image data) and numerical computing. Specifically, we highlight a modified form of deep learning that enables continuous learning via dynamically expanding the information content within the network with new and strategically-placed processing neurons. Finally, we also demonstrate neural algorithms that have been optimized to directly compute numerical functions such as matrix operations, and demonstrate how to map these functions into neuromorphic hardware to improve speed and power-consumption of the computations.

Conrad James

Sandia National Laboratories  
cdjame@sandia.gov

Craig Vineyard

Sandia National Laboratories, USA  
cmviney@sandia.gov

## MS77

### High Performance Computing Methods for Scheduling Transportation Network Maintenance

Major disasters like earthquakes, tsunamis, and floods may cause substantial disruption to essential services, and particularly to the transportation network. One of the challenges following such disasters is the need to schedule roadway repair activities, in a way that maximized the benefits to the society. Predicting and managing traffic flows in regional transportation networks is a complex modelling and planning problem, requiring advanced computational methods. The need to examine various schedules of reconstruction actions enhance the computational challenges considerably, due to the combinatorial nature of the problem. In this research we focus on the effect of construction projects schedule on Total System Travel Time  $TSTT$ . In this sense our approach is related to the literature on bi-level programming for the Network Design Problem  $NDP$ , where the classical static traffic assignment is used to predict flows at a lower level problem, while network properties, and particularly capacities, are determined at the upper level. The network restoration scheduling problem is more closely related to the Work-zone Scheduling Problem  $WSP$ , which we have been studying recently. The computational challenge is addressed by a novel solution method that combines branch-and-price concepts for the upper level problem with state-of-the-art solver for the lower level problem.

Hillel Bar-Gera

Ben-Gurion Univ. of the Negev  
Department of Industrial Engineering and Management  
bargera@bgu.ac.il



David Rey  
UNSW  
d.rey@unsw.edu.au

### MS77

#### **Cyber-Physical System and Industrial Applications of Large-Scale Graph Analysis and Optimization Problem**

In this talk, we present our ongoing research project. We have started the research project for developing the Urban OS (Operating System) on a large-scale city from 2013. The Urban OS, which is regarded as one of emerging applications of cyber-physical system (CPS), gathers big data sets of people and transportation movements by utilizing different sensor technologies and storing them to the cloud storage system. We have another research project whose objective is to develop advanced computing and optimization infrastructures for extremely large-scale graphs on post peta-scale supercomputers. For example, our project team was a winner of at the eighth, and 10th to 14th Graph500 benchmark. The Urban OS employs the graph analysis system developed by this research project and provides a feedback to a predicting and controlling center to optimize many social systems and services. We also focus on the Hierarchical Data Analysis and Optimization System (HDAOS) based on CPS. First, we gather a variety of data sets on a physical space and generate mathematical models for analyzing the social mobility of real worlds. In the next step, we apply optimization and simulation techniques to solving them, and obtain solutions and check the validity of them on cyber space. We finally feedback these solutions to real world.

Katsuki Fujisawa  
Institute of Mathematics for Industry  
Kyushu University  
fujisawa@imi.kyushu-u.ac.jp

### MS77

#### **Simulating Traffic Flow and Demand Pattern by a Parallel Computer**

A road network can be severely congested after a major disaster. Its capacity can be significantly reduced by damages on infrastructure and buildings. Excess traffic demand such as relief transport after a disaster will also make congestion much worse. Assessing congestion of a road network in a recovery process from a disaster is essential to maintain disaster resilience of a city. Network traffic assignment is a popular approach to assess congestion in a road network with a demand profile given. It models delays on each road and assign vehicles on roads assuming that they choose shortest paths from their origins to their destinations. A number of studies of network traffic assignment have been made in transportation sciences. An issue in assessing road congestions after a major disaster is how to estimate demand pattern before the disaster occurs. Because there is a lot of uncertainties in estimating demand pattern, it is essential to use various possible demand patterns to estimate congestions. It naturally incurs a lot of calculations of traffic assignment, and consequently each calculation must be performed as fast as possible. This presentation introduced a parallelised algorithm for fast calculations of dynamic traffic flow in a network and its implementation on a high-performance computer. An issue in the parallelisation is how to synchronise traffic patterns between different subsections of road networks. Its details

are to be given in the session.

Takamasa Iryo  
Kobe University  
Department of Civil Engineering  
iryoy@kobe-u.ac.jp

Junji Urata, Genaro Peque Jr., Kazuki Fukuda, Riki Kawase  
Kobe University  
urata@person.kobe-u.ac.jp, gpequejr@panda.kobe-u.ac.jp, 163t130t@stu.kobe-u.ac.jp, kawase@stu.kobe-u.ac.jp

Lalith Wijerathne, Wasuwat Petprak  
University of Tokyo  
Earthquake Research Institute  
lalith@eri.u-tokyo.ac.jp, wasuwat@eri.u-tokyo.ac.jp

### MS77

#### **Dynamic Programming of Firms' Activities and Interactions in Disaster Recovery Process**

This study aims at numerical simulation of disaster recovery process of firms where they control stocks of production facilities by recovery investment after they are damaged by disaster. Each firm is also faced with environmental state variables such as infrastructure, workers, and prices on which firms transact with one another. Disaster recovery problem is inherently a target of dynamic programming of large dimensions of state variables although "Curse of dimensionality" has prevented us from computing those problems. This study solves the dynamic optimization problem with a number of state variables in the framework of a system of Bellman (recursive) equations, from which dynamic path of economy is derived applying a rule of sequential equilibrium of market prices. Computing process is composed of two steps as follows. 1) Solve Bellman equation of each firm where probabilistic change of prices is taken into account and make a list of optimal behaviors that correspond to each state vector. Parallel computing is applied to optimization process under each state. 2) Introduce dynamic path of economy where prices of the next period are determined based on the gap of demand and supply in the current period. A case study of Tokyo Metropolitan area figures out the expected time for industries to recover their activities after disaster as well as disparity among industries.

Muneta Yokomatsu  
Kyoto University  
Disaster Prevention Research Institute  
yoko@drs.dpri.kyoto-u.ac.jp

Yuki Akiyama  
CSIS, The University of Tokyo  
aki@csis.u-tokyo.ac.jp

Yoshiki Ogawa, Ryosuke Shibasaki  
CSIS, The University of Tokyo  
ogawa@csis.u-tokyo.ac.jp, shiba@csis.u-tokyo.ac.jp

### MS78

#### **Data Analysis with SVD for Physical Experiments, Application to the Cherenkov Telescope Array**

The Ground based gamma-ray detector, as H.E.S.S., MAGIC or VERITAS, reconstruct gamma-rays directions and energies from pictures recorded by telescopes on

the ground. The best reconstruction methods use pre-computed pixel images comparison. However, this kind of analysis imposes great constraints for the Cherenkov Telescope Array that will produce hundreds of peta-bytes. To remedy this problem, we propose a new parallel SVD-based algorithm to extract each pictures information allowing stereoscopic reconstruction and analyze then obtained performances.

Pierre Aubert

IN2P3

pierre.aubert@lapp.in2p3.fr

Thomas Vuillaume

CNRS/IN2P3/LAPP

Savoie Mont-Blanc University

vuillaume@lapp.in2p3.fr

Florian Gaté

CNRS/IN2P3/LAPP

florian.gate@lapp.in2p3.fr

Gilles Maurin, Jean Jacquemier

CNRS/IN2P3/LAPP

Savoie Mont-Blanc University

gilles.maurin@lapp.in2p3.fr,

jean.jacquemier@lapp.in2p3.fr

Nahid Emad

PRiSM laboratory and Maison de la Simulation,

University of

Nahid.Emad@uvsq.fr

Giovanni Lamanna

CNRS/IN2P3/LAPP

Savoie Mont-Blanc University

lamanna@lapp.in2p3.fr

## MS78

### Scalable Computation of Many Eigenpairs of a Large Hermitian Matrix

There are continual and compelling needs for computing many eigenpairs of a large sparse Hermitian matrix in computational science and engineering and data analysis. Most of codes developed in the past are not adequate or efficient for such needs, even when vast computing resources are available. In recent years, there are a number of efforts to address this challenging task by algorithm and software developers. In this talk, we will first review the current efforts, and then discuss a number of practical techniques we are working on for numerical stability and robustness, communication-avoiding and scalability.

Luan Nguyen

Department of Computer Science

University of California, Davis

lounghuyen@ucdavis.edu

Zhaojun Bai

Departments of Computer Science and Mathematics

University of California, Davis

bai@cs.ucdavis.edu

## MS78

### Contour Integral Method to Evaluate Self-Energy Matrices for Large-Scale Electron Transport Cal-

## culations

Solving the nonlinear eigenvalue problem constitutes the most computationally expensive part of the computation of self-energy matrices used in electron transport calculations. We present an efficient and scalable method to solve this nonlinear eigenvalue problem. Key ingredients of the algorithm include pseudopotentials implemented on real-space finite-difference scheme and the use of the nonlinear Sakurai-Sugiura method. A remarkable feature of the algorithm replaces the computations of inverting the Hamiltonian matrix and the large dense eigenvalue-problem by solving sparse linear equations that can be solved by shifted biconjugate gradient method quickly. The procedure is therefore hundreds to thousands times faster than the most efficient method proposed for real-space grid approach. In addition, inherently rich parallelisms in Sakurai-Sugiura method and domain-decomposition technique for sparse real-space Hamiltonian matrix make the code accessible to highly challenging first-principles calculations that were not feasible before. We here discuss the parallel performance of the proposed method on Oakforest-PACS. The presented numerical tests demonstrate that the method is robust, accurate, efficient, and scales very well to a large number of processors.

Shigeru Iwase

Graduate School of Pure and Applied Sciences

University of Tsukuba

iwase@ccs.tsukuba.ac.jp

Yasunori Futamura

University of Tsukuba

futamura@cs.tsukuba.ac.jp

Akira Imakura

Department of Computer Science

University of Tsukuba

imakura@cs.tsukuba.ac.jp

Tetsuya Sakurai

University of Tsukuba

sakurai@cs.tsukuba.ac.jp

Shigeru Tsukamoto

Forschungszentrum Jülich

s.tsukamoto@fz-juelich.de

Tomoya Ono

University of Tsukuba

ono@ccs.tsukuba.ac.jp

## MS78

### Scalable Eigensolver for Large-scale Electronic Structure Calculations

Electronic structure calculations based on Kohn-Sham density functional theory requires solving a sequence of Hermitian eigenvalue problems. The number of eigenpairs to be computed is proportional to the number of atoms in the nanosystem. For problems that contains thousands or tens of thousands of atoms, the number of eigenpairs can be very large even though it is a small percentage of the dimension of the matrix to be diagonalized. We discuss a number of techniques for compute a large number of eigenpairs efficiently on massively parallel computers.

Chao Yang, Eugene Vecharynski

Lawrence Berkeley National Lab

cyang@lbl.gov, eugene.vecharynski@gmail.com

Yousef Saad  
Department of Computer Science  
University of Minnesota  
saad@cs.umn.edu

Ruipeng Li  
Center for Applied Scientific Computing  
Lawrence Livermore National Laboratory  
li50@llnl.gov

Yuanzhe Xi  
University of Minnesota  
yxi@umn.edu

#### MS79

##### **Productive Graph Analytics Enabled by Graph-BLAS**

We present the design, implementation and experimental results of a distributed-memory parallel algorithm for finding connected components in an undirected graph. Our algorithm is based on the Awerbuch-Shiloach algorithm and deploys highly optimized GraphBLAS primitives to parallelize performance-critical kernels. We also highlight several other graph algorithms following the same design principle. In-house Combinatorial BLAS library enables quick development of these algorithms, boosting the productivity of many graph-based applications from computational biology and scientific computing.

Ariful Azad, Aydin Buluc  
Lawrence Berkeley National Laboratory  
azad@lbl.gov, abuluc@lbl.gov

#### MS79

##### **Mining for Communities in Large-scale Graphs**

Community detection is a widely used graph operation aimed at discovering tightly-knit subsets of vertices from an input graph. It finds application in a number of scientific and social computing disciplines. However, implementing the operation at scale remains a daunting challenge owing to inherent irregularity coupled with the lack of parallel architectures ideally suited for graph data structures. In this talk we will present challenges and recent algorithmic advances for efficient parallel community detection.

Ananth Kalyanaraman  
Boeing Centennial Chair, Associate Professor  
School of EECS, Washington State University  
ananth@eecs.wsu.edu

Mahantesh Halappanavar  
Pacific Northwest National Laboratory  
hala@pnnl.gov

Hao Lu  
Oak Ridge National Laboratory  
luh1@ornl.gov

#### MS79

##### **Connected Components and Spanning Trees on Multi-GPU Systems**

The problem of computing spanning trees and connected

components for graphs is a fundamental problem in a number of scientific applications. The current fastest sequential algorithm is Rems algorithm, which follows the classical disjoint-set scheme. We present a new version of this algorithm suitable for multi-GPU computers. Through a number of tests we show that our implementation is competitive compared to other suggested GPU algorithms for these problems.

Fredrik Manne  
University of Bergen, Norway  
Fredrik.Manne@ii.uib.no

Md Naim  
University of Bergen  
Norway  
md.naim@uib.no

Jean Blair  
United States Military Academy, West Point  
jean.blair@usma.edu

#### MS79

##### **Parallel Algorithm Design Via Approximation**

We describe a paradigm for designing parallel algorithms via approximation techniques. Instead of solving a problem exactly, we seek a solution with provable approximation guarantees via the design of algorithms that have high degrees of parallelism. We use matchings and edge covers in graphs as examples of this paradigm. We show that many of the earlier approximation algorithms designed for matching and edge-covers are inherently serial, and show how these algorithms could be adapted to increase parallelism. We describe a  $1/2$ -approximation algorithm for maximum weighted b-matching, and  $3/2$ - and  $2$ -approximation algorithms for minimum weighted b-edge covers. We show that this technique leads to fast and scalable parallel algorithms for these problems on shared memory and distributed memory computers. We will also describe a few applications of matchings and edge covers.

Alex Pothén  
Purdue University  
Department of Computer Science  
apothén@purdue.edu

Arif Khan  
Pacific Northwest National Lab  
ariful.khan@pnnl.gov

#### MS80

##### **Parallel Adaptive Mesh Refinement in Sam(oa)<sup>2</sup> – Load Balancing vs. Work Stealing**

One of the governing assumptions when developing parallel applications is that assigning equal work to each CPU results in roughly equal runtime per CPU. However, dynamic hardware variability through different clock speeds and increasingly heterogeneous systems limit the accuracy of this performance model. Expecting that these effects worsen on future exascale systems, we present first results of reactive and performance-introspective work stealing in the parallel AMR framework sam(oa)<sup>2</sup>. Our hybrid MPI+OpenMP work stealing implementation mitigates imbalances between MPI ranks by selectively stealing grid sections without making use of a separate programming model, e.g., when variations of the clock frequency are enforced. We also discuss to what extent work steal-

ing can lead to improved performance when compared to a classical load balancing approach of chains-on-chains partitioning.

Philipp Samfass  
TU Munich  
samfass@in.tum.de

## MS80

### Phase Asynchronous AMR Algorithm

A straightforward Adaptive Mesh Refinement (AMR) algorithm typically contains many synchronization points which prevents good scalability of the algorithm on a large number of cores. We propose a phase asynchronous AMR algorithm that reduces most of the synchronization costs without introducing too much programming overhead. In the phase asynchronous algorithm, each subgrid at different AMR levels is considered as a task, where a task within a level can perform computation independent of other tasks at the same level as soon as its boundary data is available. Our method enables legacy applications implemented using the synchronous AMR algorithm to get the benefits of the communication and computation overlap. We discuss the implementation of the phase asynchronous algorithm in the context of the AMReX framework that is being developed as a part of the Block-Structured AMR Co-Design Center in DOE's Exascale Computing Project. We demonstrate performance results from Castro, a large-scale compressible astrophysics simulation code built on the AMReX framework.

Didem Unat  
Lawrence Berkeley National Laboratory  
dunat@lbl.gov

Nufail Farooqi  
Koc University  
mfarooqi14@ku.edu.tr

Tan Nguyen  
LBNL  
tannguyen@lbl.gov

Weiqun Zhang  
Lawrence Berkeley National Laboratory  
Center for Computational Sciences and Engineering  
weiqunzhang@lbl.gov

Ann S. Almgren  
Lawrence Berkeley National Laboratory  
asalmgren@lbl.gov

## MS80

### Comparison Between Different Versions of AMR

Structured adaptive mesh refinement can be implemented in many different ways, including cell-based, patched-based, or octree-based. Each implementation has its own strength and weaknesses, and solution fidelity, computational efficiency and memory footprint vary among them. We use a few representative applications that have differing demands of fidelity, symmetry, and memory to characterize the behavior of different versions of AMR.

Klaus Weide  
University of Chicago

klaus@flash.uchicago.edu

## MS80

### The Triumvirate of Adaptivity in the Exahype Hyperbolic Equation Solvers: Mesh, Discretisation and Parameter Space Adaption

ExaHyPE is a EU-funded H2020 project building an exascale engine for hyperbolic equation systems. The developer consortium demonstrates the engine's potential at hands of two grand challenges: seismic risk assessment and the numerical search for gravitational waves. The challenges are that demanding in terms of computational cost and required memory that the use of adaptivity becomes mandatory. In the talk we distinguish three different notions of adaptivity and discuss their interplay as well as their impact on performance and scalability: Classic mesh (h)adaptivity is provided by a spacetime mesh. The ADER-DG discretisation replaces oscillating and physically inadmissible solutions with high resolution Finite Volume schemes, i.e. it provides adaptivity in terms of the mathematical models. Finally, our code base supports support for parameter and sensitivity studies, i.e. we can adapt in the parameter space, too.

Tobias Weinzierl  
Durham University  
tobias.weinzierl@durham.ac.uk

## MS81

### Electrostatic Field Analyses of Voxel-based Biological Models by Indirect-BEM and Static-MoM Combined with FMM

Voxel-based indirect boundary element method (IBEM) and voxel-based static method of moments (MoM) are electrostatic field calculation methods designed for analyzing voxel-based biological models. Voxel-based human models are typically produced from images obtained using magnetic resonance imaging (MRI); they are composed of several tens of biological organs and millions, or even billions of cubic voxels. Voxel-based IBEM handles square boundary elements, which are square surfaces of voxels, defined on the boundary of organs. The number of boundary elements are generally  $O(N^{2/3})$ ,  $N$  being the number of voxels. Voxel-based IBEM combined with the fast multipole method is an efficient method for electrostatic field analyses. Diffusion tensor imaging is a type of MRI that can provide approximate reconstructions of anisotropic conductivities of biological tissues. Voxel-based static-MoM is designed for handling the anisotropic conductivities by emulating a volume element with six boundary elements surrounding a voxel. It also uses the fast Fourier transform algorithm to decrease the calculation time. The combined use of IBEM and MoM improves computing performance in analyzing models including both isotropic and anisotropic conductors. Examples of analyses of human models using these methods will be presented. One such example involved handles 188 million unknowns using a computer with 64 GB main memory; the relevant simultaneous linear equations could be solved in 83 minutes.

Shoji Hamada  
Department of Electrical Engineering, Kyoto University  
shamada@kuee.kyoto-u.ac.jp

## MS81

### Boundary Integral Equations for Calculating Com-

### plex Eigenvalues for Open Domains

Determining complex eigenvalues is of interest in many applications related to wave problems in open domains. In waveguides for example, there exist anomalous frequencies near which the behavior of the solution changes suddenly. Some of these anomalies are related to complex eigenfrequencies at which there exist non-trivial solutions to the homogeneous boundary value problems for waveguides. One can compute such eigenvalues with boundary integral methods by finding frequencies at which the homogeneous BIEs have non-trivial solutions. In the proposed talk, we discuss a method of solving complex eigenvalue problems in transmission problems using fast boundary integral methods and the Sakurai-Sugiura projection method, a non-iterative solver of non-linear eigenvalue problems. We note that eigenvalues of a BIE may not necessarily be those of the original BVP. These spurious eigenvalues of BIEs are called fictitious eigenvalues. We show that the true and fictitious eigenvalues can be distinguished with a modification of BIEs. This approach is tested in both waveguides for the Helmholtz equation in 2D and standard Maxwell's equations in 3D. The proposed method provides a new insight into the relation between the accuracy of integral equations and the distributions of fictitious eigenvalues. We show that a properly formulated single integral equation has fictitious eigenvalues with larger imaginary parts and are more accurate than other BIEs tested.

Naoshi Nishimura

Department of Advanced Mathematical Sciences, Kyoto University  
nchml@i.kyoto-u.ac.jp

### MS81

#### Generalizations to the Superfast Divide-and-Conquer Eigenvalue Algorithm

We present some important results on the structure of the eigenvectors of various classes of matrices. As generalizations and extensions of the superfast divide-and-conquer eigensolver for symmetric hierarchically semiseparable (HSS) matrices, we look at how these techniques relying on hierarchical low rank structure and highly structured eigenvector representations can be applied to generalized, non-symmetric, nonlinear, high-dimensional, and sparse eigenvalue problems. All results will be supported by theoretical complexity and stability analysis, as well as by numerical tests from a parallel implementation. We will discuss applications of these algorithms to problems from computational physics and electrical engineering.

James Vogel

Purdue University  
vogel13@purdue.edu

### MS81

#### Accelerating Hierarchical-Matrix Based Linear Solver on a GPU Cluster

The hierarchical low-rank compression of the matrix is a powerful tool to improve the performance of many solvers. In this talk, we discuss our recent efforts to port such a solver onto GPU clusters, Reedbush-H and Tsubame3. We used a batched GPU kernel to utilize the compute power on each node. We have also investigated several techniques to address the GPU-GPU communication including studying the effectiveness of GPUDirect for our particular communication patterns within a node and between the nodes, and

then hiding such communication by overlapping it with computation.

Ichitaro Yamazaki

University of Tennessee, Knoxville  
iyamazak@icl.utk.edu

Satoshi Ohshima

The University of Tokyo  
Information Technology Center  
ohshima@cc.kyushu-u.ac.jp

Akihiro Ida

The University of Tokyo  
ida@cc.u-tokyo.ac.jp

Rio Yokota

Tokyo Institute of Technology  
rioyokota@gsic.titech.ac.jp

Jack J. Dongarra

University of Tennessee, Oak Ridge National Laboratory, USA  
dongarra@icl.utk.edu

### MS82

#### Spectral Graph Analysis with Unite and Conquer Approach

We present the unite and conquer approach for the restarted Krylov methods such as implicitly restarted Arnoldi/Lanczos methods to extract information from graphs representing social networks. These restarted techniques allow decreasing the global number of restart cycles by coupling either synchronously or asynchronously several restarted methods. We show that our solvers issued from this approach are very efficient. This is particularly the case on undirected graph in the context of clustering applications.

Nahid Emad

PRiSM laboratory and Maison de la Simulation, University of  
Nahid.Emad@uvsq.fr

Serge Petiton

University of Lille  
CNRS  
serge.petiton@univ-lille1.fr

Alexandre Fender, Joe Eaton, Maxim Naumov  
NVIDIA

afender@nvidia.com, featon@nvidia.com,  
maxim.a.naumov@gmail.com

### MS82

#### Production Implementations of Pipelined and Communication-Avoiding Iterative Linear Solvers

Many authors have developed algorithms for pipelined and communication-avoiding Krylov subspace methods for solving linear systems. However, implementing them efficiently in a software library designed for production use can be challenging. Fast implementations of pipelined and communication-avoiding Krylov methods demand features both of parallel programming models, and of local computational kernels that the software library developers may not have time or expertise to implement efficiently. This

talk will cover these issues that appear low level, but can influence users' decisions whether to try out new Krylov solvers in production.

Mark Hoemmen

Sandia National Laboratories  
mhoemme@sandia.gov

Ichitaro Yamazaki

University of Tennessee, Knoxville  
iyamazak@icl.utk.edu

## MS82

### Development of Large-Scale Scientific & Engineering Applications on Post-Peta/Exascale Systems Using ppOpen-HPC

"ppOpen-HPC" is an open source infrastructure for development and execution of optimized and reliable simulation code on post-peta-scale (pp) parallel computers based on many-core architectures, and it consists of various types of libraries, which cover general procedures for scientific computation. "ppOpen-HPC" is part of a five-year project (FY.2011-2015) spawned by the "Development of System Software Technologies for Post-Peta Scale High Performance Computing" funded by JST-CREST. The framework covers various types of procedures for scientific computations, such as parallel I/O of data-sets, matrix-assembly, linear-solvers with practical and scalable preconditioners, visualization, adaptive mesh refinement and dynamic load-balancing, in various types of computational models, such as FEM, FDM, FVM, BEM and DEM. Automatic tuning (AT) technology enables automatic generation of optimized libraries and applications under various types of environments. In 2016, the team of ppOpen-HPC joined ESSEX-II (Equipping Sparse Solvers for Exascale) project, which is funded by JST-CREST and DFG-SPPEXA, Germany until FY.2018. In ESSEX-II, we develop pK-Open-HPC (extended version of ppOpen-HPC, framework for exa-feasible applications), preconditioned iterative solvers for quantum sciences, and a framework for automatic tuning (AT) with performance model. In the presentation, various types of achievements of ppOpen-HPC, ESSEX-II, and pK-OpenHPC will be shown.

Kengo Nakajima

The University of Tokyo  
Information Technology Center  
nakajima@cc.u-tokyo.ac.jp

## MS82

### Algebraic Multigrid Methods, Data Structures and Performance

The hypre software library provides high performance preconditioners and solvers for the solution of large sparse linear systems on massively parallel computers. One of its attractive features is the provision of conceptual interfaces, which include a structured, a semi-structured interface, and an unstructured interface. Each of these interfaces uses very different matrix data structures, which affect the performance of the provided multigrid methods. Structured multigrid solvers can take advantage of additional information in the structured matrix data structures, but are confined to structured problems, whereas unstructured multigrid solvers can be applied to more general problems. In addition, work is in progress to design a semi-structured multigrid method that can take advantage of the structured parts of a problem, but is capable to solve more gen-

eral problems. This presentation discusses these methods, their implementation and their performance.

Ulrike Meier Yang

Lawrence Livermore National Laboratory  
yang11@llnl.gov

## MS83

### A Semi-Implicit Solver for Two Phase Flow with Moving Contact Line on Rough Surface

In this talk, we develop an efficient lattice Boltzmann model for the two-phase moving contact line problem with variable density. The NavierStokes and CahnHilliard equations are recovered from the lattice Boltzmann model proposed by Fakhari and Rahimian. To improve numerical stability, we present a semi-implicit lattice Boltzmann method together with a mixed finite difference scheme. In order to describe the behavior of the contact line motion on the boundary, we incorporate the generalized Navier boundary condition by the nonequilibrium extrapolation method. The proposed method is easy to implement and retains the advantage of the standard lattice Boltzmann method. Numerical tests are carried out to verify the proposed method. Our numerical results show that the present approach is able to model two-phase flows with variable density and moving contact line.

Jizu Huang

Chinese Academy of Sciences  
huangjz@lsec.cc.ac.cn

## MS83

### A Parallel Full Space Lagrange-Newton-Krylov Algorithm for Trajectory Optimization Problems in Space Missions

The objectives of this work are to apply and study the full-space quasi Lagrange-Newton-Krylov (FQLNK) algorithm for solving trajectory optimization problems arising from aerospace industrial applications. As its name suggests, we first convert the constrained optimization problem into an unconstrained one by introducing the augmented Lagrangian parameters. The next step is to find the optimal candidate solution by solving the Karush-Kuhn-Tucker (KKT) condition with the Newton-Krylov method. To reduce the computational cost of constructing the KKT system, we employ the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula to build an approximation of the (1,1) subblock of the KKT matrix, which is the most expensive part of the overall computation. The BFGS-based FQLNK algorithm exhibits a superior speedup compared to some of the alternatives. We demonstrate our FQLNK algorithm to be a practical approach for designing an optimal trajectory of the launch vehicle in the space missions.

Feng-Nan Hwang

National Central University  
Department of Mathematics  
hwangf@math.ncu.edu.tw

## MS83

### Parallel Monolithic Algorithm for Simulation of Patient-Specific Coronary Artery

We develop a scalable parallel algorithm for the simulation of 3D coronary artery based on patient-specific geometry, physical parameters, and boundary conditions. The problem is discretized by a stabilized finite element method in



space and a fully implicit backward finite difference scheme in time. A parallel Newton-Krylov-Schwarz method is employed to solve the fully coupled large, sparse nonlinear system at each time step. In this talk, we show the parallel performance of the coupled fluid and solid solver obtained on a parallel computer with thousands of processor cores.

Wen-Shin Shiu

Shenzhen Institute of Advanced technology,  
China  
xuw@siat.ac.cn

Rongliang Chen

Chinese Academy of Sciences  
rl.chen@siat.ac.cn

Zhengzheng Yan, Jia Liu

Shenzhen Institute of Advanced technology, China  
zz.yan@siat.ac.cn, jia.liu@siat.ac.cn

Xiao-Chuan Cai

University of Colorado Boulder, Boulder, USA  
cai@colorado.edu

#### MS83

##### **A Fully Implicit Constraint-Preserving Simulator for Multiphase Flow in Porous Media**

Due to the rapid advancement of supercomputing resource, there is a growing interest in developing parallel algorithms for the large-scale reservoir simulation. In this paper, we present a parallel and fully implicit simulator for the multiphase flow in porous media based on the variational inequality (VI) framework, which can be used to enforce important mathematical and physical properties to obtain accurate constraint-preserving solutions. In the proposed parallel simulator, the multiphase flow problem is reformulated as a variational inequality system that naturally satisfies the mass conservation, and then a fully implicit finite method is applied to discretize the model equations. In addition to that, some nonlinear and linear fast solver technologies, including a variant of inexact Newton methods and the domain decomposition based preconditioners, are employed to strengthen the robustness and parallel scalability of the simulator. Several numerical results pertaining to the problems are presented to illustrate the efficiency, robustness, and the overall performance of the fully implicit constraint-preserving simulator.

Haijian Yang

Hunan University  
China  
yanghaijian2@sina.com

#### MS84

##### **Multi-Level Spectral Nested-Dissection Graph Ordering**

Abstract Not Available At Time Of Publication.

Mathias Jacquelin

Lawrence Berkeley National Lab  
mjacquelin@lbl.gov

#### MS84

##### **Enhancing Performance of Sparse Matrix Factor-**

##### **izations Via Ordering Refinements**

In this work, we propose a low complexity heuristic which aims at improving the block structure of non zeros below the diagonal blocks by bringing non-zero rows closer together. This is of particular importance to achieve higher computational efficiency on GPU platforms for instance. We compare our heuristic to the state-of-the-art heuristic based on solving the Traveling Salesman Problem in terms of reordering time, quality of the block structure, and finally impact on numerical factorization time.

Esmond G. Ng

Lawrence Berkeley National Laboratory  
eng@lbl.gov

#### MS84

##### **Variable Clustering Techniques for Low-rank Factorization-based Linear Solvers and Preconditioners**

Sparse linear solvers and preconditioners based on low-rank approximation techniques have gained popularity in recent years. Many follow a standard multifrontal or supernodal approach, that splits the sparse factorization into dense subproblems, typically by ordering the problem using Nested Dissection. A good ordering of the variables of each subproblem is critical to reveal and exploit low-rankness. In this talk, we explore different techniques using our partitioner LS-GPart, for problems coming from the multiphysics code LS-Dyna.

Francois-Henry Rouet

Livermore Software Technology Corp.  
frouet@lstc.com

#### MS84

##### **Title Not Available At Time Of Publication**

Abstract Not Available At Time Of Publication.

Yousef Saad

Department of Computer Science  
University of Minnesota  
saad@cs.umn.edu

#### MS85

##### **Raising the Level of Abstraction of Developing Optimized Parallel Applications**

Developing performance-oriented parallel applications for modern supercomputing platforms can be challenging for several programmers. Therefore, high-level tools that capture the knowledge of expert parallel programmers are needed. One such high-level tool Interactive Parallelization Tool (IPT) will be presented in this talk. IPT is designed to assist users in the process of developing parallel (MPI/OpenMP/CUDA/MPI+X) applications using existing serial versions. IPT reduces the effort required for developing efficient parallel applications by more than 90%.

Ritu Arora

TACC  
rauta@tacc.utexas.edu

#### MS85

##### **Performance Engineering Scientific Applications**

### Using the Tau Performance System

The complex nature of HPC platforms and the application development environment, combining multiple languages, programming paradigms, hardware, and compilers, make effective performance engineering a challenging task. To meet the needs of computational scientists in performance engineering their codes, we present the TAU Performance System. TAU is a powerful profiling and tracing toolkit that covers multiple aspects of performance instrumentation, measurement, and analysis. After describing and demonstrating how performance data is collected using TAUs automated instrumentation, the talk will present ways to analyze the performance data collected and to drill down to find performance bottlenecks. Topics will cover generating performance profiles and traces using the TAU Commander interface with memory and system load utilization metrics, I/O, communication, and hardware performance counter data using PAPI, compiler-based instrumentation, callsite-profiling, generating OTF2 traces natively, using TAU for evaluating CUDA and Python based applications. The talk will cover instrumentation of Hybrid MPI and OpenMP codes, and the use of TAU for profiling Kokkos based applications using the Kokkos Profiling Interface. Analysis of performance data using ParaProf, TAU's 3D profile browser and PerfExplorer will show how to assess the scalability of a parallel application. TAU may be downloaded from <http://tau.uoregon.edu>.

Sameer Shende

Dept. of Computer and Information Science  
University of Oregon  
[sameer@cs.uoregon.edu](mailto:sameer@cs.uoregon.edu)

### MS85

#### User-Defined Code Transformation for Separation of Performance-Awareness from Application Codes

While performance-aware programming is indispensable to exploit the potential of any modern HPC system, it often overcomplicates an application code and thus severely degrades the code maintainability. Therefore, we are exploring an effective way of expressing performance awareness separately from an application code, and developing a user-defined code transformation framework, Xevolver. This talk briefly introduces several case studies to discuss the benefits and limitations of the code transformation approach.

Hiroyuki Takizawa

Tohoku University/JST CREST  
[takizawa@tohoku.ac.jp](mailto:takizawa@tohoku.ac.jp)

### MS85

#### Performance Engineering for Sparse Linear Algebra Kernels: Navigating Between Models and Expectations

Irregular and indirect data access patterns of kernels and algorithms involving sparse matrix operations are a hard problem for analytic performance modelling. A simple, well-known intensity-based performance model for sparse matrix-vector multiplication yields an upper performance bound for this basic but important kernel. We demonstrate how this coarse performance model can be enhanced by heuristics to get more accurate modelling results and apply it to related scenarios.

Gerhard Wellein

Erlangen Regional Computing Center (RRZE), Germany  
[Gerhard.Wellein@rrze.uni-erlangen.de](mailto:Gerhard.Wellein@rrze.uni-erlangen.de)

### MS86

#### Integrating Multiple Random Sketches for Singular Value Decomposition

The singular value decomposition (SVD) is a key tool in dimension reduction and data analytics. The rapid growth in the dimension and sample size of data has increased the need for developing efficient large-scale SVD algorithms. Randomized SVD based on one-time sketching has been studied, and its potential has been demonstrated for computing a low-rank SVD. Instead of exploring different single random sketching techniques, we propose a Monte Carlo type integrated SVD algorithm based on multiple random sketches. The proposed integration algorithm takes multiple random sketches and then integrates the results obtained from the multiple sketched subspaces. The integrated SVD can achieve higher accuracy and lower stochastic variations. The main component of the integration involves an optimization problem with a matrix Stiefel manifold constraint. Our theoretical analysis shows that the leading singular vectors can be induced by population averaging and ensure the consistencies between the computed and true subspaces and singular vectors. Statistical analysis further proves a strong Law of Large Numbers and gives a rate of convergence by the Central Limit Theorem.

Ting-Li Chen

Institute of Statistical Science  
Academia Sinica  
[tlchen@stat.sinica.edu.tw](mailto:tlchen@stat.sinica.edu.tw)

Dawei D. Chang

Institute of Applied Mathematical Sciences  
National Taiwan University  
[davidzan830@gmail.com](mailto:davidzan830@gmail.com)

Su-Yun Huang

Institute of Statistical Science  
Academia Sinica  
[syhuang@stat.sinica.edu.tw](mailto:syhuang@stat.sinica.edu.tw)

Hung Chen

Institute of Applied Mathematical Sciences  
National Taiwan University  
[hchen@math.ntu.edu.tw](mailto:hchen@math.ntu.edu.tw)

Chienyao Lin

Institute of Statistical Science  
Academia Sinica  
[youyuioims94@gmail.com](mailto:youyuioims94@gmail.com)

Weichung Wang

National Taiwan University  
Institute of Applied Mathematical Sciences  
[wwang@ntu.edu.tw](mailto:wwang@ntu.edu.tw)

### MS86

#### Communication Avoiding Approach for Reduction to Tri-Diagonal, Bi-Diagonal, and Hessenberg Forms

For eigenvalue and singular-value decomposition, the three transformations of tri-diagonalization, bi-diagonalization, and reduction to a Hessenberg form are expensing tremendous communication cost in massively parallel processing.

The communication avoiding (CA) approach proposed for Householder tridiagonalization offers the obvious effect of removing 80% of ‘the number of collective communications (startup overhead)’ reaching 15 to 20% of the total computation time. The similar idea of CA can be applied to other transformation methods. Proposal of two reconstructed transformation methods, its complexity analysis, and experimental results of K-computer are presented in the mini-symposium.

Toshiyuki Imamura  
RIKEN Advance Institute for Computational Science  
imamura.toshiyuki@riken.jp

Yusuke Hirota  
RIKEN AICS  
yusuke.hirota@riken.jp

Susumu Yamada, Masahiko Machida  
Japan Atomic Energy Agency  
yamada.susumu@jaea.go.jp,  
machida.masahiko@jaea.go.jp

#### MS86

##### Unconstrained Functionals for Efficient Parallel Scaling of Conjugate Gradient Eigensolvers

A large number of scientific applications require the solution of eigenvalue problems. For applications where some small percentage of the eigenpairs are required, rather than the full spectrum, iterative eigensolvers are typically used. Often, the scaling of these types of iterative eigensolvers on modern massively parallel multi-core computers is limited by the reorthogonalization step, which typically uses direct diagonalization of the subspace matrix, Cholesky or QR decomposition. Algorithms that reduce to as few as possible the number of explicit matrix diagonalizations are expected to display much more favorable scaling and therefore being the preferred choice for very large systems, compared to the actual state-of-the-art parallel implementation. We focus on the unconstrained energy functional formalism for electronic structure calculations, which offers the possibility to fulfilling that requirement. Unconstrained functionals are more complex than the standard functionals but avoid the reorthogonalization step so that the trial vectors are only orthogonal at the minimum. In the presentation, we compare the performance of the unconstrained energy functional formalism with that of existing approaches on realistic problems, and discuss opportunities for performance improvement.

Osni A. Marques  
Lawrence Berkeley National Laboratory  
Berkeley, CA  
oamarques@lbl.gov

Andrew M. Canning, Mauro Del ben  
Lawrence Berkeley National Laboratory  
acanning@lbl.gov, mdelben@lbl.gov

#### MS86

##### Validated Solution of Linear Systems for Real Symmetric and Positive Definite Matrices

This talk concerns verified numerical computations for linear systems. If the coefficient matrix is symmetric positive definite, then a very fast verification method was proposed by Rump and Ogita. Their important technique is to obtain a lower bound of the minimum singular value using

diagonal shift and Cholesky decomposition. If a matrix size is large or a condition number of the matrix is high, the Rump-Ogita method may fail to verify accuracy of the computed results. We improve the diagonal shift by Rump and Ogita. Numerical results illustrate that our method succeeds to work for a wide range of problems compared to the Rump-Ogita method.

Takeshi Terao  
Shibaura Institute of Technology  
mf15059@shibaura-it.ac.jp

Katsuhisa Ozaki  
Shibaura Institute of Technology / JST  
ozaki@sic.shibaura-it.ac.jp

#### MS87

##### ParILUT - A New Parallel Threshold ILU

We present a parallel algorithm for computing a threshold incomplete LU factorization. The main idea is to interleave a parallel fixed-point iteration that approximates an incomplete factorization for a given sparsity pattern with a procedure that adaptively changes the pattern. We describe and test a strategy for identifying fill-in candidates and adaptively adding and removing nonzeros from the sparsity pattern. The resulting pattern may be different and more effective than that of existing threshold ILU algorithms. Also, in contrast to other parallel threshold ILU algorithms, much of the new algorithm has fine grained parallelism.

Hartwig Anzt  
Innovative Computing Lab  
University of Tennessee  
hartwig.anzt@kit.edu

Edmond Chow  
School of Computational Science and Engineering  
Georgia Institute of Technology  
echow@cc.gatech.edu

Jack J. Dongarra  
University of Tennessee, Oak Ridge National Laboratory,  
USA  
dongarra@icl.utk.edu

#### MS87

##### Enhancing a Parallel Iterative Solver Through Randomization and GPU Computing

We illustrate how we can enhance convergence and execution time in the parallel distributed version of the Algebraic Recursive Multilevel Solver (pARMS). First we apply Random Butterfly Transformation (RBT) in the ARMS preconditioner to solve more efficiently the last Schur complement system in the recursive multilevel process. Then we show that performing the local preconditioning phase (ILU or ARMS) on the GPU can result in significant performance gain on the test problems considered in our experiments. Finally we study the effect of soft fault errors on the convergence of flexible GMRES preconditioned with ARMS or ILUT.

Marc Baboulin  
University of Paris-Sud  
marc.baboulin@lri.fr

Evan Coleman

Old Dominion University  
ecole028@odu.edu

Aygul Jamal, Amal Khabou  
University of Paris-Sud  
aygul.jamal@lri.fr, amal.khabou@lri.fr

Masha Sosonkina  
Old Dominion University  
msosonki@odu.edu

## MS87

### Performance Property of Preconditioned Chebyshev Basis CG Solver for Multiphase CFD Simulations

To improve the convergence property of the communication avoiding conjugate gradient (CA-CG) method is needed for applying it to ill conditioned problems such as the pressure Poisson equation in the multiphase CFD code JUPITER. In the CA-CG method, one can avoid more communication by increasing the number of CA steps. However, this makes the CA-CG method less robust against numerical errors. To resolve this problem, we apply the Chebyshev basis CG (CBCG) method to JUPITER.

Akie Mayumi, Yasuhiro Idomura, Takuya Ina, Susumu Yamada  
Japan Atomic Energy Agency  
mayumi.akie@jaea.go.jp, idomura.yasuhiro@jaea.go.jp,  
ina.takuya@jaea.go.jp, yamada.susumu@jaea.go.jp

Toshiyuki Imamura  
RIKEN Advance Institute for Computational Science  
imamura.toshiyuki@riken.jp

## MS87

### Shifted Cholesky QR for Computing the QR Factorization for Ill-conditioned Matrices

The Cholesky QR algorithm is an algorithm for computing the QR factorization of a tall-skinny matrix and ideal from the viewpoint of high performance computing. Unfortunately it has the inherent difficulty regarding numerical instability and breakdown when the matrix is ill-conditioned. A recent work establishes that the instability can be cured significantly by repeating the algorithm twice (called CholeskyQR2). However, the applicability of CholeskyQR2 is still limited by the requirement that the Cholesky factorization of the Gram matrix runs to completion, which means it does not always work for ill-conditioned matrices. In this work we extend the applicability to such matrices by introducing a shift to the computed Gram matrix so as to guarantee the Cholesky factorization  $R^T R = A^T A + sI$  succeeds numerically. We show that the resulting  $AR^{-1}$  has reduced condition number, for which CholeskyQR2 safely computes the QR factorization. We thus obtain a backward stable QR factorization of ill-conditioned matrices: both the residual and deviation from orthogonality are of order the unit roundoff.

Yuka Yanagisawa  
Waseda University  
yuuka@aoni.waseda.jp

Takeshi Fukaya  
Hokkaido University  
fukaya@iic.hokudai.ac.jp

Yuji Nakatsukasa  
University of Oxford  
yuji.nakatsukasa@maths.ox.ac.uk

Yusaku Yamamoto  
The University of Electro-Communications, Japan  
yusaku.yamamoto@uec.ac.jp

Ramseshan Kannan  
University of Manchester  
yuuka@ruri.waseda.jp

## MS88

### Imaging Earth's Mantle: Global Full-Waveform Inversion Based on Spectral-Element Simulations & Adjoint Methods

Having accurate images of Earth's interior is crucial to improve our understanding of the inner dynamics of our planet. Recent advances in numerical methods combined with developments in high-performance computing have enabled unprecedented simulations of seismic wave propagation in realistic 3D Earth models. Global adjoint tomography is one of the extreme projects in seismology due to the intense computational requirements and vast amount of data that can potentially be assimilated in inversions. The first-generation global adjoint tomography model, GLAD-M15, was constructed using data from 253 earthquakes with transverse isotropy confined to the upper mantle. The results suggest that we are approaching the resolution of continental-scale studies in densely covered regions despite using a limited data set. We are now performing the second-generation global adjoint model including surface-wave azimuthal anisotropy while increasing the database in a complementary inversion. We perform our simulations using the GPU version of SPEC-FEM3D-GLOBE on the Oak Ridge Leadership Computing Facility's Cray XK7 Titan system, a computer with 18 688 GPU accelerators. We will perform 9 s simulations (currently 17 s) on Oak Ridge's next generation supercomputer Summit. The ultimate aim is to go down to 1 Hz in global simulations and assimilate all available seismic data from all global CMT earthquakes within the magnitude range of 5.5 to 7.0 in the construction of global models.

Ebru Bozdag  
Colorado School of Mines  
bozdag@mines.edu

Matthieu Lefebvre, Wenjie Lei  
Princeton University  
ml15@princeton.edu, lei@princeton.edu

Ridvan Orsvuran  
University of Cote d'Azur  
ridvan.orsvuran@geoazur.unice.fr

Daniel Peter  
King Abdullah University of Science and Technology  
daniel.peter@kaust.edu.sa

Yuyi Ruan, James Smith  
Princeton University  
yuyir@princeton.edu, jas11@princeton.edu

Dimitri Komatitsch  
Laboratory of Mechanics and Acoustics  
CNRS, Marseille, France

komatitsch@lma.cnrs-mrs.fr

Jeroen Tromp  
Princeton University  
Department of Geosciences  
jtromp@princeton.edu

## MS88

### Exploring Planetary and Stellar Convection Using the Rayleigh Code

Whether originating within a planetary or stellar interior, the fundamental building blocks of a dynamo are the same. Stars and planets typically possess an interior source of heat, and that heat is transported across some fraction of the interior via the bulk motion of an electrically-conducting fluid. As the currents driven by this process occur deep within the interiors of stars and planets, and so remain inherently inaccessible to direct observation, our understanding of the dynamo process derives heavily from theoretical and computational models. Spectral algorithms are a popular choice for modeling such systems, due in part to their inherent numerical accuracy and also, as in the case of the sphere, geometrical considerations. These advantages must be weighed against the high cost of communication, however, as a single time step taken by a spectral method will typically require multiple, global transposes of the distributed system variables. I will discuss the parallelization of a community-sourced spectral code designed to mitigate this problem by minimizing the number of messages initiated within a single time step. This code, named Rayleigh, has been optimized for the study of planetary and stellar dynamos, and it scales efficiently to  $O(10^5)$  cores for simulations with  $O(10^3)$  grid points. After discussing the parallelization of Rayleigh, I will present initial results achieved using the code when applied to Earth, the Sun, and Jupiter.

Nicholas A. Featherstone  
University of Colorado, Boulder  
feathern@colorado.edu

## MS88

### Advances in Mantle Convection Modelling: Non-linear Solvers, Multiphysics, Linking Scales

The past years have seen significant improvements in the evolution of mantle convection modelling, both in the size of performed models and complexity of utilized techniques and processes. Three of the main driving factors for this evolution are constantly increasing computing power, improved numerical algorithms, and a better understanding of how to engineer mantle convection modelling software. In this talk I will present examples of novel techniques and applications using the mantle convection software ASPECT, which is developed as a modular, massively parallel, open-source code for the geodynamic modelling community. In particular, I will present approaches that allow different equations, solvers, and solver schemes to coexist in a single software, thus allowing direct comparisons and benchmarks. I will discuss how a modular software design allows multiphysics computations at the example of coupled grain-size evolution and mantle convection simulations. Finally, I will point out how this approach can cross spatial and temporal scales from years to hundreds of millions of years and from hundreds of metres to thousands of kilometers. In conclusion, I will discuss some of the challenges on the path ahead, including the coupling of software packages, and the verification of increasingly

complex models.

Rene Gassmoeller  
Colorado State University  
rene.gassmoeller@mailbox.org

Wolfgang Bangerth  
Colorado State University  
Texas A&M University  
bangerth@colostate.edu

Juliane Dannberg  
Colorado State University  
judannberg@gmail.com

Timo Heister  
Clemson University  
Mathematical Sciences  
heister@clemson.edu

## MS88

### Frontiers of Scientific Computation in Geodynamics

Scientific computation has long played a key role in solid Earth geosciences; quantitative numerical models provide a critical link between observations at Earth's surface and dynamic processes in the interior. With increases in computational power and sophistication of algorithms, solid Earth geoscience has seen a proliferation of powerful and predictive models; seismology and geodynamics regularly drive development of new frontiers in parallel computing. Examples include the propagation of seismic waves through Earth's structure, origin and evolution of planetary dynamos, multi-physics modeling of fluids and faults in the crust and lithosphere, and long-term thermal evolution of the planet. Each of these scientific problems represents a unique computational grand challenge. To improve our understanding of these problems using scientific computation and large scale parallel computing, the Computational Infrastructure for Geodynamics (CIG) was established as scientific community of practice and partnership between geophysics and computational science. CIG provides computational tools to enable geoscientists to more effectively explore and understand the dynamics of the solid Earth. Scientific software is developed with performance and accuracy benchmarking, and incorporates training and community-building. This talk will discuss mathematical and computational issues and directions for development in computational geodynamics.

Louise Kellogg  
University of California, Davis  
lhkellogg@ucdavis.edu

## MS89

### Multi-GPU Scaling of Eigenvalue Computations for Spectral Partitioning

Spectral methods are powerful tools for finding hidden information in graphs, such as communities which are weakly connected or overlapping. These methods are based on computing some portion of the full set of eigenpairs, which can be computationally challenging, requiring special attention to avoid non-convergence or poor quality results. In this talk we will present our work on acceleration of large scale eigensolvers, scaling to a few billion vertices. The talk will include results from new methods such as optimal subspace adaptation and weighted operators to im-

prove convergence and partition quality.

Joe Eaton, Alex Fender, Maxim Naumov  
NVIDIA

jeaton@nvidia.com, afender@nvidia.com,  
maxim.a.naumov@gmail.com

Nahid Emad  
PRiSM laboratory and Maison de la Simulation,  
University of  
Nahid.Emad@uvsq.fr

Serge Petiton  
University of Paris  
serge.petiton@uvsq.fr

## MS89

### Addressing Scalability in SLEPc Eigensolvers

SLEPc stands for Scalable Library for Eigenvalue Problem Computations, but how scalable is it? We discuss the properties of the different algorithms implemented in SLEPc, in terms of parallel efficiency and scalability. We focus on topics that we have been developing in the last years, such as block orthogonalization, GPU computing, polynomial filtering, two-level MPI parallelization, etc. We will show performance data from different scientific computing applications.

Jose E. Roman  
Universidad Politecnica de Valencia  
jroman@dsic.upv.es

## MS89

### A Parallel Generator of Non-Hermitian Matrices Computed from Known Given Spectra

We often face the problem of evaluating the convergence of iterative or restarted methods to solve very large Non-Hermitian linear algebra problems on parallel and/or distributed machines. The convergence often depends on the degree of clustering of the dominant eigenvalues. It is crucial to have large sets of matrices to experiment on post-petascale machine. In this talk, we introduce a parallel generator which uses the spectra given by users to build very large band matrices and to ensure their eigenvalues to be the given ones. Experiments on several supercomputers will be presented.

Xinzhe Wu  
?CNRS  
xinzhe.wu@cea.fr

Serge Petiton  
University of Lille  
CNRS  
serge.petiton@univ-lille1.fr

Hervé Galicher  
King Abdullah University of Science and Technology  
herve.galicher@kaust.edu.sa

Christophe Calvin  
CEA/DRF

christophe.calvin@cea.fr

## MS89

### Distributed Parallel Implementation of the Sakurai-Sugiura Method for Large Dense Generalized Eigenvalue Problems

We consider utilization of the Sakurai-Sugiura method (SSM) for large-scale dense generalized eigenvalue problems on many core clusters. In the method, solutions of shifted linear systems are required and dominate the computational cost. In this talk, we propose a distributed parallel implementation of the Block BiCGrQ method [Du et al., 2014] as a linear solver used in the SSM. Performance comparisons of the proposed implementation and ScaLAPACK will be shown with a practical problem from an electronic structure calculation. The performance evaluation is performed on a Knights Landing cluster Oakforest-PACS.

Takahiro Yano  
Department of Computer Science  
University of Tsukuba  
yano@mma.cs.tsukuba.ac.jp

Yasunori Futamura  
University of Tsukuba  
futamura@cs.tsukuba.ac.jp

Tetsuya Sakurai  
Department of Computer Science  
University of Tsukuba  
sakurai@cs.tsukuba.ac.jp

## MS90

### Performance Portable Sparse Matrix Matrix Multiplication with Applications in Scientific Computing and Graph Analytics

We approach the problem of sparse matrix-sparse matrix multiplication (SpGEMM) from three perspectives. We, first, consider the problem of writing performance-portable SpGEMM kernel for many-core architectures. Second, we identify and address the performance requirements of SpGEMM in scientific computing applications. Then, we extend this linear-algebra method to two graph analytic problems: triangle enumeration and jaccard. We demonstrate the performance of our SpGEMM-based approaches on various architectures w.r.t. architecture- and problem-specific methods.

Mehmet Deveci, Michael M. Wolf, Jonathan W. Berry,  
Siva Rajamanickam  
Sandia National Laboratories  
mndevec@sandia.gov, mmwolf@sandia.gov,  
jberry@sandia.gov, srajama@sandia.gov

Erik G. Boman  
Center for Computing Research  
Sandia National Labs  
egboman@sandia.gov

Christian Trott  
Sandia National Laboratories  
crtrott@sandia.gov

Neil Butcher  
University of Notre Dame  
nbutcher@nd.edu



Simon D. Hammond  
 Scalable Computer Architectures  
 Sandia National Laboratories  
 sdhammo@sandia.gov

Stephen Olivier  
 Sandia National Laboratories  
 slolivi@sandia.gov

## MS90

### Performance of a Scalable Programming Environment for Graph Algorithms

Analytic workflows are comprised of diverse methods with different processing, memory, and synchronization requirements. Moreover, data types are heterogeneous and can be either dense or sparse with and without structure. These facts make it impossible to design a homogeneous, extreme scale system that can provide performance and scalability throughout the workflow. The introduction of heterogeneity increases significantly development costs. In this talk, I will present HAGGLE, a hybrid, multithreaded analytic platform that we are developing.

John T. Feo  
 Pacific Northwest Laboratory  
 john.feo@pnl.gov

## MS90

### Asynchronous, One-Sided Communication: Programming Support for Distributed-Memory Scientific Applications

Decoupling communication and synchronization is beneficial in parallel scientific computing, including graph analytics. Meanwhile, most scientific codes use MPI and need convenient features to support one-sided communication, or Remote Memory Access (RMA), models. We present RMACXX, a set of C++ bindings to MPI-3 RMA, to ease the use of MPI RMA. Unlike PGAS (Partitioned Global Address Space) models, which may have interoperability issues with MPI, RMACXX is written on top of MPI and uses the same runtime as MPI. It provides an intuitive API for building scientific applications while maintaining performance close to handwritten MPI-3 RMA codes. We demonstrate its efficacy using various microbenchmarks and applications, including a graph analytics example.

Assefaw Gebremedhin, Sayan Ghosh  
 Washington State University  
 assefaw@eecs.wsu.edu, sghosh1@eecs.wsu.edu

Yanfei Guo, Pavan Balaji  
 Argonne National Laboratory  
 yguo@anl.gov, balaji@mcs.anl.gov

## MS90

### Balanced Partition Refinement with the Graph P-Laplacian

Balanced graph partitioning is a fundamental and ubiquitous technique in scientific computing. The classical spectral approach generates an approximate solution from the second (Fiedler) eigenvector of the graph Laplacian matrix, but computation of this vector is considered too costly for large-scale application. Algorithms that perform recursive coarsening and local refinement to address this complexity eventually superseded it both in terms of partition quality

and computational efficiency. The Fiedler vector is minimiser of the Rayleigh quotient of the graph Laplacian in the 2-norm. Minimisation in the p-norm leads to the second eigenvector of the nonlinear graph p-Laplacian operator. It has been shown to provide a sharp approximation to the Ratio Cheeger Cut metric but again, the complexity of initialisation from the Fiedler eigenvector, and of the constrained nonlinear optimisation problem that results, make it unsuitable for large graphs. We initialise the partitioning problem from a synthetic eigenvector generated from METIS multilevel partitioning software and handle the constraint in a novel way. Computational complexity is thus reduced, becoming linear in graph edges and this allows a large-scale investigation in which the p-Laplacian refines a METIS partition, improving quality consistently by 4% for 3D tetrahedral meshes with up to 100 million edges. When tested on benchmark large-scale irregular graphs, separator quality is improved by up to 60%.

Toby Simpson  
 Università della Svizzera italiana  
 toby.simpson@usi.ch

Olaf Schenk  
 USI - Università della Svizzera italiana  
 Institute of Computational Science  
 olaf.schenk@usi.ch

## MS91

### Implementation of Massively Parallel Hyperbolic Multiresolution Methods in the Block-structured Mesh Refinement Framework AMROC

The AMROC framework provides a generic interface for implementing block-structured dynamically adaptive finite volume methods for hyperbolic problems on massively parallel distributed memory systems. Higher-order shock-capturing schemes as well as lattice Boltzmann methods are nowadays available and employ the same block-based mesh refinement algorithm, level-set-based geometry embedding and even fluid-structure coupling routines. While dynamic mesh adaptation for hyperbolic problems usually depends on simplistic refinement indicators, like scaled gradients or heuristic error estimation by Richardson extrapolation, we report in here about the incorporation of a new class of generic wavelet-based mesh adaptation criteria into the AMROC system. Built on the multiresolution principle they promise mesh refinement with defined precision. First comprehensive tests indicate that for similar numerical error mesh savings of 10-20% are typically obtained. In this presentation we will report on details of the implementation and discuss benchmark multiresolution results on several hundreds cores for shock-capturing as well as lattice Boltzmann methods. In particular, we will explain the approach that we have devised to systemically quantify and compare the efficiency of massively parallel AMROC computations using vastly different adaptive meshes.

Ralf Deiterding  
 University of Southampton  
 r.deiterding@soton.ac.uk

Margarete Domingues  
 Instituto Nacional de Pesquisas Espaciais  
 margarete.domingues@inpe.br

## MS91

### Extreme-Scale Block-Structured Adaptive Mesh

**Refinement**

Abstract Not Available At Time Of Publication.

Florian Schornbaum

University of Erlangen-Nuremberg

florian.schornbaum@informatik.uni-erlangen.de

**MS91****An AMR Framework for Realizing Effective High-Resolution Simulations on Multiple GPUs**

Recently grid-based physical simulations with multiple GPUs require effective methods to adapt grid resolution to certain sensitive regions of simulations. In the GPU computation, an adaptive mesh refinement (AMR) method is one of the effective methods to compute certain local regions that demand higher accuracy with higher resolution. The AMR method on the GPU supercomputers is, however, complicated and it is necessary to apply various optimizations suitable for the GPU supercomputers in order to obtain high performance. To develop the applications using the AMR method on the GPU supercomputers effectively, we are developing a block-based AMR framework for grid-based applications written in C++ and CUDA. Programmers just write the stencil functions that update a grid point on Cartesian grid. The framework executes these functions over a tree-based AMR data structure effectively. It also provides the efficient GPU-suitable methods for halo exchange and mesh refinement. In this talk, we will explain the implementation and programming model of the proposed AMR framework, and show some computational results and performance evaluation of AMR applications running on GPU supercomputers.

Takashi Shimokawabe

Global Scientific Information and Computing Center

Tokyo Institute of Technology

shimokawabe@cc.u-tokyo.ac.jp

Takayuki Aoki

Tokyo Institute of Technology

taoki@gsic.titech.ac.jp

Naoyuki Onodera

Center for Computational Science & e-Systems

Japan Atomic Energy Agency

onodera.naoyuki@jaea.go.jp

**MS91****Performance Modeling of Adaptive Mesh Refinement on Modern Supercomputers**

It is relatively easy to model the performance of the steady-state computation of uniform meshes. On the contrary, the speedup promised by AMR is subjected to the dynamic changes in the mesh structure in addition to the overhead of mesh management. The problem of performance analysis is further complicated with choices made by high-level frameworks. To that end, we introduce an analytical and experimental performance model that allows the developers of AMR frameworks to understand the efficiency of AMR, i.e., how close the achieved speedup is to the maximum achievable speedup for a specific application. We show how this performance model was used to expose scalability limitations in neighborhood exchange for the AMR codes generated by our AMR framework, Daino.

Mohamed Wahib

RIKEN, Japan

mohamed.attia@riken.jp

**MS92****Integer Matrix Approximation and Its Applications**

The integer low-rank approximation of integer matrices has received attention recently due to its capacity of naturally representing parts of integer data sets. Different from the general low-rank approximations, the integer approximation is naturally discrete, therefore, the conventional techniques for matrix approximation, such as SVD and non-negative matrix approximation, are inappropriate and unable to solve this problem. To the best of our knowledge, a numerical method for finding a low-rank integer approximation of an integer matrix has not been proposed in the literature earlier. In this talk, we want to propose a block coordinate descent method to obtain the integer low-rank approximation of integer matrices. This method consists of recursively finding integer solutions of integer least square problems. Applications on the real world problems such as the market basket transactions, association rule mining, cluster analysis, and pattern extraction will be given. Numerically, we show that our method can find a more accurate solution than any other existing methods designed for continuous data sets.

Matthew M. Lin

Department of Mathematics

National Cheng Kung University

mhlin@mail.ncku.edu.tw

**MS92****Parallel Butterfly-Based Direct Solvers for Highly Oscillatory Problems**

Fast direct integral equation (IE) solvers for oscillatory problems governed by the Helmholtz and Maxwells equations constitute an active research area. Existing direct solvers leveraging hierarchical (semi-separable) matrices rely on low-rank (LR) compressions of off-diagonal blocks in the discretized IE operator and its inverse to achieve low memory and CPU complexities. These solvers are quite successful in low-frequency regimes, however, they fail to provide reduced-complexity solutions in high-frequency regimes due to the lack of LR compressibility of high-frequency IE blocks. Here I present a new class of direct IE solvers based on so called butterfly schemes to reduce memory and CPU complexities. The proposed solvers construct hierarchical (semi-separable) matrices with pertinent blocks represented as butterfly factorizations via fast randomized schemes. The CPU and memory complexities of these solvers are estimated and numerically validated to be  $\mathcal{O}(N \log^2 N)$  and  $\mathcal{O}(N \log^{1.5} N)$ , respectively. In addition, a distributed-memory parallelization scheme that partitions the workload at both the matrix and block levels are presented.

Yang Liu

Computational Research Division, Lawrence Berkeley

National

liuyangzhuan@lbl.gov

**MS92****Parallelization of the IFMM-based Preconditioner**

### for 3D Helmholtz BEM

We will present a parallel algorithm of the inverse fast multipole method (IFMM), which is a fast direct solver for dense linear systems usually obtained from integral equations in classical physics. The IFMM realizes the low computational complexity by (i) extending the underlying dense system with the relationships of the fast multipole method (FMM) and then (ii) factorizing the resulting larger but sparse system by means of redirecting new fill-ins, which correspond to well-separated interactions, to the existing fill-ins through a low-rank approximation. To parallelize the IFMM, we focus on the parallelism with respect to nodes of the FMM hierarchy (tree structure). From the serial factorization algorithm of the IFMM, we can say that two nodes (at a level of the hierarchy) can be processed concurrently if the distance between them is four or more times of their size. In this grain of parallelization, each of concurrent nodes can process multiple matrix-matrix products (BLAS 3 operations), whose arithmetic intensity is generally high. Therefore, we can expect a high parallel efficiency unless the memory bandwidth is limited. The parallel algorithm can be implemented with OpenMP. We apply the code to accelerate the IFMM-based preconditioner to solve the boundary integral equation for the 3D Helmholtz equation, which was investigated in our previous work. We will show the measured speedup and analyze the result.

Toru Takahashi  
Nagoya University  
ttaka@nuem.nagoya-u.ac.jp

Chao Chen  
Stanford University  
cchen10@stanford.edu

Eric F. Darve  
Stanford University  
Mechanical Engineering Department  
darve@stanford.edu

### MS92

#### Asynchronous Communication in Hierarchical Low-Rank Solvers

Hierarchical low-rank approximation methods combine the robustness of a direct solver with the efficiency of approximate solvers. They were originally developed for dense matrices and remain to be more competitive for such problems, including Schur complements in sparse direct solvers. Recently, there has been a lot of work on hierarchical low-rank approximation methods, i.e. H-matrix, HSS, and recursive skeletonization, but the work has been limited to serial computation or shared memory parallelism with the exception of a very few cases. The current work focuses on the distributed memory parallelism for hierarchical low-rank solvers, and how asynchronous communication can benefit these problems. Tests are run on the TSUB-AME 3.0 supercomputer with 4 GPUs per node. The bulk synchronous variants are compared with the asynchronous methods, and the breakdown of the communication time, load balance, and idle time are analyzed for a few different cases. We considered various configurations for the intra- & inter-node communications along with low-latency RDMA.

Rio Yokota  
Tokyo Institute of Technology

rioyokota@gsic.titech.ac.jp

### MS93

#### Efficient Monte Carlo Transport Methods on Manycore Architecture

The classical Monte Carlo (MC) neutron transport employs energy lookup on long tables to compute the cross sections needed for the simulation. This process has been identified as an important performance hotspot of MC simulations, because poor cache utilization caused by random access patterns and large memory footprint makes it unfriendly to modern architectures. We will present different solutions to overcome this problem, especially on many core architectures. Two different approaches have been studied. The 1st one consists in optimizing energy lookup algorithms when data have been pre-computed and stored in memory. The 2nd one consists in reconstructing cross-section on the fly. The reconstruction converts the problem from memory-bound to compute-bound. These solutions have been analyzed and compared through 3 main axes: performances, memory footprint and energy consumption.

Christophe Calvin  
CEA/DRF  
christophe.calvin@cea.fr

Yunsong Wang  
CEA  
yunsong.wang@cea.fr

Emeric Brun, Fausot Malvagi  
CEA/DEN  
emerik.brun@cea.fr, fausto.malvagi@cea.fr

### MS93

#### Efficient Load Balancing for High Performance Multi-Physics Computations

High performance computing on large scale architectures allow running complex multi-physics simulation codes. However, to achieve a good parallel efficiency is complicated, mainly due to the difficulty to keep a balance workload on each computing unit. We will present our work in Scotch to efficiently load balance mesh based multi-physics simulations by solving multi-criteria graph partitioning problems. We will present algorithms, highlighting differences with concurrent approaches such as Metis and PaToH. We will also show results on parallel applications, requiring static and dynamic load balancing.

Cédric Chevalier  
French Alternative Energies & Atomic Energy  
Commission (CEA)  
cedric.chevalier@cea.fr

Remi Barat  
CEA/DAM  
baratr@ocrea.cea.fr

Francois Pellegrini  
University of Bordeaux  
francois.pellegrini@labri.fr

### MS93

#### Fast Linear Solvers Exploiting Hierarchical Matrix

## Structures and Hardware Organization

As current and future architectures become increasingly hierarchical and offer abundant intra-node fine-grained parallelism, developing high-performance applications become more onerous. We present our recent research results of new algorithm development for factorization based sparse solvers and code optimizations targeted for the new architectures. We use OpenMP's task and nested parallel constructs to reduce load imbalance and increase parallelism for sparse computations. We restructure the code to increase vectorization for both computations and scatter/gather operations. We exploit hierarchical matrix structures to reduce communication cost, and employ asynchronous broadcast to reduce process idle time. We will quantitatively assess the performance effects of these techniques on the new manycore machines, such as Intel KNL.

X. Sherry Li  
Lawrence Berkeley National Laboratory  
xsli@lbl.gov

Gustavo Chavez  
Computational Research Division, Lawrence Berkeley National  
gichavez@lbl.gov

Pieter Ghysels  
Lawrence Berkeley National Laboratory  
pghysels@lbl.gov

Chris Gorman  
UC Santa Barbara  
gorman@math.ucsb.edu

Yang Liu  
Computational Research Division, Lawrence Berkeley National  
liuyangzhuan@lbl.gov

## MS93

### Efficient Sparse General Matrix-Matrix Multiplication Algorithms for Many-Core Processors

Sparse general matrix-matrix multiplication (SpGEMM) is one of the key kernels of preconditioners such as algebraic multigrid method or graph algorithms. However, the performance of SpGEMM is quite low on modern processors due to random memory access to both input and output matrices. Efforts for improving locality are typically hindered by the absence of compile time information about the number and the pattern of non-zero elements in the output matrix. Moreover, the state-of-the-art implementations of SpGEMM require large amounts of memory for temporary results, diminishing cache locality and limiting the matrix size computable on fast GPU device memory, or MCDRAM of Intel KNL architecture. We propose a new fast SpGEMM algorithm that requires small amount of memory and achieves high performance. Calculation of the pattern and value in output matrix is optimized using a hash table with limited index range. Performance evaluation using matrices from the Sparse Matrix Collection of the University of Florida and synthetic matrices for graph processing shows that our approach achieves significant speedups compared to existing SpGEMM libraries on both NVIDIA's Pascal generation GPU and Intel KNL

architectures.

Yusuke Nagasaka  
Tokyo Tech  
Tokyo, Japan  
nagasaka.y.aa@m.titech.ac.jp

Akira Nukada  
Tokyo Institute of Technology  
nukada@smg.is.titech.ac.jp

Ariful Azad, Aydin Buluc  
Lawrence Berkeley National Laboratory  
azad@lbl.gov, abuluc@lbl.gov

Satoshi Matsuoka  
Tokyo Institute of Technology  
matsu@is.titech.ac.jp

## MS94

### Towards Petascale Genome Analytics

Genome Analytics is widely recognized to be one of the original data analytics problems. The width of problems that genome analytics tries to address, ranging from graph algorithms to I/O algorithms to irregular memory access problems, has baffled researchers for years. With the recent push in scaling genome analytics to petascale volumes (with respect to the size of data), however, a new set of problems have emerged that need to be urgently addressed. In this talk, I will discuss some of the recent advances in genome analytics and a vision and path forward towards reaching petascale genome analytics by the time the first Exascale supercomputers are deployed.

Pavan Balaji  
Argonne National Laboratory  
balaji@mcs.anl.gov

## MS94

### Using Workflows to Automate Big Data Processing

Physical sciences have been dealing with the collection, management, and processing of Big Data for decades. The key to success to discoveries in fields such as earthquake science, gravitational-wave physics, and others is automation of the data processes. Although some projects have developed project-specific tools, others have relied on workflow technologies that are domain agnostic. This talk focuses on the Pegasus Workflow Management System that is designed to support a broad range of sciences and that aims to improve the performance of workflows through a number of techniques. The talk will describe these solutions and demonstrate their impact on various types of workflows.

Ewa Deelman  
Information Science Institute  
University of Southern California  
deelman@isi.edu

## MS94

### Experiences with In-Situ Analytics at Extreme Scale

Data-related challenges are quickly dominating computational and data-enabled sciences, and threaten to limit the potential impact of emerging extreme scale computing environments. Data staging and in-situ/in-transit data ana-

lytics are emerging as attractive approaches for addressing these challenges and extreme scale scientific workflows. In this talk I will explore solutions based on the DataSpaces data staging service, and presents experiences in supporting in-situ/in-transit data analytics for extreme scale application workflows in combustion and fusion.

Manish Parashar

Department of Computer Science  
Rutgers University  
parashar@rutgers.edu

#### MS94

##### **Transitioning Data Analytics of Md Simulations Toward the Exascale Era**

Molecular dynamics (MD) simulations studying the classical time evolution of a molecular system at atomic resolution are widely recognized in the fields of chemistry, material sciences, molecular biology and drug design; these simulations are one of the most common simulations on supercomputers. Next-generation supercomputers will have dramatically higher performance than do current systems, generating more data that needs to be analyzed (i.e., in terms of number and length of molecular dynamics trajectories). The coordination of data generation and analysis cannot rely on manual, centralized approaches as it is predominately done today. In this talk I will discuss how the combination of machine learning and data analytics approaches, workflow management methods, and high performance computing techniques can transition the runtime analysis of larger and larger MD trajectories towards the exascale era.

Michela Taufer

University of Delaware  
taufer@udel.edu

#### MS95

##### **Matrix Multiplication Based Algorithm for Inverse Eigenvalue Problems and Its Quadratic Convergence**

Inverse eigenvalue problems arise in a variety of applications. The additive inverse symmetric eigenvalue problem is a typical example, which arises in inverse Sturm-Liouville problems, inverse vibration problems, nuclear spectroscopy, signal processing, and so forth. In this talk, we focus on numerical algorithms for solving the above inverse eigenvalue problems. There are various Newton-like methods for solving the inverse eigenvalue problems, where the quadratic convergence is theoretically guaranteed. With such a background, we propose a new quadratically convergent algorithm based on simple matrix equations. This algorithm has the following nice features. First, the proposed algorithm primarily comprises matrix multiplications, though a standard Newton's method requires for solving an eigenvalue problem in each iteration. Second, although the proposed algorithm is similar to the Cayley transform method, the proposed algorithm does not require the Cayley transform, which reduces the operations in each iteration. Third, our idea can be extended to inverse generalized eigenvalue problems, while the Cayley transform method cannot be applied to such problems. Moreover, the proposed algorithm is applied to multiple eigenvalues. For an arbitrary set of prescribed eigenvalues, the quadratic convergence is ensured under a mild condition.

Kensuke Aishima

Graduate School of Information Science and Technology,  
University of Tokyo  
Kensuke\_Aishima@mist.i.u-tokyo.ac.jp

#### MS95

##### **A Nonlinear Semi-NMF Based Method and Its Parallel Implementation for Deep Neural Networks**

For computing weight matrices of deep neural networks (DNNs), the backpropagation (BP) method has been widely used as a de-facto standard algorithm. As another type of algorithm for computing weight matrices, in this talk, we propose an alternating optimization method using linear and nonlinear semi-nonnegative matrix factorizations (semi-NMFs) and its parallel implementation. We also evaluate the performance compared with the BP method for fully-connected DNNs.

Akira Imakura

Department of Computer Science  
University of Tsukuba  
imakura@cs.tsukuba.ac.jp

Ryosuke Arai

University of Tsukuba  
arai@mma.cs.tsukuba.ac.jp

Tetsuya Sakurai

Department of Computer Science  
University of Tsukuba  
sakurai@cs.tsukuba.ac.jp

#### MS95

##### **A New High Performance and Scalable SVD Algorithm on Distributed Memory Systems**

Computing the SVD of a matrix is an important problem in scientific computing and many applications such as information retrieval, PCA and signal processing, etc. In order to computing SVD fast on supercomputers, we implement and test some algorithms on Tianhe2, one of the fastest computers in the world. We find the polar decomposition (PD) recently exhibits much advantage in computing the singular value decomposition (SVD). This work introduces a high performance implementation of Zolo-SVD algorithm on the distributed memory systems, which is based on the PD algorithm via the Zolotarev's function (Zolo-PD), originally proposed by Nakatsukasa and Freund [SIAM Review, 2016]. Our implementation highly relies on the routines of ScaLAPACK and therefore it is portable. Comparing with other PD algorithms such as the QR-based dynamically weighted Halley method (QDWH-PD), Zolo-PD is naturally parallelizable and has better scalability though performs more floating-point operations. When using more processes, Zolo-PD can be up to three times faster than QDWH-PD algorithm, and Zolo-SVD can be two times faster than the ScaLAPACK routine PDGESVD. These numerical experiments are performed on Tianhe2 supercomputer, and the tested matrices include some sparse matrices from particular applications and some randomly generated dense matrices with different dimensions.

Shengguo LI

National University of Defense Technology  
China  
nudtli@nudt.edu.cn

Jie Liu, Xinbiao Gan

National University of Defense Technology

liujie@nudt.edu.cn, xinbiaogan@163.com

## MS95

### Fast Solution of Nonnegative Matrix Factorization Via a Matrix-Based Active Set Method

For the solution of nonnegative matrix factorization (NMF), which is a low rank matrix approximation problem with nonnegative constraints, we propose a new alternating nonnegative least squares (ANLS) method by utilizing modulus-type inner outer iteration method to solve the nonnegative constrained least squares problem with multiple right hand sides at each iteration. Theoretical convergence that the limit of the sequence generated by the proposed method is a stationary point of NMF can be guaranteed. We also propose a new matrix-based active set method which utilize the componentwise operation in the inner iterations to improve the convergence performance. Finally, we show that the proposed methods can be extended to solve the sparse NMF and regularized NMF. Numerical experiments on the synthetic data, face image data and text data show that the proposed methods converges faster than the state-of-the-art methods and active set methods.

Ning Zheng

SOKENDAI (The Graduate University for Advanced Studies)  
nzheng@nii.ac.jp

Ken Hayami  
National Institute of Informatics  
hayami@nii.ac.jp

Nobutaka Ono  
National Institute of Informatics  
SOKENDAI (The Graduate University for Advanced Studies)  
onono@nii.ac.jp

## MS96

### Towards Large Scale Unfitted Adaptive Finite Element Simulations

The use of unfitted finite element methods (FEMs) is an appealing approach for different reasons. They are interesting in coupled problems or to avoid the generation of body-fitted meshes. One of the bottlenecks of the simulation pipeline is the body-fitted mesh generation step and the unstructured mesh partition. The use of unfitted methods on background octree Cartesian meshes avoids the need to define body-fitted meshes, and can exploit efficient and scalable space-filling curve algorithms. In turn, such schemes complicate the numerical integration, imposition of Dirichlet boundary conditions, and the linear solver phase. The condition number of the resulting linear system does depend on the characteristic size of the cut elements, the so-called small cut cell problem. In this work, we will present a parallel unfitted framework that relies on adaptive octree background meshes and space-filling curve partitioners. In order to solve the small cut cell problem, we will pursue two different lines. The first one is a re-definition of the finite element spaces that solves this issue, leading to condition number bounds as the ones for body-fitted schemes without any kind of perturbation/stabilization of the Galerkin formulation. Another approach will be to define appropriate iterative linear solvers based on domain decomposition preconditioning that are robust with respect

to the small cut cell problem.

Santiago Badia

Universitat Politècnica de Catalunya, Spain  
sbadia@cimne.upc.edu

Alberto Martin

International Center for Numerical Methods in Engineering,  
Universitat Politècnica de Catalunya, Barcelona, Spain  
amartin@cimne.upc.edu

Francesc Verdugo

CIMNE  
fverdugo@cimne.upc.edu

## MS96

### Performance Tuning of DDM Subdomain Local FE Solvers for Many-Core Architectures

Exa-scale supercomputers will appear around 2020, 2021, 2022. To obtain high intra node performance, efficient utilization of processor cache memory should be considered. The traditional memory access-intensive approach, which prefers less computing and more storage on main memory, might not be effective for supercomputers in near future. The Domain Decomposition Method (DDM) is one of the effective parallel finite element schemes. We have been developing an FE-based parallel structural analysis code, ADVENTURE Solid, based on DDM, with the Balancing Domain Decomposition (BDD) pre-conditioners. The re-design of the subdomain local FE solver part, which is a performance sensitive kernel in the DDM code, is required. Here in this work, an on-cache iterative solver based on the DDM framework is developed. The subdomain local FE solver of the DDM code is implemented using CG solvers with various types of pre-conditioners, such as diagonal scaling, SSOR and ICT. These iterative solvers are parallelized using OpenMP, so that each subdomain can be solved by multiple cores. By adjusting the subdomain size so that the footprint fits within the last-level cache of a processor, this DDM code can be considered as a kind of an on-cache iterative solver. Performance benchmark results are shown on various kinds of HPC platform, such as Skylake Xeon, Knights Landing and Fujitsu PRIMEHPC FX100.

Hiroshi Kawai

Tokyo University of Science-Suwa  
kawai063@toyo.jp

Masao Ogino  
Nagoya University  
masao.ogino@cc.nagoya-u.ac.jp

Ryuji Shioya  
Toyo University  
shioya@toyo.jp

Tomonori Yamada, Shinobu Yoshimura  
The University of Tokyo  
yamada@race.u-tokyo.ac.jp, yoshi@sys.t.u-tokyo.ac.jp

## MS96

### New Aspects on Nonlinear Domain Decomposition Solvers on the Extreme Scale

Parallel Newton-Krylov domain decomposition methods



(Newton-Krylov-DDM) are fast and robust solvers, e.g., for nonlinear implicit problems in solid and fluid mechanics. Here, the nonlinear problem is first linearized and then each linearized problem is decomposed, i.e., is solved by a Krylov space method using a DD preconditioner. In nonlinear domain decomposition methods (NL-DDM), by changing the order of these operations, first the nonlinear problem is decomposed and then the nonlinearly decomposed problem is linearized. This allows for the design of new parallel nonlinear FETI-DP and BDDC domain decomposition methods with increased locality and reduced communication. Such NL-DDM have been successful in reducing the time to solution for different nonlinear problems compared to Newton-Krylov-DDM. Since the NL-DDM are based on a decomposition of the global nonlinear problem into many local nonlinear problems, strongly local nonlinearities can be resolved by a small number of local nonlinear problems or computational cores, respectively. The remaining cores which are not busy with those local, strongly nonlinear effects have the potential to wait and save energy, speed up the computations of the busy cores, or serve to recover subdomain problems on faulty cores. In classical Newton-Krylov-DDM, this is not possible since all nonlinear effects, whether local or global, interfere with the global convergence of Newton's method and keep all cores busy all the time.

Axel Klawonn  
Universitaet zu Koeln  
axel.klawonn@uni-koeln.de

Martin Lanser  
Universitaet zu Koeln  
Mathematisches Institut  
martin.lanser@uni-koeln.de

Oliver Rheinbach  
Technische Universität Bergakademie Freiberg  
oliver.rheinbach@math.tu-freiberg.de

## MS96

### Large-Scale Nonlinear Solid Mechanics Analysis Using Domain Decomposition Method Based on Nonlinear Solver

To discretize partial differential equations of solid mechanics, finite element method is frequently used. When the discretized problem is large-scale and nonlinear, a scalable nonlinear solver should be adopted. In this study, a domain decomposition method that is based on a nonlinear solver such as a quasi-Newton method is presented. To hold the scalability, balancing domain decomposition (BDD) preconditioner, which is one of multi-grid methods, is applied to the present nonlinear solver. The BDD preconditioner was proposed and has generally been used for linear solvers. In the presentation, the methodology is explained in comparison with a conventional domain decomposition method, followed by a benchmark analysis to test the weak scaling.

Yasunori Yusa, Hiroshi Okada  
Tokyo University of Science  
yyusa@rs.tus.ac.jp, hokada@rs.noda.tus.ac.jp

Tomonori Yamada, Shinobu Yoshimura  
The University of Tokyo

yamada@race.u-tokyo.ac.jp, yoshi@sys.t.u-tokyo.ac.jp

## MS97

### Sparse Tensor Decomposition on EMU Platform

Analysis of data represented as sparse tensors poses many challenges for typical computer architectures. Data locality appears in short bursts but overall the analysis algorithm jumps around the data structure. The EMU platform strives to alleviate these issues by migrating lightweight threads to the data directly and emphasising fine-grained memory accesses. We analyse the effects of such an architecture to accelerate the CANDECOMP/PARAFAC algorithm popularly used to analyse multi-linear data models.

Srinivas Eswar  
Computational Science & Engineering  
Georgia Tech  
srinivas.eswar@gatech.edu

Jiajia Li  
College of Computing  
Georgia Institute of Technology  
jiajiali@gatech.edu

Richard Vuduc  
Georgia Institute of Technology  
richie@cc.gatech.edu

Patrick Lavin  
Computational Science & Engineering  
Georgia Institute of Technology  
plavin3@gatech.edu

Young Jeffrey  
School of Computer Science  
Georgia Institute of Technology  
jyoung9@gatech.edu

## MS97

### Design and Implementation of Deep Learning Kernel Library on APU

Deep Learning (DL) is booming in various areas, like computer vision, natural language processing and scientific computing. All DL algorithms are composed of training step and inference step. Training step includes forward propagation (FP) and back propagation (BP), which are both time and energy consuming processes, and tends to run on cluster servers. Energy efficient APU acts as a potential rival to NVIDIA GPU in future exascale supercomputer system. On the other hand, which seems to be a compute intensive DL algorithm on multi-GPUs platform presents to be a communication intensive algorithm when it is deployed on a distributed environment. Our work aims to design and implement a set of DL interfaces on a distributed environment with APU. Not only auto tuning network awareness kernels are developed to improve the performance on Vega10 architecture, but also communication avoiding BP are utilized to explore the performance of exascale platform.

Haoqiang Guo  
Institute of Computing Technology  
Chinese Academy of Sciences

guohaoqiang@ncic.ac.cn

#### MS97

##### Accelerating Scientific Kernels on Energy-Efficient FPGAs

Iterative stencil algorithms find applications in a wide range of domains. FPGAs have long been adopted for computation acceleration due to its advantages of dedicated hardware design. However, efficient implementation of iterative stencil algorithms on FPGAs is very challenging due to the data dependencies between iterations and elements in the stencil algorithms, programming hurdle of FPGAs, and large design space. In this talk, I will introduce a comprehensive framework that synthesizes iterative stencil algorithms on FPGAs efficiently. We leverage the OpenCL-to-FPGA toolchain to generate accelerator automatically and perform design space exploration at the high level. We propose to bridge the neighboring tiles through pipe and enable data sharing among them to improve computation efficiency. We also develop analytical performance models to explore the complex design space. Experiments using a wide range of stencil applications demonstrate that on average our heterogeneous implementations achieve 1.65X performance speedup but with less hardware resource compared to the state-of-the-art.

Yun (Eric) Liang

School of Electronics Engineering and Computer Science  
Peking University  
ericlyun@pku.edu.cn

#### MS97

##### Exploring and Analyzing the Real Impact of Modern On-Package Memory on HPC Scientific Kernels

High-bandwidth On-Package Memory (OPM) innovates the conventional memory hierarchy by augmenting a new on-package layer between classic on-chip cache and off-chip DRAM. Due to its relative location and capacity, OPM is often used as a new type of LLC. Despite the adaptation in modern processors, the performance and power impact of OPM on HPC applications, especially scientific kernels, is still unknown. In this paper, we fill this gap by conducting a comprehensive evaluation for a wide spectrum of scientific kernels with a large amount of representative inputs, including dense, sparse and medium, on two Intel OPMs: eDRAM on multicore Broadwell and MCDRAM on many-core Knights Landing. Overall, we demonstrate that the applications' memory footprint size plays an important role in determining the effectiveness of OPM usage, while other factors such as power and algorithm features should also be taken into consideration as they may increase costs or even degrade performance.

Shuaiwen Song

Pacific Northwest National Laboratory, USA  
shuaiwen.song@pnnl.gov

#### MS98

##### An Approach to Accelerating the SpMV Kernel by Exploiting Specific Sparse Structures

We aim at accelerating the sparse matrix-vector (SpMV) kernel, which is one of important building blocks of iterative solvers. We consider exploiting certain sparse structures that are beneficial for improving the performance

of the SpMV kernel (e.g. avoiding indirect memory access). We construct a mechanism in which specified (partial) sparse structures in an input matrix are automatically extracted and stored in a hybrid storage format. We examine the effectiveness of our approach on a recent multicore CPU.

Takeshi Fukaya, Koki Ishida, Akie Miura, Takeshi Iwashita

Hokkaido University  
fukaya@iic.hokudai.ac.jp, koki-skater@eis.hokudai.ac.jp,  
shiro125@eis.hokudai.ac.jp, iwashita@iic.hokudai.ac.jp

Hiroshi Nakashima  
ACCMS

Kyoto University  
h.nakashima@media.kyoto-u.ac.jp

#### MS98

##### Solving the $k$ -th Eigenvalue Problem in Large-Scale Electronic Structure Calculations

In this talk, we consider computing the  $k$ -th eigenvalue and its corresponding eigenvector of large sparse generalized Hermitian eigenvalue problems. In electronic structure calculations, several material properties are governed by an eigenpair with a material-specific index  $k$ . Building on spectral bisection, we present a framework for computing the  $k$ -th eigenpair with validation of its index. Numerical results demonstrate the efficiency and accuracy of the proposed framework.

Dongjin Lee

Nagoya University  
dongjin-i@na.nuap.nagoya-u.ac.jp

Takeo Hoshi

Tottori University  
hoshi@damp.tottori-u.ac.jp

Tomohiro Sogabe, Yuto Miyatake, Shao-Liang Zhang  
Nagoya University  
sogabe@na.nuap.nagoya-u.ac.jp,  
miyatake@na.nuap.nagoya-u.ac.jp,  
zhang@na.nuap.nagoya-u.ac.jp

#### MS98

##### Resilience for Fine-Grained Preconditioning Methods

We investigate the use of various mechanisms for improving the resilience of the fine-grained parallel preconditioning, such as those for computing an incomplete LU factorization, proposed in [Chow and Patel, Fine-grained parallel incomplete LU factorization, SIAM J. on Scientific Computing, 37(2), pp 169–193, 2015]. These include several checkpointing variants as well as a self-stabilizing periodic correction of an algorithm. Results concerning convergence of all of the self-stabilizing variants of the algorithm with respect to the occurrence of faults, and the impact of any sub-optimality in the produced incomplete L and U factors in Krylov subspace solvers are given. Numerical tests show that the simple algorithmic changes suggested here can ensure convergence of the fine-grained parallel incomplete factorization, and improve the performance of the resulting factors as preconditioners in Krylov subspace solvers in the presence of transient soft faults.

Masha Sosonkina, Evan Coleman

Old Dominion University  
msosonki@odu.edu, ecole028@odu.edu

### MS98

#### Adaptive Multipreconditioned Domain Decomposition Methods

Domain Decomposition methods are a family of solvers designed for very large linear systems that require parallel computers. They proceed by splitting the computational domain into subdomains and then approximating the inverse of the original problem with local inverses coming from the subdomains. I will present a classical domain decomposition method and show that for realistic simulations (with heterogeneous materials for instance) convergence usually becomes very slow. Then I will explain how this can be fixed by injecting more information into the solver. In particular I will show how using multiple search directions within the conjugate gradient algorithm makes the algorithm more reliable. Efficiency is also taken into account since our solvers are adaptive.

Nicole Spillane  
Centre de Mathématiques Appliquées  
Ecole Polytechnique  
nicole.spillane@cmap.polytechnique.fr

Christophe Bovet  
LMT, ENS Cachan, CNRS, Université Paris-Saclay,  
94235 Cachan  
christophe.bovet@lmt.ens-cachan.fr

Augustin Parret-Fréaud  
Safran Tech, Rue des Jeunes Bois, Châteaufort  
CS 80112 78772 Magny-Les-Hameaux  
augustin.parret-freaud@safrangroup.com

Pierre Gosselet  
LMT, ENS Cachan  
gosselet@lmt.ens-cachan.fr

### MS99

#### First-Principles Investigation of Thermal Transport Property of Earth's Lowermost Mantle

Lattice thermal conductivity,  $\kappa_{\text{lat}}$ , is a fundamental physical parameter in controlling the activity of heat transfer in a solid. It attracts researchers from wide ranging fields such as materials science, engineering, and Earth and planetary science. The determination of  $\kappa_{\text{lat}}$  of Earth's constituent materials is key to understanding the dynamics and thermal evolution of the Earth's interior. Recently, I established an accurate computational scheme to predict  $\kappa_{\text{lat}}$  at any  $P$  and  $T$  conditions based on the density functional theory (DFT) and the many body quantum perturbation theory for the phonon-phonon interaction in Boltzmann transport theory. The technique was applied to end-member of major LM minerals,  $\text{MgSiO}_3$  bridgmanite (Dekura+2013PRL), and  $\text{MgO}$  periclase (Dekura+2017PRB). Next, I extended our techniques to more realistic Fe-bearing systems in conjunction with the internally consistent DFT+ $U$  method (Wang+2015NatureGeosci) to deal with the electronic structures of such strongly-correlated systems. In this presentation, I introduce the current situation of the research on  $\kappa_{\text{lat}}$  and thermal transport property of the core-mantle boundary region.

Haruhiko Dekura

Ehime University  
dekura.haruhiko.mf@ehime-u.ac.jp

### MS99

#### Solar Convection Simulation with Reduced Speed of Sound Technique

The sun has sunspots and the number of sunspot has 11-year cycle. The complicated interaction between the turbulence and magnetic field maintains the magnetic cycle. Thus, precise understanding of the turbulence from large-scale to small-scale is important to understand generation and maintenance mechanism of the magnetic field. To this end, we need to carry out large-scale simulations of the solar convection zone. In the solar convection zone, the speed of sound is much faster than convection velocity, i.e., velocity of turbulence. Thus, the speed of sound severely restricts the time step through the CFL condition. In order to avoid this severe restriction, anelastic approximation in which the speed of sound is assumed infinity is broadly adopted. The anelastic approximation requires frequent global communication due to elliptic equation for the pressure. In order to minimize the communication cost, we newly adopt the Reduced Speed of Technique (RSST) in which the speed of sound is artificially reduced and only local communication is required. In the talk, we present latest results of high resolution calculation with using the RSST. In addition, the RSST allows us to cover regions with low and high Mach number simultaneously. Taking advantage of this, we can cover whole convection zone in one calculation. We also show this new result.

Hideyuki Hotta  
Chiba University  
hotta@chiba-u.ac.jp

### MS99

#### Large Scale Seismic Simulations with SW4

The current and future generation of supercomputers allow seismic simulations with 100,000' s to 1,000,000' s of cores with mixed CPU & GPU architectures. However, direct application of current algorithms on these large machines will not achieve satisfactory performance. In order to move toward exascale computing we are modifying the SW4 seismic simulation code for the latest and future machines. In this talk we report on the porting of SW4 to the Cori-II supercomputer at the National Energy Research Scientific Computing Center (NERSC), located at LBNL. The Cori-II machine has 9,688 nodes, each having a 68-core Intel Xeon Phi processor. While each node on this machine has more cores than previous (Intel based) supercomputers, each individual core is slower. To achieve good computational performance it is thus necessary to make the spatial decomposition finer, i.e., assign fewer grid points to each core. We have implemented a hybrid programming model in SW4 where a MPI-task decomposition in the horizontal directions is augmented by an OpenMP-threaded decomposition in the vertical direction. We report on the performance of the hybrid programming model for some smaller test cases, and then demonstrate how this technique has enabled a significant increase in frequency content for regional ground motion simulations in the San Francisco Bay Area.

Anders Petersson  
Lawrence Livermore National Laboratory  
petersson1@llnl.gov

Bjorn Sjogreen, Arthur Rodgers  
Center for Applied Scientific Computing  
Lawrence Livermore National Laboratory  
sjogreen2@llnl.gov, rodders7@llnl.gov

#### MS99

##### **Global Simulations of Mantle Convection and Plate Tectonics on Massively Parallel Computers using StagYY**

The coupled system of convection of the Earth's solid mantle plate tectonics is the driver of geological change on our planet, including continental drift, volcanoes, earthquakes, crustal production, atmospheric degassing, the recycling, and cooling of the core, which drives the geodynamo. Modelling of this process is challenging due to the wide range of length scales (from faults to continents) and time scales (seconds to billions of years) and the complex rheology of rocks, which exhibit visco-elasto-plastic behaviour with strongly temperature-dependent viscosity varying by orders of magnitude over short length scales. Nevertheless, it is now routine to perform global-scale 3-D spherical simulations that span the 4.5 billion year age of our planet and contain complex effects such as partial melting and crustal production and solid-solid phase transitions. StagYY is one of the leading codes for performing such simulations. It uses a finite-volume discretization on a yin-yang spherical grid. Both built-in geometric multigrid, and the range of solvers available through PETSc, can be used. Here, technical details will be discussed, including recent enhancements made with funding from the Swiss Platform for Advanced Scientific Computing (PASC) project such as use of hybrid (GPU-CPU) architectures. Some recent scientific results published in Science and Nature will then be summarised.

Paul J. Tackley  
ETH Zurich  
ptackley@ethz.ch

Patrick Sanan  
USI - Università della Svizzera italiana  
ICS  
patrick.sanan@gmail.com

#### MS100

##### **Workload Simulations with TraceR-CODES**

Design space exploration and procurement process for next-generation high performance computing systems (HPC) is often guided by the expected performance and cost tradeoffs offered by various alternative options. Networks are a major component of HPC systems, which pose a primary performance concern. The success or usefulness of performance predictions made for future systems is contingent upon several factors such as the accuracy of the prediction methodology, the speed of predictions, and the complexity of workloads being analyzed. These requirements have led to the development of the CODES-TraceR simulation framework. CODES-TraceR addresses the former issues by combining optimistic parallel discrete event simulations with detailed packet-level modeling of many practically relevant topology variants. In order to reproduce real-world scenarios, dynamically configurable simulations of full application traces and synthetic workloads is supported.

Nikhil Jain  
Lawrence Livermore National Laboratory

nikhil@llnl.gov

Misbah Mubarak  
Argonne National Laboratory  
mmubarak@anl.gov

Abhinav Bhatele  
Lawrence Livermore National Laboratory  
bhatele@llnl.gov

Christopher Carothers  
Rensselaer Polytechnic Institute  
chrisc@rpi.cs.edu

Rob Ross  
Argonne National Laboratory  
ross@mcs.anl.gov

#### MS100

##### **Implementation and Evaluation of NSIM-ACE: An Interconnect Simulator for Remote Direct Memory Access**

Remote Direct Memory Access (RDMA) is a technology for access to memory on a remote machine independently of main processors. Network Interface Cards (NICs) used for interconnection networks of high-performance parallel computers are capable of RDMA. RDMA has advantages over conventional communications on networks in latency and memory usage. In addition, RDMA efficiently parallelizes communication using a NIC and computation using a main processor. However, existing network simulators lacked the function of directly handling RDMA. NSIM-ACE is an interconnection network simulator for RDMA. In NSIM-ACE, the target network is simply modeled without significant accuracy loss. NSIM-ACE is constructed on a discrete event simulator, where time step changes dynamically according to subsystems of the target network and their statuses in order to reduce the simulation cost. Thus, the simulator achieves both simulation accuracy and speed. NSIM-ACE accepts flexible simulation settings for targeting a wide range of networks. Target access patterns are written like RDMA-based parallel programs, which makes simulator use easy. By comparing simulation results and real measurements, it was confirmed that the simulator is sufficiently accurate for predicting the better algorithm in RDMA-based parallel programs. The evaluation also showed that simulations of RDMA on large-scale networks are possible in practical simulation times even on a single-processor machine.

Ryutaro Susukita, Yoshiyuki Morie  
Kyushu University  
susukita.ryutaro.491@m.kyushu-u.ac.jp,  
morie.yoshiyuki.404@m.kyushu-u.ac.jp

Takeshi Nanri  
Kyushu University, Japan  
JST CREST, Japan  
nanri@cc.kyushu-u.ac.jp

Hidetomo Shibamura  
Information Technologies and Nanotechnologies

shibamura@isit.or.jp

## MS100

### Simulating MPI Applications: the SMPI Approach

Developers of High Performance Computing applications and runtime systems face many challenges related to performance evaluation, debugging, and optimization on the road to Exascale. These challenges make them strive for controlled and replicable results, access to many different hybrid/large scale platforms (that sometimes do not even exist), and shorter time to conduct large experimentation campaigns. Simulation and emulation are a possible answer to address these needs. In this talk, we will present the recent efforts made within the SimGrid project to target two kinds of HPC applications: classical MPI parallel codes with our simulated SMPI runtime and task based linear algebra applications. Details about the implementation principles, the underlying computing, network, and energy models, as well as encouraging results will be given.

Frédéric Suter  
CC-IN2P3 / CNRS  
frederic.suter@cc.in2p3.fr

Arnaud Legrand  
INRIA Grenoble Rhône-Alpes  
arnaud.legrand@imag.fr

Martin Quinson  
ENS Rennes  
martin.quinson@ens-rennes.fr

## MS100

### Flexible and Scalable Applications Models for the Structural Simulation Toolkit (SST)

The Structural Simulation Toolkit (SST) provides a discrete event simulation framework for architecture design studies ranging from single processor to full system. Full system simulation of distributed (exascale) architectures fundamentally relies on scalable endpoint and network models to be feasible. This scalability relies on a combination of cost-saving approximations and brute-force parallelization of the simulator. While many application endpoint models rely on basic state machine models or off-line traces, the SST macroscale (SST/macro) toolkit aims to provide flexible, on-line simulation of real application code running on actual runtime software stacks. SST/macro therefore strives for not just hardware design but co-design of software algorithms. We review the coarse-grained approximations in SST/macro covering both network models and so-called skeleton applications. We further discuss the parallel discrete event algorithms required to support on-line simulations within SST.

Jeremiah Wilke  
Sandia National Laboratories  
jjwilke@sandia.gov

## MS101

### HPC Benchmarking; Perspective from Power and Sustained Performance

Since HPC systems/architectures and applications executed on them have rapidly diversified, it is getting hard to reproduce performance behaviors of all real scientific applications using only existing benchmark programs. There-

fore, it is necessary to reconsider the metric for evaluating HPC systems. Besides, the future HPC system is required to achieve a certain level of sustained performance with limited power and hardware resources; it is essential to have benchmark programs capable of reproducing the sustained performance of real applications. To this end, co-operation between computational and computer scientists is imperative. In this talk, we look back and discuss on research activities related to HPC benchmarking, especially focusing on power issues and sustained performances of the HPC systems.

Ryusuke Egawa  
Tohoku University  
egawa@tohoku.ac.jp

## MS101

### HPC Benchmarking: Problem Size and Other Pitfalls

Most HPC benchmarks intended for ranking supercomputers adopt very simple setups for their execution. For instance, a single run of HPL with arbitrary problem size is used for the most widely recognized ranking. Similar practise of using a single, or very few, problem sizes is adopted in two other major HPC benchmarks, namely HPCG and HPGMG. We report benchmarking result for all three benchmarks covering a wide range of problem sizes on six distinct hardware architectures. We find that the data holds significantly more information on the performance of the underlying hardware as compared to just the maximum performance observed. We therefore argue that an aggregate value derived from a whole range of problem sizes can significantly improve the sensitivity of a given benchmark to relevant hardware properties. To illustrate the point, we present analysis of the HPGMG benchmark with respect to the influence of the network and the vectorization capabilities of the architecture. The results show a complex pattern of frequently alternating communication and computation phases. The measured communication times depend heavily on latency and message rate of the network. We have further analyzed the influence of vectorization on the benchmarks. HPGMG turns out to be the most adapt for vectorization. Giving small hints to the compiler, we were able to increase vectorization and locality and thus achieved higher fractions of peak performance for HPGMG than on other architectures.

Jose Gracia  
High Performance Computer Center, Stuttgart (HLRS)  
University of Stuttgart  
gracia@hlrs.de

Vladimir Marjanovic  
High Performance Computing Center Stuttgart (HLRS)  
hpcvmrj@hlrs.de

## MS101

### Expectation to Supercomputer Benchmarks from the Viewpoint of Performance Optimization

Modern supercomputers are highly parallel machine combining inter-nodes process parallelism and inter-core thread parallelism. And the memory hierarchy including the cache in the node is also complicated. On the other hand, applications that run on supercomputers cannot fully utilize the performance of hardware unless high parallelization and individual node tuning are performed according to that hardware. In this presentation, we describe the expectations for

benchmark programs from the standpoint of implementing performance optimization such as high parallelization of applications and intra-node tuning.

Kazuo Minami  
RIKEN  
minami\_kaz@riken.jp

#### MS101

##### **Comparing Signatures of Common Supercomputer Benchmarks and Their Utility for the Future**

How to measure the performance of supercomputers adequately has been a topic of research papers and hot debates since the beginning of the field. Many benchmarks formulated for narrow fields of science have come and gone quickly. More common and general benchmarks, which can be used for several decades and across a larger variety of computer architectures, are hard to design effectively. If they are too specific, they will not stand the test of time. If they are too general they might not provide additional discriminative power beyond simple architectural numbers such as peak performance and will not be accepted. The popular supercomputer ranking TOP500 has used the performance of the High Performance Linpack (HPL) for the last 25 years. HPL is probably the longest-lived benchmark in active use. It has often criticized for being too simplistic and for not providing enough differentiating power between different computer architectures. Two new benchmarks have been developed with the goal to provide additional data to the TOP500, the High Performance Conjugate Gradients (HPCG) benchmark and the High-performance Geometric Multigrid (HPGMG) benchmark. In this presentation we will introduce these benchmarks and compare their most important features. We will also address if any of these benchmarks can provide performance measures for new application fields such as deep learning neural networks or help us to compare quantum computing to traditional computing systems.

Erich Strohmaier  
Lawrence Berkeley National Laboratory  
estrohmaier@lbl.gov

#### MS102

##### **TerraNeo - A Finite Element Multigrid Framework for Extreme-Scale Earth Mantle Convection Simulations**

Earth mantle convection simulations require extremely large grids for a sufficient spatial resolution and many time steps to represent geological time scales of millions of years. We present a framework for such large-scale time-dependent simulations on a thick spherical shell with variable viscosity. The nonlinear multiphysics problem of Stokes equation coupled to the energy equation is solved using finite elements. Transient simulation results of the temperature distribution are presented, as well as the stationary flow field for variable temperature-dependent viscosity. Moreover, scaling results show that our approach facilitates solving systems in excess of ten trillion ( $10^{13}$ ) unknowns on Peta-Scale systems using compute times of a few minutes.

Dominik Bartuschat  
Department of Computer Science  
FAU Erlangen-Nuernberg  
dominik.bartuschat@cs.fau.de

Ulrich J. Ruede  
University of Erlangen-Nuremberg  
Department of Computer Science (Simulation)  
ulrich.ruede@fau.de

#### MS102

##### **Simulation Studies on Mantle Dynamics of Terrestrial Planets: Theoretical Backgrounds, Tools and Outcrops**

Numerical simulations of mantle dynamics of terrestrial planets strongly rely on advanced techniques for solution of ill-conditioned elliptic equations stemming from Stokes flow of highly viscous 'solid' rocks. In this presentation, the theoretical backgrounds of mantle convection will be reviewed from mathematical, computational and fluid-dynamical aspects, followed by a brief overview of the research targets and recent outcrops of the simulation studies on mantle dynamics.

Masanori Kameyama  
Geodynamics Research Center (GRC)  
Ehime University, Matsuyama, JAPAN  
kameyama@sci.ehime-u.ac.jp

#### MS102

##### **Thermal Inversion in Subduction Zones**

Seismic observations are insufficient to resolve the full slab shape in the Alaskan subduction zone, necessitating inversion incorporating multi-modal data to determine an accurate slab structure. Advances in technology and measurement techniques provide a rich suite of geophysical databases to constrain the problem. These include sea floor age, surface heat flow, seismic tomography, shear wave splitting, GPS surface velocities, earthquake hypocenters, and petrological temperature constraints. We incorporate many different data modalities in a large scale inverse problem, for which a multilevel formulation and solution strategy is employed. In order to understand the contribution of the many different data modalities to the ultimate posterior distribution, we use a specialized cross-sensitivity measure.

Matthew G. Knepley  
Rice University  
knepley@gmail.com

Margarete Jadamec  
University at Buffalo  
mjadamec@buffalo.edu

Tobin Isaac  
Georgia Tech  
tisaac@cc.gatech.edu

#### MS102

##### **An Extreme-Scale Implicit Solver for Highly Nonlinear and Heterogeneous Flow in Earth's Mantle**

The physics of the Earth's mantle are as fascinating as they are computationally challenging. While being a fundamental geophysical process, enormous knowledge gaps about Earth's mantle convection remain. The reasons: Realistic mantle models pose computational challenges due to highly nonlinear rheologies, severe heterogeneities, anisotropies, and wide ranges of spatial scales. We present new advances on various levels of a nonlinear implicit solver for Earth's



instantaneous mantle flow governed by nonlinear instantaneous Stokes PDEs: (1) Heterogeneity-robust Schur complement preconditioning (weighted BFBT), (2) Hybrid spectral-geometric-algebraic multigrid (HMG), and (3) Nonlinear preconditioning of an inexact Newton-Krylov method (primal-dual Newton). These methods operate on aggressively adapted meshes and mixed continuous-discontinuous discretizations with high-order accuracy. Our goals are the realistic representation of mantle physics, maximizing accuracy and minimizing runtime, and achieving optimal algorithmic performance. While targeting mantle flow problems, the mentioned advances can be applied to many problems at the intersection of computational science & engineering with high-performance & large-scale computing.

Johann Rudi

Institute for Computational Engineering and Sciences  
The University of Texas at Austin  
johann@ices.utexas.edu

Georg Stadler

Courant Institute for Mathematical Sciences  
New York University  
stadler@cims.nyu.edu

Omar Ghattas

The University of Texas at Austin  
omar@ices.utexas.edu

### MS103

#### Security Constrained Optimization of Large Scale Energy Systems on High Performance Computers

The electrical power grid is a critical infrastructure and in addition to economic dispatch, the grid operation should be resilient to failures. Increased penetration of the renewable energy sources is placing greater stress on the grid, shifting operation of the power grid equipment towards their operational limits. Thus, any unexpected contingency could be critical to the overall operation. Consequently, it is essential to operate the grid with a focus on the security measures. Security constrained OPF (SCOPF) imposes additional security constraints to the OPF problem. It aims for minimum adjustments in the pre-contingency operating state, such that in the event of any contingency, the operation will remain secure and within operating limits. For a realistic power network, however, with numerous contingencies considered, the overall problem size becomes intractable for single-core optimization tools in short time frames for real-time industrial operations, such as real-time electricity market responses to electricity prices. We propose distributed interior-point framework exploiting the block-structured KKT linear system arising from the SCOPF problem using a Schur complement technique. In order to utilize a node-level parallelism, an incomplete augmented multicore sparse factorization is used, which further exploits sparse structure of the problem. The performance of the implementation is evaluated on the "Piz Daint" supercomputer.

Juraj Kardos

Institute of Computational Science  
Università della Svizzera italiana  
juraj.kardos@usi.ch

Drosos Kourounis

USI - Università della Svizzera italiana  
Institute of Computational Science

drosos.kourounis@usi.ch

Olaf Schenk

USI - Università della Svizzera italiana  
Institute of Computational Science  
olaf.schenk@usi.ch

### MS103

#### Solving Steiner Tree Problems and Variants to Optimality on a Supercomputer

The Steiner tree problem in graphs is a classical combinatorial optimization problem that commonly arises in practical applications as one of many variants. Despite the strong relationship between the different variants, solution approaches employed so far have been prevalently problem-specific. In contrast, the general-purpose solver SCIP-Jack can compute optimal solutions to both the classical Steiner tree problem and 11 related problems without modification. Furthermore, the solver comes with shared and distributed parallelization extensions by means of the UG framework. In this talk we report on the current state of SCIP-Jack with special focus on the parallel extension, which has already allowed to solve several well-known instances for the first time to optimality.

Thorsten Koch

Zuse Institute Berlin  
koch@zib.de

Daniel Rehfeldt, Yuji Shinano

Zuse Institute Berlin  
Takustr. 7 14195 Berlin  
rehfeldt@zib.de, shinano@zib.de

### MS103

#### Towards the New Generation of Multiperiod Optimal Power Flow Solvers

Distributed energy storage devices are commonly employed as an effective approach for addressing operational challenges associated with the integration of renewable energy sources in modern power grids. The integration of storage device equations however, over the time period of interest, enters the optimal power flow (OPF) problem as an intertemporal linear inequality constraint that couples the OPF problems defined for each subdivision of the overall time period. The computational work for the solution of the so called multiperiod optimal power flow problem (MPOPF), is growing rapidly as the number the desired time period subdivisions increase. This work proposes a parallel structure-exploiting data-compression interior point algorithm that minimizes the computational work for the solution of the MPOPF problem. A Schur-complement-based approach is tailored to the particular structure of the linear systems associated with the optimality conditions allowing significant savings in both computational time and memory even for large-scale MPOPF problems. We provide benchmark cases for power grids of increasing complexity and compare the overall runtime performance and memory consumption against standard optimization methods like KNITRO, MIPS provided in the software package MATPOWER.

Drosos Kourounis

USI - Università della Svizzera italiana  
Institute of Computational Science  
drosos.kourounis@usi.ch

Juraj Kardos  
Institute of Computational Science  
Università della Svizzera italiana  
juraj.kardos@usi.ch

Alexander Fuchs  
Research Center for Energy Networks,  
Swiss Federal Institute of Technology (ETH Z)  
fuchs@fen.ethz.ch

Olaf Schenk  
USI - Università della Svizzera italiana  
Institute of Computational Science  
olaf.schenk@usi.ch

### MS103

#### **Parallel Computing in a Hierarchical Mixed-Integer Linear Programming Method for Optimal Design of Distributed Energy Systems**

To attain the highest performance of energy supply systems from the viewpoint of reductions in cost, energy consumption, and CO<sub>2</sub> emission, it is necessary to determine design specifications optimally in consideration of operational strategies corresponding to seasonal and hourly variations in energy demands. Mixed-integer linear programming (MILP) methods have been applied widely to such optimal design problems. The authors have proposed a hierarchical MILP method to solve such optimal design problems efficiently. The original problem is solved by dividing it into a relaxed optimal design problem at the upper level and optimal operation problems which are independent of each other at the lower level. In addition, some strategies have been adopted to enhance the computation efficiency furthermore. In this work, parallel computing is adopted as another strategy to enhance the computation efficiency. Multiple design solution candidates are searched simultaneously in the relaxed optimal design problem at the upper level by allowing multiple parallel threads in the MILP solver CPLEX. In addition, multiple optimal operation problems are solved simultaneously at the lower level by utilizing Message Passing Interface. Through a case study on the optimal design of energy supply systems, it is clarified how the parallel computing is effective to enhance the computation efficiency in comparison and/or combination with the previous strategies.

Ryohei Yokoyama  
Department of Mechanical Engineering  
Osaka Prefecture University  
yokoyama@me.osakafu-u.ac.jp

Yuji Shinano  
Zuse Institute Berlin  
Takustr. 7 14195 Berlin  
shinano@zib.de

### MS104

#### **Exploring Intrusiveness of Parallel-in-Time Methods**

One important aspect when choosing a time-parallel method is its intrusiveness, i.e., the effort one has to put into implementing the method when aiming at adding parallelism to an existing time-stepping code. The multigrid-reduction-in-time (MGRIT) algorithm of Falgout et al. is a non-intrusive approach that, similarly to time stepping, uses an existing time propagator to integrate from one time

to the next. The corresponding open source code XBraid is flexible, allowing for a variety of time stepping, relaxation, and temporal coarsening options. Comparing the performance of MGRIT with the more invasive space-time multigrid (STMG) algorithm by Horton and Vandewalle, the speedup in comparison to space-parallel methods with sequential time stepping is larger for the invasive STMG approach than for the non-invasive MGRIT approach. In this talk, we discuss incorporating some of the more intrusive, but highly efficient, aspects of the STMG algorithm into XBraid. We develop an STMG-like method in the MGRIT framework and present results which demonstrate performance improvements for a parabolic model problem.

Stephanie Friedhoff  
University of Wuppertal  
friedhoff@uni-wuppertal.de

Robert D. Falgout, Jacob B. Schroder  
Lawrence Livermore National Laboratory  
rfalgout@llnl.gov, schroder2@llnl.gov

### MS104

#### **Communication-Aware Adaptive Parareal (CAAP) with Application to a Hyperbolic PDE System**

Parallel integration of hyperbolic evolution problems using Parareal has for a long time been considered impractical. Theoretical results have shown that it is not possible to give a general guarantee on the stability of the algorithm when applied to this class of problems, and practical experiments have likewise suggested issues of slow convergence. In this talk we present our experiences using Parareal to accelerate a tsunami simulation application complete with inundation modelling. The underlying PDE is the shallow water wave equation. The system is in many ways typical of a purely hyperbolic system of coupled non-linear PDEs, the solutions of which typically contain both shocks and smooth regions interacting in a non-trivial manner. Contrary to what one might expect, we find that with sufficient care in constructing the coarse operator, and a clever way of choosing the time-subdomain length, we obtain parallel speed-up beyond what is possible using conventional spacial domain-decomposition techniques alone. A critical aspect in achieving this was to decouple the time-subdomain length from the total time to be integrated, combined with theoretical considerations on how to choose the time-subdomain length optimally so to balance the impact of the communication of solution states and the convergence rate.

Allan Nielsen, Jan S. Hesthaven  
EPFL  
allan.nielsen@epfl.ch, jan.hesthaven@epfl.ch

### MS104

#### **Performance Analysis of the PinT Computation of the Large-Scale Simulation for the Diffusion Equation**

The number of computational core of recent super computers amounts to one million. In such a high parallel environment, it is difficult to fulfill the high efficiency computation. Parallel-in-time integration method is in a practical use level. However, high performance can not be guaranteed in the high parallel environment because there are few arguments of HPC. In this study, we apply Parareal method to the three-dimensional diffusion equation, and

measure the performance in high parallel environment.

Yoshiharu Ohi

RIKEN Advanced Institute for Computational Science  
yoshiharu.ohi@riken.jp

#### MS104

##### On Implementations of Parallel-in-Time Integration Methods

With million-way concurrency at hand, the efficient use of modern HPC systems has become one of the key challenges in computational science and engineering. Novel mathematical concepts are needed to fully exploit massively parallel architectures and for the numerical solution of time-dependent partial differential equations, time-parallel methods have shown to provide a promising way to extend prevailing scaling limits of numerical codes. While many ideas, algorithms and proofs of concept exist in this domain, the number of actual large-scale time-parallel application codes is small. Interestingly, the same is also true for stand-alone parallel-in-time libraries, which are mainly used for showcasing performance or for testing new ideas. In particular, codes which can deal with parallelization in time as well as in space are rare. In this talk, we will give an overview of what is currently available and describe some of the issues arising when implementing parallel-in-time integration methods. In addition, we will trace the evolution of one particular approach, the parallel full approximation scheme in space and time (PFASST), and discuss the different implementations this algorithm has inspired.

Ruth Schöbel, Robert Speck

Juelich Supercomputing Centre  
Forschungszentrum Juelich GmbH  
r.schoebel@fz-juelich.de, r.speck@fz-juelich.de

#### MS105

##### A DSL for Hybridization and Static Condensation Techniques in Firedrake

The static condensation technique for the reduction of global matrix systems is a powerful tool for the efficient solution of PDEs using the finite element method. Particularly when hybridizing finite element methods, the reduction of global matrix systems together with element-local solves is a distinguishing feature of such discretizations. This solution approach permits a highly efficient means for solving indefinite problems in a computationally efficient manner. However, such techniques require manual intervention in intricate numerical code in order to execute the necessary dense linear algebra operations. This process must be repeated each time the model is extended, modified, or de-bugged. In contrast, the Firedrake Project at Imperial College London automates much of the finite element process by composing several layers of abstraction. Firedrake takes the discretized equations in symbolic form as input, and automatically generates high-performance parallel code from this mathematical specification. In this work, we introduce a new domain-specific language for describing localized linear algebra on finite element tensors. From a high-level specification of the local linear algebra operations, parallelizable C++ code is automatically generated from them. Together with this new DSL, we aim to extend static condensation and hybridization techniques within Firedrake for use in automated simulation.

Thomas H. Gibson

Department of Mathematics

Imperial College London

t.gibson15@imperial.ac.uk

#### MS105

##### Optimizing Auto-Generating Kernel-Code for Low Order FEM on Locally Structured Meshes

This talk presents a python based approach to generate kernel code for locally block structured Finite Element methods. The local refinements increase the arithmetic intensity of the kernel, improving the efficiency of low order FEM. Based on a weak formulation in UFL we construct the loop structure of the kernel represented in Loo.py, allowing transformations and optimizations within the polyhedral model. We present initial results and discuss different optimisations for 2d and 3d meshes.

Marcel Koch

University of Münster  
marcel.koch@uni-muenster.de

#### MS105

##### Efficient Matrix-Free Discontinuous Galerkin Assembly: To Hand-Write Or to Generate?

Medium to high order Discontinuous Galerkin (DG) methods have received a fair bit of interest recently, as their high computational intensity and structured memory access patterns are a good fit for current HPC architectures. However, implementations are rather complex and highly sensitive to small implementation details. In this talk, we present and compare two different approaches to developing high-performance implementations of matrix-free DG: In our earlier efforts, we developed a C++ implementation for convection-diffusion problems and Navier-Stokes on top of the DUNE framework. More recently, our UFL-based code generation framework gained a sum-factorized DG backend based on the earlier work, which allowed us to quickly add support for other types of problems and hardware architectures. There are tradeoffs between the two approaches, and we discuss how their strengths combine for a better framework than would be possible with either one alone.

Steffen Müthing

Heidelberg University  
steffen.muething@iwr.uni-heidelberg.de

Peter Bastian

Interdisciplinary Center for Scientific Computing  
University of Heidelberg  
peter.bastian@iwr.uni-heidelberg.de

René Heß, Dominic Kempf, Marian Piatkowski

IWR, Heidelberg University  
rene.hess@iwr.uni-heidelberg.de, dominic.kempf@iwr.uni-heidelberg.de, marian.piatkowski@iwr.uni-heidelberg.de

#### MS105

##### Achieving Portable Performance Across Architectures with PETSc

We present our approach for achieving portable performance across different hardware and software architectures in the PETSc library. Our approach is based on compact and architecture-specific kernels, which are composed in an architecture-agnostic manner to preserve runtime configurability and maintainability. Backed by the concrete

examples of sparse matrix-vector products (MatMult) and sparse matrix-matrix products (MatMatMult) we claim that performance portability requires a continuous evolution of the code base and cannot be achieved by only offloading the task to external tools such as compilers.

Karl Rupp

Freelance Computational Scientist  
me@karlrupp.net

Richard T. Mills  
Intel Corporation  
richardtmills@gmail.com

Hong Zhang, Hong Zhang  
Argonne National Laboratory  
hzhang@mcs.anl.gov, hongzhang@anl.gov

Barry F. Smith  
Argonne National Lab  
MCS Division  
bsmith@mcs.anl.gov

#### MS106

##### **Solving a Newton Equation on the Stiefel Manifold with Matrix-Free Krylov Subspace Methods**

Optimization methods for unconstrained optimization problems on the Euclidean space can be generalized to those on Riemannian manifolds. We take up the truncated singular value decomposition, which is one particular problem defined as an optimization problem on the product of two Stiefel manifolds. For this problem, an algorithm of the Riemannian Newtons method has been designed. However, this is not easy to implement in its original form because the Newton equation is expressed by a system of matrix equations that is difficult to solve directly. In this talk, we use matrix-free Krylov subspace methods for solving linear systems of equations into which the Newton equation is rewritten. Numerical experiments show the effectiveness of the presented method. Our approach can be used on other problems as well.

Kensuke Aihara

Tokyo City University, Japan  
aiharak@tcu.ac.jp

Hiroyuki Sato  
Department of Applied Mathematics and Physics  
Graduate School of Informatics, Kyoto University  
hsato@amp.i.kyoto-u.ac.jp

#### MS106

##### **Parallel Solution for Multi-Group Radiation Transfer Equations**

In inertial confinement fusion (ICF) numerical simulations, it is necessary and very important to solve multi-group radiation transfer (MGRT) equations. Usually, the solution cost for MGRT equations is more than 90% of the whole simulation. The source iteration method is often used to solve MGRT equations, and one typical parallel solution method for MGRT equations is the spatial domain decomposition (spatial parallelism) method. In numerical simulations, the scalability of application code is limited if only spatial parallel strategy is used. By exploring the character of the source iteration method, a two-level parallel strategy is designed for solving MGRT equations. In this strategy,

the equations are firstly divided into several sets, and in each source iteration, the equations in different sets are solved in parallel. And then for solving each equation, the spatial parallel strategy is used. The scalability of the application code is improved dramatically with the two-level parallel strategy.

Hengbin An

Institute of Applied Physics and Computational Mathematics  
an\_hengbin@iapcm.ac.cn

#### MS106

##### **A Parallel Algorithm for Tridiagonal Quasi-Toeplitz Linear Systems**

We consider the parallel solution of tridiagonal quasi-Toeplitz linear systems. By exploiting the special quasi-Toeplitz structure, we give a decomposition form of the coefficient matrix. Then, we propose a parallel algorithm for solving the tridiagonal quasi-Toeplitz linear systems. Numerical results are presented to illustrate the performance of our algorithm.

Jintao Zhang

School of Mathematics and Physics Science  
Dalian University of Technology at Panjin  
zjt@dlut.edu.cn

Lei Du

Dalian University of Technology  
China  
dulei@dlut.edu.cn

#### MS106

##### **Block GMREs Methods for Linear Systems with Multiple Right-Hand Sides**

Solving a sequence of large linear systems with several right-hand sides given simultaneously is at the core of many problems in the computational sciences, such as in radar cross section calculation in electromagnetism, wave scattering and wave propagation in acoustics, various source locations in seismic and parametric studies in general. In that framework, block Krylov approaches appear as good candidates for the solution as the Krylov subspaces associated with each right-hand side are shared to enlarge the search space. They are attractive not only because of this numerical feature (larger search subspace), but also from a computational view point as they enable higher data reusability consequently locality (BLAS3-like implementation). These nice data features comply with the memory constraint of modern multicore architectures. In this talk, we present recent work on block Gmres methods for “simultaneous” right-hand side linear systems.

Yanfei Jing

University of Electronic Science and Technology of China  
yanfeijing@uestc.edu.cn

#### MS107

##### **An Implicit, Multiscale, Conservative Hybrid Kinetic-Ion/Fluid-Electron Algorithm for Plasma Simulation**

We introduce an implicit, conserving hybrid kinetic-ion/fluid-electron electromagnetic solver. Ions are described by particles, while electrons are described by an

Eulerian fluid model, including Hall effects, resistivity and viscosity for well-posedness. An implicit formulation, together with careful spatial discretization of the electron model and ion gather/scatter interpolations, enables (to our knowledge, for the first time) simultaneous conservation of mass, momentum, and energy. We precondition our system by addressing fast electron timescales. We will demonstrate the algorithm on problems ranging from the simple (Landau damping of an ion acoustic wave) to the more complex (magnetic reconnection).

Luis Chacón, Adam Stanier  
Los Alamos National Laboratory  
chacon@lanl.gov, stanier@lanl.gov

## MS107

### A Partitioned Coupling Analysis of FEM and SPH for Large-Scale Fluid-Structure Interaction Problems

This research focuses on a partitioned coupling analysis for 3-dimensional fluid-structure interaction problems, adopting an incompressible smoothed particle hydrodynamics method for fluid dynamics involving free surface flow and the finite element method for structural dynamics. A coupling analysis of a particle-based method and a grid-based method has been considerable researched. However, most of these are developed by in-house codes and then become a function-specific application software. In this research, to save cost in software developments whose software has the general versatility, the open source software is utilized. Especially, a general-purpose finite element analysis system, named ADVENTURE, and a general-purpose coupling analysis platform, named REVOCAP\_Coupler, are adopted. In here, the use of the REVOCAP\_Coupler has forewarned that it unifies software based on the grid-based method. To solve this issue, techniques of a virtual particle on fluid-structure boundaries and a dummy mesh for fluid analysis domain is proposed. To verify a developed system, a 3-dimensional dam break problem with an elastic obstacle is solved, and the result is compared with the results calculated by other methods.

Masao Ogino, Takuya Iwama  
Nagoya University  
masao.ogino@cc.nagoya-u.ac.jp,  
iwama@hpc.itc.nagoya-u.ac.jp

Mitsuteru Asai  
Kyushu University  
asai@doc.kyushu-u.ac.jp

## MS107

### Matrix-Free Parallel Multigrid for Fe Systems with a Trillion Unknowns

Abstract Not Available At Time Of Publication.

Ulrich J. Ruede  
University of Erlangen-Nuremberg  
Department of Computer Science (Simulation)  
ulrich.ruede@fau.de

Dominik Thoennes, Nils Kohl  
University of Erlangen-Nuremberg

dominik.thoennes@fau.de, nils.kohl@fau.de

## MS107

### Leveraging Application Structure Within Next Generation Multigrid Solvers

Many applications have a specific structure that can be leveraged within a linear solver to improve convergence or to take advantage of architecture characteristics. This talk focuses on structure within an underlying grid including vertically extruded meshes (for ice sheet modeling), block structured meshes (for hypersonic flow), and uniformly refined unstructured meshes for magneto-hydrodynamics. The benefits of a structured approach include lower memory bandwidth requirements, greater vectorization potential, less indirect memory access, simpler setup procedures, and modest nonzero growth on coarse grids of an AMG hierarchy. Further, special structured multigrid algorithms can exhibit superior convergence rates when compared with their unstructured multigrid counterparts.

Most Sandia applications, however, are not fully structured. For example, applications may require complex unstructured meshes within only small sub-regions to accurately capture important features. In other cases, an underlying unstructured mesh is needed, but it is natural to refine this mesh in a uniform fashion. This talk highlights our software and algorithm developments to leverage partial structure that is present within different application scenarios. While some additional application information is required, our goal is to provide interfaces that are convenient and natural for those coming from an unstructured finite element or block structured finite volume perspective.

Raymond Tuminaro  
Sandia National Laboratories  
rstumin@sandia.gov

Luc Berger-Vergiat, Matthias Mayr  
Sandia National Labs  
lberge@sandia.gov, mmayr@sandia.gov

## MS108

### Accelerating Irregular Applications on Heterogeneous Computer Systems

Efficient execution of irregular applications in heterogeneous systems such as accelerated processing units (APUs) has been a challenge due to their irregular parallelism, access patterns and load imbalance. This talk will discuss various techniques to address these issues using graph workloads as a case study. The techniques include algorithmic and runtime optimizations, libraries and networking improvement by leveraging the latest hardware features.

Shuai Che  
AMD Research  
shuai.che@amd.com

## MS108

### A Novel Parallel SPGEMM Algorithm on Sunway TaihuLight

In this work we first adjust the sparse matrix-matrix multiplication (SpGEMM) algorithm to make full use of the architecture of Sunway TaihuLight. Then, based on this adjustment we develop a novel parallel SpGEMM implementation on the considered supercomputer. Our implementation delivers an excellent performance than the avail-

able ones in terms of the execution time, memory usage and scalability.

Xin He

Institute of Computing Technology  
Chinese Academy of Science  
hexin@ncic.ac.cn

#### MS108

##### **When Sparse Matrices Met Heterogeneous Processors: Opportunities and Challenges**

In recent years, low-throughput GPUs have been integrated onto the same chip as the CPU. AMD APUs are representatives in this trend. The newest hardware progress, such as unified virtual address space, makes tightly coupled CPU-GPU heterogeneous processors a promising tool for scientific computing. This talk will focus on our empirical study of performance behaviors of sparse matrix algorithms (e.g., SpTRANS, SpMV, SpTRSV and SpGEMM) on emerging heterogeneous processors. A performance comparison with modern multi-core and many-core processors will be presented. Through an analysis on collected performance data, several opportunities and challenges will be discussed.

Weifeng Liu

Department of Computer Science  
Norwegian University of Science and Technology  
weifeng.liu@ntnu.no

#### MS108

##### **Refactoring Several Scientific Kernels on SW26010 Many-core Processor**

SW26010 is the new-generation customized heterogeneous many-core processor equipped in Sunway TaihuLight supercomputer. The peak performance of SW26010 is over 3TFLOPS while its compute-to-memory bandwidth ratio can only achieve 22, which makes tuning memory-intensive applications even challenging. In this talk, I would like to present the recent progress on refactoring memory intensive scientific kernels such as stencil, sparse triangular solvers (SpTRSV) as well as sparse matrix-vector multiplication (SpMV). The numerical experiments show that the satisfying performance can be gained on these kernels by using more aggressive and finer-grained redesign.

Wei Xue

Tsinghua University  
xuewei@tsinghua.edu.cn

#### MS109

##### **A Distributed-memory Hierarchical Solver for General Sparse Linear Systems**

We present a parallel hierarchical solver for solving general sparse linear systems on distributed-memory machines. The fully algebraic algorithm can be faster and more memory-efficient than sparse direct solvers by exploiting the low-rank structure of fill-in blocks. With high-accuracy low-rank approximations, the algorithm behaves similarly to a direct solver, and it can also be used as a pre-conditioner when the approximation accuracy is low. The parallel algorithm is based on data decomposition, and only local communication is required for updating boundary data on every processor. Moreover, the computation-to-communication ratio of the parallel algorithm is approximately the volume-to-surface-area ratio of the subdomain

owned by every processor. We show various numerical results to demonstrate the versatility and scalability of the parallel algorithm.

Chao Chen

Stanford University  
cchen10@stanford.edu

Siva Rajamanickam

Sandia National Laboratories  
srajama@sandia.gov

Erik G. Boman

Center for Computing Research  
Sandia National Labs  
egboman@sandia.gov

Eric F. Darve

Stanford University  
Mechanical Engineering Department  
darve@stanford.edu

#### MS109

##### **Tacho: Task Parallel Sparse Direct Cholesky Factorization with Bounded Memory Constraint**

We present task-parallel sparse direct Cholesky factorization with a bounded memory constraint. The solver uses multi-frontal algorithm and two-level parallelism. The factorization requires workspace to store partial factorization of front matrices. In general, fixed size of workspace is allocated per thread. This can be problematic when we use a large number of threads, which is characteristic in modern architectures. We remedy this problem by using Kokkos dynamic task-dag parallelism with a variable-size block memory pool. When a task requires workspace, it allocates memory from the pool. The memory is deallocated after dependent tasks are executed. Several numerical experiments are presented and compared with Intel MKL Pardiso to demonstrate the merits of the approach.

Kyungjoo Kim, Siva Rajamanickam, H. Carter Edwards  
Sandia National Laboratories

kyukim@sandia.gov, srajama@sandia.gov, hcedwar@sandia.gov

#### MS109

##### **ACHILES : An Asynchronous Iterative Sparse Linear Solver**

Asynchronous methods are becoming more popular for solving linear systems offering faster performance in many cases. In such methods, local updates are performed using the latest available information from other sub problems eliminating the need for synchronization. In this talk, we present ACHILES, an asynchronous iterative linear solver that uses MPI Remote Memory Access (RMA) to asynchronously communicate data. We further show the performance of ACHILES for asynchronous Restricted Additive Schwarz.

Paritosh P. Ramanan, Edmond Chow

School of Computational Science and Engineering  
Georgia Institute of Technology  
paritoshpr@gatech.edu, echow@cc.gatech.edu

Siva Rajamanickam

Sandia National Laboratories  
srajama@sandia.gov



Erik G. Boman  
Center for Computing Research  
Sandia National Labs  
egboman@sandia.gov

#### MS109

##### **Multilevel Iterative Methods and Their Application in Field-Scale Petroleum Reservoir Simulation**

We examine linear algebraic solvers for a general purpose compositional simulator. In particular, the decoupling stage of the constraint pressure residual preconditioner for linear systems arising from the fully implicit scheme is evaluated. An asymptotic analysis of the convergence behavior is given when time stepsize approaches zero. Based on this analysis, we propose an analytical decoupling technique, from which the pressure equation is directly related to an elliptic equation and can be solved efficiently. We show that this method ensures good convergence behavior of the algebraic solvers in a two-stage preconditioner. We also propose a semi-analytical decoupling strategy that combines the analytical method and alternate block factorization method. Numerical experiments demonstrate the superior performance of the analytical and semi-analytical decoupling methods compared to existing methods. This talk is based on a collaboration with Prof. Jinchao Xu and Dr. Changhe Qiao from Penn State University.

Chensong Zhang  
Chinese Academy of Sciences  
zhangcs@lsec.cc.ac.cn

#### MS110

##### **Design Experience on Intel HARP2 Platform using OpenCL**

Recently, Moores Law is approaching to the end and the semiconductor technology is going to stop improving power consumption and performance. Towards post-Moore era, FPGA (Field Programmable Gate Array) is expected to gain good efficiency by reconfiguration as an application-specific accelerator on the fly. Intel HARP2 (Hardware Accelerator Research Program Ver.2) platform is a prototype system using a preproduction version of Intel Xeon processor including in-package FPGA. In this presentation, design experience using OpenCL is described for HARP2 platform, and preliminary evaluation using several applications written by OpenCL is demonstrated on the HARP2 system.

Toshihiro Hanawa  
Information Technology Center  
The University of Tokyo  
hanawa@cc.u-tokyo.ac.jp

Taisuke Boku  
Center for Computational Sciences, University of Tsukuba  
taisuke@cs.tsukuba.ac.jp

#### MS110

##### **Towards Compute Everywhere: Computing in the Network with FPGA-Accelerated Clouds and Clusters**

In his SIAM-PP16 keynote, Paul Coteus described the future of HPC as being, in part, Compute Everywhere. Our take is as follows: As Moores Law ends, performance gains will come from having compute capability wherever data

moves. Also, designs will be, not only application driven, but adaptable to the application. FPGAs are the ideal devices to explore this future. Since the hardware adapts to the application, rather than the reverse, we can obtain very high efficiency at low power. And since FPGAs are hybrid communication-computation processors that can be interconnected directly chip-to-chip, large-scale communication can proceed with both high bandwidth and low latency. Our work with FPGA-centric clouds and clusters includes the following: custom, process-specific, communication fabrics; scheduled routing to eliminate congestion and reduce switching overhead; compute-offload to shorten instruction paths; and application-specific primitives for computing in the network. We illustrate these ideas with case studies from Machine Learning and also Molecular Dynamics including the FFT, grid-particle mapping with tricubic interpolation, and use of collectives in the range-limited computation. We conclude by discussing some possible impacts of these additional degrees of freedom, from mundane offload of data transformations, to novel algorithm strategies.

Martin Herbordt  
Boston University  
herbordt@bu.edu

#### MS110

##### **Scalable Inter-FPGA Direct Communication for Parallel FPGA Applications**

FPGAs have been emerging in HPC research area because their computation and communication capabilities have been dramatically strengthened in recent years thanks to improvements of semiconductor integration technologies depending on Moores Law. In addition to the FPGA performance improvements, OpenCL-based FPGA development toolchains have been developed and offered by FPGA vendors, which makes its programming effort lower compared to the past. These backgrounds show possibilities of realizing a concept to enable on-the-fly offloading computation parts, where CPUs/GPUs are weak, to FPGAs while performing low-latency data movement. We think that this concept is one of the keys to more improve performance of modern heterogeneous supercomputers using accelerators like GPUs. Today, we talk about how to design high-performance inter-FPGA Ethernet communication controlled from OpenCL kernel. This technique gives us opportunity to facilitate parallel FPGA applications offering good strong scalability.

Ryohei Kobayashi  
University of Tsukuba  
kobayashi@cs.tsukuba.ac.jp

#### MS110

##### **Data-Flow HPC with Custom Hardware on FPGA Cluster**

For hardware resources distributed on many chips in a large system to efficiently be utilized under high network latency, data-flow computing is promising due to its decentralized synchronization feature. Although software implementation is expected to have high overhead for data-flow computing with many threads, we focus on offloading some parts of them to customized hardware on FPGAs. We introduce our big picture for data-flow HPC and our research progress including a data-flow compiler for FPGAs. We show our research progress of FPGA shell designs and

application case studies using our data-flow compiler.

Kentaro Sano  
Tohoku University  
kentaro.sano@gmail.com

#### MS111

##### Analyzing Exascale Memory Architectures Using the SST Toolkit

Designing the memory subsystem of future Exascale-class supercomputing systems is an extremely challenging activity. In part, the challenges come from the wide variation found in potential media for memory (e.g. High Bandwidth Memory, conventional DDR or the emerging use of non-volatile technologies) as well as the multiple different processor designs being proposed. In this talk we will present the memory modeling capabilities of Sandia's Structural Simulation Toolkit (SST) including its well-exercised cache, network-on-chip, memory controller and memory-backend simulation components. We present an overview of the modeling capabilities found in the toolkit and initial results for strawman Exascale memory subsystem designs demonstrating, in particular, the sensitivity of several important classes of DOE application kernels to the use of non-volatile technologies as a memory medium. Our results show negligible performance impact for some kernels implying future use of NV technology could result in lower node-energy use, while for other kernels, the impact on performance is significant and will lead to considerable increases in execution times. SST is a foundational technology for the DOE's Exascale Computing Project (ECP) Design Space Exploration (DSE) task and is available as an open source toolkit for broader HPC community use enabling replication of results and use by academic, industry and other laboratory users.

Clayton Hughes  
Sandia National Laboratories  
chughes@sandia.gov

Simon D. Hammond  
Scalable Computer Architectures  
Sandia National Laboratories  
sdhammo@sandia.gov

Arun Rodrigues  
Sandia National Laboratories  
afrodri@sandia.gov

#### MS111

##### Unifying Software Models with Hardware Emulation

Traditionally, hardware designers developed high-level software models that simulate the behavior of future hardware in order to perform chip- or system-level evaluations, and then implemented a few hardware designs in a hardware description language (HDL). This, however, requires implementing the same circuitry twice and ensuring the functionality between models is equivalent, sometimes a huge task. In this talk, I will discuss a HDL which our group is utilizing for hardware research. From a common code-base, Chisel generates both C++ and Verilog models from a single code base and has a development effort comparable to functional programming. Using Chisel, we have developed OpenSoC Fabric, a comprehensive on-chip network generator. OpenSoC Fabric has a powerful list of configuration parameters and state-of-the-art internal ar-

chitecture. I will also cover other upcoming projects where we apply the same concept to, such as quantum computing control.

Farzad Fatollahi-Fard, David Donofrio,  
George Michelogiannakis, John Shalf  
Lawrence Berkeley National Laboratory  
ffard@lbl.gov, ddonofrio@lbl.gov, mihelog@lbl.gov,  
jshalf@lbl.gov

#### MS111

##### Exploring Extreme Heterogeneity with Analytical Modeling

Most proposed exascale architectures include heterogeneous processing units, ranging from traditional CPUs with attached accelerators to specialized and reconfigurable processors. A variety of new memory technologies and configurations are also being proposed. As architectures become more complex and heterogeneous, the combined possibilities for partitioning and mapping computation, data, threads, and processes onto various parts of the system are exploding into an extremely large design space. Proposed exascale architectures are currently being evaluated using a combination of analytical modeling and architectural simulation approaches, but these approaches are limited in scope due to high overheads. The results can also vary greatly depending on how an application is mapped to the architecture, with the risk of possibly misleading conclusions. We propose low-overhead techniques for capturing architecture-independent application characteristics in the form of an abstract application model (AAM). We map the AAM to an abstract machine model (AMM) so as to satisfy specified constraints. The constraints determine the feasible space of possible solutions which can then be optimized for different objective functions. We give preliminary results on development of a mapping component for the ASPEN analytical modeling framework that formalizes and automates how application requirements are mapped onto target heterogeneous architectures.

Shirley Moore  
Oak Ridge National Laboratory  
mooresev@ornl.gov

Jeffrey S. Vetter  
Oak Ridge National Laboratory  
Georgia Institute of Technology  
vetter@computer.org

#### MS111

##### Modeling Massively Multithreaded GPUs with GPGPU-Sim

This talk will present the simulation infrastructure and related tools for the GPGPU-Sim simulator. GPGPU-Sim is a detailed, cycle-level simulator that models a modern GPU running applications written in CUDA. CUDA is an extension of the familiar C programming language used for programming GPUs and related many-core architectures. The rapidly growing number of applications written for these devices typically contain tens of thousands of threads making them an interesting workload for future many-core architecture research. The simulator can be downloaded from the github <https://github.com/gpgpu-sim/gpgpu-sim.distribution>. The focus of this talk is on what is modeled in GPGPU-Sim, an overview of the research based on using GPGPU-Sim and future plans for

development.

Tim Rogers  
Purdue University  
timrogers@purdue.edu

## MS112

### A Parallel-Adaptive Mesh Refinement Numerical Approach for Generalized Phase Field Simulations

Revealing the underlying physics during solidification of metal alloys requires development of powerful physical and numerical approaches. The state-of-art phase field model is becoming more popular nowadays to simulate microstructure transition during solidification. In this study, a numerical robust algorithm, namely Para-AMR, comprising of adaptive mesh refinement (AMR) and parallel (Para-) computing capabilities was developed to solve the phase field equations. Numerical tests revealed that this algorithm could improve the computational efficiency for 3 orders of magnitude. Microstructure transition in response to external forces, including electromagnetic fields, ultrasound and mechanical stirring was simulated and studied. Accordingly, important solidification behaviors, including dendrite growth (for both aluminum alloy and magnesium alloy), fragmentation, fluid flow, and coarsening were studied with new physics revealed.

Zhipeng Guo  
Tsinghua University  
zhipeng\_guo@mail.tsinghua.edu.cn

## MS112

### An Optimized and Parallelized Phase-Field Solver for Sintering Using the Pace3d Framework

Simulations allow to improve the development of new high performance materials with tailored microstructures and defined properties. To fulfill these conditions a deep understanding the underlying microstructure evolution processes are required. For ceramic materials, the process of sintering is of outstanding high interest. During sintering, a pressed body of loose powder, called green body, containing different sized particles is heated, meanwhile the body is densifying and grain growth occurs, which leads to a solid ceramic component. To efficiently investigate the microstructure evolution in large scale 3D domains during the sintering process the phase-field model is presented in this talk. The model is implemented in the massive parallel phase-field solver framework PACE3D. It is optimized on various levels starting from the model and parameters down to the hardware. The solver allows to resolve and calculate an arbitrary number of individual grains, using a local reduction technique and a material class based parameterization concept. The evolution equations for the grains and the concentration are explicitly vectorized using intrinsics. Based on a light weight macro layer for the vectorization, the solver is also able to exploit NEC SX-ACE CPUs. Performance results on a single core, single node and with 96100 processes on the German supercomputers Hazel Hen (x86), ForHLR II (x86) and Kabuki (NEC SX-ACE) are shown and discussed.

Johannes Hötzer  
Hochschule Karlsruhe Technik und Wirtschaft  
Institute of Materials and Processes

johannes.hoetzer@kit.edu

## MS112

### Parallel GPU Phase-Field Simulations of Dendrite Competitive Growth

Phase-field method has been emerged as a powerful numerical tool to accurately simulate the dendrite growth during solidification of pure materials and alloys. On the other hand, due to employing a diffuse interface model, the large computational cost is a major drawback in the phase-field simulation. To overcome the drawback, we have developed a parallel GPU computational scheme for the large-scale phase-field simulation. In this talk, we show two topics regarding the large-scale phase-field simulation. One is the competitive growth of multiple dendrites during directional solidification of an alloy. Second is the investigation of effects of natural convection on the competitive growth of multiple dendrites during directional solidification of an alloy. These simulations have been firstly enabled by the large-scale parallel GPU computation.

Tomohiro Takaki  
Kyoto Institute of Technology  
Mechanical and System Engineering  
takaki@kit.ac.jp

Munekazu Ohno  
Hokkaido University  
mohno@eng.hokudai.ac.jp

Yasushi Shibuta  
The university of Tokyo  
shibuta@material.t.u-tokyo.ac.jp

Shinji Sakane  
Kyoto Institute of Technology  
d7821004@edu.kit.ac.jp

Takashi Shimokawabe  
Global Scientific Information and Computing Center  
Tokyo Institute of Technology  
shimokawabe@cc.u-tokyo.ac.jp

Takayuki Aoki  
Tokyo Institute of Technology  
taoki@gsc.titech.ac.jp

## MS112

### Multiple-Gpu Computing of Phase-Field Crystal Simulation of Grain Growth

Abstract Not Available At Time Of Publication.

Akinori Yamanaka  
Tokyo University of Agriculture and Technology  
a-yamana@cc.tuat.ac.jp

## MS113

### Coupling Field and Particle Methods for Massively Parallel Simulations of the Earth Mantle

Particle-in-cell methods couple mesh-based methods for the solution of continuum mechanics problems, with the ability to advect and evolve particles. They have a long history and many applications in scientific computing. However, they have most often only been implemented for ei-

ther sequential codes, or parallel codes with static meshes that are statically partitioned. In contrast, many mesh-based codes today use adaptively changing, dynamically partitioned meshes, and can scale to thousands or tens of thousands of processors. Consequently, there is a need to revisit the data structures and algorithms necessary to use particle methods with modern, mesh-based methods. Here we review commonly encountered requirements of particle-in-cell methods, and describe efficient ways to implement them in the context of large-scale parallel finite-element codes that use dynamically changing meshes. We also provide practical experience for how to address bottlenecks that impede the efficient implementation of these algorithms and demonstrate with numerical tests both that our algorithms can be implemented with optimal complexity and that they are suitable for very large-scale, practical applications. We discuss a reference implementation in ASPECT, an open source code for geodynamic mantle-convection simulations built on the deal.II library.

Wolfgang Bangerth  
Colorado State University  
Texas A&M University  
bangerth@colostate.edu

Rene Gassmoeller  
Colorado State University  
rene.gassmoeller@mailbox.org

#### MS113

##### **A Stencil Scaling Approach for Accelerating Matrix-Free Finite Element Implementations**

We present a novel approach to fast on-the-fly low order finite element assembly for scalar elliptic partial differential equations of Darcy type with variable coefficients optimized for matrix-free implementations. Our approach introduces a new operator that requires roughly one third of the floating-point operations compared to a classical finite element assembly. An a priori analysis shows that solutions obtained by this approach have asymptotically optimal order convergence in the  $H^1$ - and  $L^2$ -norm on hierarchical grids.

Daniel Drzisga  
Technical University of Munich  
drzisga@ma.tum.de

Simon Bauer, Marcus Mohr  
Department fuer Geo- und Umweltwissenschaften  
Universitaet Muenchen  
simonb@geophysik.uni-muenchen.de,  
mohr@geophysik.uni-muenchen.de

Ulrich J. Ruede  
University of Erlangen-Nuremberg  
Department of Computer Science (Simulation)  
ulrich.ruede@fau.de

Christian Waluga, Barbara Wohlmuth  
Technical University of Munich  
christian.waluga@ma.tum.de,  
bara.wohlmuth@ma.tum.de

#### MS113

##### **Efficient, Matrix-based Approaches to Solving**

##### **Landscape Evolution Models in Parallel**

Landscape evolution models depend on computing river flows which cut into the surface and carry sediment to low ground. This generally requires the calculation of total upstream precipitation for every point on the land surface. This computation can be made very efficiently by traversing a graph of the downhill relationships between all the points. Commonly the graph-traversal algorithm is parallelised catchment-by-catchment which is conceptually extremely simple but problematic for landscapes dominated by a small number of large river systems. It is more obvious how to form a general parallel algorithm if we approach the problem from the equivalent matrix form from the start. Writing the graph traversal in terms of sparse-matrix multiplication allows the use of existing parallel libraries. It also allows generalisation of the flow calculation to cases with multiple flow pathways and to compute the effect of damming or the formation of swamps.

Louis Moresi  
School of Earth Sciences  
University of Melbourne  
louis.moresi@unimelb.edu.au

#### MS113

##### **Hybrid Operators and Composable Software within Lithospheric Dynamic Simulation**

We describe recent work synthesizing software components within a geophysical application. This work demonstrates the incorporation of several modern computational components into the FE application code pTatin3D for lithospheric dynamics. Our focus is to run realistic simulations at high efficiency on a hybrid CPU/accelerator supercomputer, solving the nonlinear Stokes equations with complex rheologies. We focus on development and composition of high-performance operator/smoothing application and integration into a large-scale rifting simulation.

Patrick Sanan  
USI - Università della Svizzera italiana  
ICS  
patrick.sanan@gmail.com

Dave A. May  
University of Oxford  
david.may@earth.ox.ac.uk

Karl Rupp  
Freelance Computational Scientist  
me@karlruupp.net

Guillaume Duclaux  
University of Bergen  
guillaume.duclaux@uib.no

#### MS114

##### **Realizing Extremely Large-Scale Scientific Applications Using Deeper Memory Hierarchy**

Abstract Not Available At Time Of Publication.

Toshio Endo  
Tokyo Institute of Technology

bar-

endo@is.titech.ac.jp

#### MS114

##### Optimization Problems Arising in Programming Environments Based on Data-Driven Tasking

Task-based programming environments have been proposed as a core technology for programming environments at exascale system size. We will discuss how avoidance of data movement, and exploiting the memory hierarchy in a data-centric programming paradigm relate to classic problems of mathematical optimization.

Utz-Uwe Haus  
CRAY EMEA Research Lab  
uhaus@cray.com

#### MS114

##### Distributed Solving of Large-scale LPs and MIPs from Energy System Models

Current LP and MIP energy system models frequently bring along problems of both high intricacy and large scale. The size and complexity of these problems often prove to be intractable even for commercial LP and MIP solvers. This talk describes an approach to tackle these problems by using massively parallel environments. For solving the LPs we have extended the parallel interior point solver PIPS-IPM to handle and efficiently exploit their structure (block-diagonal with linking-constraints and linking-variables). Although our main focus has been on LPs so far, we have also employed the UG framework to solve the MIP instances at hand. We will present first results for energy system problems from several European research institutes and universities.

Daniel Rehfeldt, Yuji Shinano  
Zuse Institute Berlin  
Takustr. 7 14195 Berlin  
rehfeldt@zib.de, shinano@zib.de

#### MS114

##### Toward Multiperiod AC Security Constrained Power Flow Optimization at Exascale

We present a multiperiod formulation of an AC security constraint optimal power flow. The parallel interior point solver for nonlinear programming serves as a scalable backend for our optimization problem. The two stage problem is expressed in Julia using the StructJuMP package for block structured problems. We present a medium sized grid with 24 periods and contingencies. Our goal is to run at scale on the exascale computer in 2021 at Argonne. For now we are targeting the KNL architecture on Theta and plan to run at full system size in the near future.

Michel Schanen  
Argonne National Laboratory  
mschanen@anl.gov

Mihai Anitescu  
Argonne National Laboratory  
Mathematics and Computer Science Division  
anitescu@mcs.anl.gov

Cosmin G. Petra  
Lawrence Livermore National  
Center for Advanced Scientific Computing

petra1@llnl.gov

#### MS115

##### Convergence Acceleration of PinT Integration of Advection Equation Using Accurate Phase Calculation Method

The convergence of the parareal method for hyperbolic PDEs such as the advection equation is worse than that of parabolic PDEs. It is reported that, theoretically, the poor convergence rate of the hyperbolic PDEs is identified by the beating of waves caused by a discrepancy of the phase between the fine and the coarse solver [1]. In this presentation, we show results of the convergence of the parareal method for the advection equation with cases of wide range parameters using the CIP (Constrained Interpolation Profile Scheme or Cubic-Interpolated Pseudo Particle method) and STRS-CIP (Space-Time Reversal Symmetry CIP) schemes that give us the accurate phase calculation.

Mikio Iizuka, Kenji Ono  
Kyushu University  
iizuka.mikio.903@m.kyushu-u.ac.jp, keno@cc.kyushu-u.ac.jp

#### MS115

##### Preliminary Application of Parareal for Two Phase Fluid Flow Simulation in Geological Media with Tough2 Code

CCS (Carbon dioxide Capture and Storage) is the promising approach for reducing the greenhouse gas emitting to the atmosphere by capturing carbon dioxide (CO<sub>2</sub>) and injecting it into deep reservoir rocks. We applied the parareal scheme to TOUGH2/ECO2N, which is a numerical code for multi-phase fluid flow in geologic media for predicting CO<sub>2</sub> behavior in the reservoirs. This study provides preliminary results of computational performance benefitted from the parareal with discussions on current issues.

Mitsuhiro Miyagi  
Taisei Corporation  
mygath00@pub.co.jp

Hajime Yamamoto  
aTaisei Corporation, 344-1 Nase-cho Totsuka-ku,  
Yokohama 245-  
hajime.yamamoto@sakura.taisei.co.jp

Toshiya Takami  
Oita University  
takami-toshiya@oita-u.ac.jp

Mikio Iizuka, Kenji Ono  
Kyushu University  
iizuka.mikio.903@m.kyushu-u.ac.jp,  
keno@cc.kyushu-u.ac.jp

Kengo Nakajima  
The University of Tokyo  
Information Technology Center  
nakajima@cc.u-tokyo.ac.jp

#### MS115

##### Accelerating Simulations of Fast Ion Trajectories

### in a Fusion Reactor

Computing trajectories of fast ions in a fusion reactor by solving the Lorentz equations numerically comes with substantial computational cost. To run simulations in a feasible amount of time, this cost needs to be mitigated by using tailored numerical methods and codes optimized for special hardware like graphic cards. The talk will describe Boris-SDC, a recent generalization of the popular Boris algorithm to arbitrary order, and its implementation and optimization for GPGPUs.

Daniel Ruprecht  
School of Mechanical Engineering  
University of Leeds  
d.ruprecht@leeds.ac.uk

Rob Akers  
Culham Centre for Fusion Energy  
rob.akers@ukaea.uk

Debasmita Samaddar  
CCFE  
UK Atomic Energy Authority  
Debasmita.Samaddar@ukaea.uk

Krasymyr Tretiak  
Kharkiv Institute of Physics and Technology  
krasymyr\_tretyak@yahoo.com

### MS115

#### Towards Parallel in Time Methods for Numerical Weather Prediction

We present recent work on a parareal algorithm for solving the rotating shallow water equations. These equations are the simplest to exhibit behaviour relevant for atmospheric modelling and as such are often used as a test bed for new numerical algorithms. The parareal algorithm has two ingredients: a cheap integrator (the coarse propagator) and a more accurate integrator (the fine propagator). The convergence rate, and hence efficiency, of this scheme is strongly affected by the accuracy of the coarse propagator. Our approach follows that of [Haut, T. and Wingate, B. An asymptotic parallel-in-time method for highly oscillatory PDEs, 2014] where it was shown that averaging the nonlinear terms over some fast oscillations includes the effects of near-resonances, essential for accuracy and hence convergence. A key component of this scheme is the computation of the exponential of the linear operator corresponding to the fast linear waves in the system. Here we use the rational approximation approach of [Haut, T. et al A high-order scheme for solving wave propagation problems via the direct construction of an approximate time-evolution operator, 2015]. This requires the solution of an elliptic problem for each term. The method is highly parallelisable as each term can be computed separately, but the solution of each term must be efficient, scalable and fast. In this talk, we describe how we construct a coarse propagator and present the latest results from various test cases.

Jemma Shipton  
Department of Mathematics  
Imperial College London  
j.shipton@ic.ac.uk

### MS116

#### Amt Runtime Scalability and Portability of Large-

#### Scale Parallel Applications Using Lessons from the Uintah Project

AMT codes and approaches such the Uintah software have been shown to both provide and offer the possibility of scalable and portable solutions to challenging large scale problems. Starting from lessons learned with Uintah and contrasting against other AMT approaches, the challenge of getting of getting new applications to work within these frameworks is addressed. AMT approaches also differ in how they address making use of third party software, which may be essential to the application. The approach taken by Uintah to achieve this will be described in the context of approaches used by other frameworks. Examples will be used based on a number of large scale problems and from a range of machines including DOE's Mira and Titan and the new Sunway Tiahulight machine in China.

Martin Berzins  
Scientific Computing and Imaging Institute  
University of Utah  
mb@sci.utah.edu

Alan Humphrey  
University of Utah  
ahumphre@sci.utah.edu

Dan Sunderland  
Sandia Labs Albuquerque  
dsunder@sandia.gov

Damodar Sahasbarude  
Univeristy of Utah  
damodars@sci.utah.edu

Zhang Yang  
IAPM Beijing  
yang\_zhang@iapm.ac.cn

### MS116

#### Enzo-P / Cello: A Charm++ Adaptive Mesh Refinement Astrophysics Application

Cello is a highly scalable "array-of-octree" based adaptive mesh refinement (AMR) software framework, implemented using Charm++, an object-oriented message-driven parallel programming system. Cello is being developed concurrently with Enzo-P, a branch of the ENZO astrophysics and cosmology application, that is built on top of the Cello scalable AMR framework. The Cello framework provides a scientific application with mesh adaptivity, data-driven ghost cell refresh, generic distributed field and particle data types, and sequencing of user methods for computing on block data. Further advanced software support, including dynamic load-balancing and checkpoint / restart, is provided directly by the Charm++ programming system. We give an overview of the Enzo-P / Cello / Charm++ software stack, discuss the design and implementation of Enzo-P / Cello, and present parallel scaling results for test problems up through 250K cores.

James Bordner  
University of California, San Diego  
jobordner@ucsd.edu

### MS116

#### Flecsale: A Task-Based, Multi-Dimensional Eulerian-Lagrangian Solver Built with the Flexible

### Computational Science Infrastructure (FLeCSI)

FLeCSALE is a computer software package developed for studying problems that can be characterized using continuum dynamics, such as fluid flow. It is specifically developed for existing and emerging large-scale, distributed-memory system architectures. For this purpose, it uses the Flexible Computational Science Infrastructure (FLeCSI) project for mesh and data structure support. We present implementation details and performance results for the Multi-Dimensional Eulerian-Lagrangian Solver built on top of FLeCSALE and FLeCSI.

Irina Demeshko, Ben Bergen  
Los Alamos National Laboratory  
irina@lanl.gov, bergen@lanl.gov

### MS116

#### Harnessing HPX to Run Billions of Tasks for a Scalable Portable Hydrodynamics Simulation of the Merger of Two Stars

We present a highly scalable demonstration of HPX a portable Asynchronous Many Task programming model and runtime system applied to a grid-based Adaptive Mesh Refinement hydrodynamic simulation of a double white dwarf merger with 14 levels of refinement that spans 17 orders of magnitude in astrophysical densities. The code uses the portable C++ Parallel Programming Model embodied in HPX that is being incorporated into the ISO C++ standard. The model represents a significant shift from existing bulk synchronous parallel programming models under consideration for exascale systems. Through the use of the Futurization technique, seemingly sequential code is transformed into wait-free asynchronous tasks. The simulation achieves a parallel efficiency of 96.8% on Cori (655,520 Knights Landing cores) using billions of asynchronous tasks, proving the applicability of our methodology.

Hartmut Kaiser  
Louisiana State University  
hkaiser@cct.lsu.edu

### MS117

#### Simulation of Volcanic Ash Transport Using Forest-Claw

As the eruption of Eyjafjallajökull demonstrated, "zero ash tolerance" policies imposed on civil aviation can lead to significant disruption for travelers and loss of revenue for airlines. To mitigate the hazards associated with volcanic eruptions, numerical volcanic ash transport and dispersal models are routinely used to forecast ash concentration loads in the atmosphere. However, delivering high fidelity predictions within operational time frames, especially for long-lasting eruptions that spread ash over wide areas with dense air traffic, places severe demands on computational resources. Ash3d, developed by the USGS, is one of several volcanic ash transport models in operational use. Available through a web-based portal, Ash3d solves a set of advection-diffusion-deposition equations to transport one or more classes of ash particles on a regional latitude/longitude grid. Current meteorological data is used to define wind fields for the transport equations and second order finite volume schemes are used to update the evolving ash plume. To improve the efficiency of the single grid, serial Ash3d code, we have ported Ash3d to our parallel, adaptive software library ForestClaw. We present results

showing the scalability of the Ash3d extension of Forest-Claw and discuss issues related to porting legacy codes to modern parallel, adaptive frameworks. Results will include tests to simulate the eruption of Mt. St. Helen's and the Icelandic volcano Eyja.

Donna Calhoun  
Boise State University  
donnacalhoun@boisestate.edu

Melody Shih  
Columbia University  
yoyoshih13@gmail.com

Hans Schweiger  
US Geological Survey  
Alaska Volcano Observatory  
hschwaiger@usgs.gov

Larry Mastin  
US Geological Survey  
Cascades Volcano Observatory  
lgmastin@usgs.gov

Roger Denlinger  
USGS Cascades Volcano Observatory  
Vancouver, WA  
roger@usgs.gov

### MS117

#### Multigrid Methods for Complex Fracture Networks in Porous Media

Numerical simulations of seismic waves in fractured rocks can result in significant advances for the indirect characterization of such environments. The discretisation of Biot's equations in the time-frequency domain gives rise to a generalised saddle-point problems with complex variables and allows to study attenuation and modulus dispersion due to fluid flow induced at different stimulation frequencies. We present an adaptive mesh refinement technique in order to simulate random fracture networks in cracked rocks. Starting from an uniform mesh of a representative elementary volume, a geometrical indicator is employed in order to refine elements intersecting fractures. This process is repeated multiple times until cracks are correctly resolved. We validate our approach on benchmark problems and show performances of a geometric multigrid solver where the intermediate meshes are employed in a multilevel solution process. We also study parallel scaling properties for this particular saddle-point problem employing point-block Gauss-Seidel smoothers.

Marco Favino  
Institute of Computational Science  
Università della Svizzera italiana  
marco.favino@usi.ch

Jürg Hunziker, Klaus Holliger  
Institute of Earth Sciences  
University of Lausanne  
jurg.hunziker@unil.ch, klaus.holliger@unil.ch

Rolf Krause  
Università della Svizzera italiana, Switzerland



rolf.krause@usi.ch

## MS117

### Free Boundary Solvers on a Forest of Octrees

We will present recent advances in the numerical solution of free boundary value problems on parallel architectures. In particular, we will present sharp numerical approximations of partial differential equations, including the level-set technology, on Cartesian adaptive grids encoded by Octree data structures.

Frederic Gibou, Arthur Guittet  
UCSB

fgibou@engineering.ucsb.edu, arthur.guittet@gmail.com

Mohammad Mirzadeh  
MIT

mirzadeh@mit.edu

Carsten Burstedde  
Universität Bonn  
burstedde@ins.uni-bonn.de

## MS117

### Applications of Space-Time Adaptive Mesh Refinement with a Block-Structured Higher-Order Cut Cell Framework

We present an adaptive, higher-order space-time finite volume cut cell discretization for PDEs on complex, moving domains. In the interior of the domain, traditional higher-order finite volume schemes are used, but near the domain boundary the finite volume stencils are modified for accurate and conservative representation of boundary or jump conditions. With careful generation of the stencils, higher-order convergence rates can be achieved, even when there are boundary kinks or topology changes or arbitrarily small cells appear. Cut cell stencil calculations are performed locally in parallel, with minimal communication and relatively high arithmetic intensity, which provides some advantages for our multigrid solver approach. We demonstrate results for model test problems, including Poisson, heat / Stefan, advection-diffusion-reaction, and Euler and Navier-Stokes equations. When combined with adaptive mesh refinement and semi-implicit time integration, it results in efficient and very accurate solution for a variety of PDE's.

Hans Johansen

Lawrence Berkeley National Laboratory  
Computational Research Division  
hjohansen@lbl.gov

Anneet Pal S. Bhalla  
Lawrence Berkeley National Laboratory  
apbhalla@lbl.gov

Dan Graves  
Lawrence Berkeley Laboratory  
dtgraves@lbl.gov

Phillip Colella  
Lawrence Berkeley National Laboratory

PColella@lbl.gov

## MS118

### A Nonlinearly Preconditioned Newton Method for a Hyperelasticity Problem

Inexact Newton is a popular technique for solving large sparse nonlinear system of equations. In this talk, we discuss a nonlinearly preconditioned inexact Newton method which is often more robust than the classical Newton when the nonlinearities in the system are not balanced. We show numerically that the method performs quite well for solving some nonlinearly difficult problems, such as the three-dimensional hyperelasticity equation which is used to model the wall of human arteries.

Shihua Gong  
university  
civar77@gmail.com

Xiao-Chuan Cai  
University of Colorado, Boulder  
Dept. of Computer Science  
cai@cs.colorado.edu

## MS118

### Scalable Newton-Krylov-Amg-Based Preconditioners for Implicit Continuum Plasma Simulations

Continuum modeling for plasma physics systems involves the solution of the governing PDEs describing conservation of mass, momentum, and total energy, along with various forms of Maxwells equations for the electromagnetic fields. We consider finite element methods on unstructured meshes and employ fully-coupled and block preconditioners for our algebraic multigrid-based preconditioned Newton-Krylov solution approach based on Trilinos solver library components. We examine solver performance and scaling on both CPU and Knights Landing platforms.

Paul Lin, John N. Shadid, Edward G. Phillips  
Sandia National Laboratories  
ptlin@sandia.gov, jnshadi@sandia.gov, egphill@sandia.gov

Eric C. Cyr  
Computational Mathematics Department  
Sandia National Laboratories  
eccyr@sandia.gov

Roger Pawlowski  
Multiphysics Simulation Technologies Dept.  
Sandia National Laboratories  
rppawlo@sandia.gov

## MS118

### Performance Comparison of GMG and BDD for Large Scale Numerical Simulation

We have been developing the open source CAE (Computer Aided for Engineering) system called ADVENTURE[ANDVENTURE project, <http://adventure.sys.t.u-tokyo.ac.jp>] since 1997. ADVENTURE is a finite element package and designed to analyze a model arbitrary shape over 100 million degrees of freedom based mainly on HDDM (Hierarchical Domain Decomposition Method) with BDD (Balancing Domain Decomposition) preconditioner. Towards future software architecture for exascale computing, our ADVENTURE group collaborates with

ExaSolvers groups of DFG SPPEXA program. UG4[UG4 framework, <http://gcsc.uni-frankfurt.de/simulation-and-modelling/ug4>] (Unstructured Grids 4.0) developed by ExaSolvers groups, is a powerful software framework for the solution of PDE (Partial Differential Equation) based models on high performance computers with GMG(Geometric Multi Grid) preconditioner. UG4 is written in C++ and can be used as a usual C++ library. Both ADVENTURE system and UG4 simulation framework have high parallel performance on the supercomputer with BDD or GMG. In this study, we used and compared those two free and open source software programs with preconditioner to solve the analysis of large scale problem with super computers.

Ryuji Shioya, Hongjie Zheng  
Toyo University  
shioya@toyo.jp, tei@toyo.jp

Hiroshi Kawai  
Tokyo University of Science-Suwa  
kawai063@toyo.jp

Amane Takei  
Faculty of Engineering  
University of Miyazaki  
takei@cc.miyazaki-u.ac.jp

## MS118

### A Reduced Iterative Domain Decomposition Method for Mixed Variational Formulations Derived from Magnetic Field Problems

An iterative Domain Decomposition Method (DDM) is applied into a mixed formulation of magnetic field problems like eddy current and magnetostatic ones. From the engineering point of view, these magnetic field problems are often formulated by neglecting the gauge conditions, where the magnetic vector potential is only one unknown function. The formulation without any gauge conditions enables us to reduce computational costs in case of the conventional one domain problem, and to formally introduce an iterative DDM. However, to the best of our knowledge, mathematical justifications of the numerical scheme without any gauge conditions such as unique solvability and convergency are not available. Therefore, we introduce a gauge condition and formulate magnetic field problems by mixed variational problems. The mixed formulation regards the magnetic vector potential and the Lagrange multiplier as two unknown functions that are approximated by the Nédélec curl-conforming edge element and by the conventional piecewise linear element, respectively. Then, we can introduce a new iterative DDM and can establish ones unique solvability and convergency. Moreover, we can reduce the iterative DDM into more efficient one by using the property of the Lagrange multiplier. Finally, some numerical results are shown in case of ultra-large computational models whose numbers of DOF are  $10^7$ – $10^9$ .

Daisuke Tagami  
Kyushu University, Japan  
tagami@imi.kyushu-u.ac.jp

Shin-ichiro Sugimoto  
Tokyo University of Science, Suwa  
sugimoto@rs.tus.ac.jp

## MS119

### Directive-based Vectorization Strategies in LLVM

### for Current and Upcoming Many-Core Processors

The Single Instruction Multiple Data (SIMD) programming model has enjoyed large success due to its simplicity and its potential for achieving significant performance improvements: executing a single operation over multiple disjoint data elements. This has the advantages of increased parallelism, good spatial locality and its ability to mask high latency memories. In particular, architectural support for SIMD Vectorization is pervasive throughout the HPC and commodity hardware landscape. Currently, this landscape is changing, and the major hardware vendors are aiming for a common goal: broadening the coverage of targetable codes. The introduction of novel architectural features such as speculative execution and predicated instructions proves so. However, some architectural innovations might have a negative impact. For instance, shrinking the reorder buffers (ROBs), will make loop unrolling ineffective, and careful statement scheduling increasingly important. In this talk we will discuss our latest advances on extending the automatic vectorization capabilities of the LLVM compiler infrastructure. More specifically, we will describe our ongoing work for handling irregular codes, the use of OpenMP as a performance portability vehicle, and new code generation strategies that allow to achieve both high productivity as well as improved performance.

Martin Kong, Lingda Li  
Brookhaven National Laboratory (BNL)  
mkong@bnl.gov, lli@bnl.gov

Barbara Chapman  
Stony Brook University (SBU)  
Brookhaven National Laboratory (BNL)  
barbara.chapman@stonybrook.edu

## MS119

### Omni Compiler: Compiler Infrastructure for Source-to-Source Translation

Most industry-level compilers and open-source compilers translate high-level source code into binary code or assembly code. However, there is still a demand for source-to-source compilers in research area, especially for parallel programming languages which does not focus on machine code generation and low-level optimization techniques. In this presentation, we will talk about our compiler infrastructure, named the Omni compiler and research activities using its source-to-source code translation. The Omni compiler have been developed by RIKEN AICS and the University of Tsukuba. The main objective is to develop and research advanced parallel languages. Currently the compiler supports OpenMP and XcalableMP in C and Fortran. C++ support at the front-end will be added in the near future. XcalableMP is a parallel programming language for cluster computing. It extends the base languages (C/Fortran) with directives, which has similar programming model to OpenMP. The directives specify parallelism in the source code including data/loop distribution among nodes. The Omni compiler translates OpenMP/XcalableMP directives into base language code with runtime calls. To enable source-to-source translation, the intermediate form preserves the high-level structures including loops and arrays in the base languages. A metaprogramming framework to help HPC application development is now under development by using the compiler features.

Jinpil Lee, Mitsuhsa Sato  
RIKEN Advanced Institute for Computational Science

jinpil.lee@riken.jp, msato@riken.jp

#### MS119

##### OpenARC: A Compiler Framework for Directive-based High-Performance Reconfigurable Computing

Heterogeneous computing with GPUs and Xeon Phis have recently become popular solutions for power-efficient high performance computing (HPC). Along these lines, reconfigurable computers, such as Field Programmable Gate Arrays (FPGAs), have had a long history of offering even more advantages in terms of performance and energy efficiency for specific workloads than other heterogeneous systems. Nevertheless, FPGAs have traditionally suffered several disadvantages that have limited their deployment in large scale HPC systems, mostly due to the challenges in terms of programmability and portability. In this talk, I will present a directive-based, high-level programming framework for high-performance reconfigurable computing. It takes a standard, portable OpenACC C program as input and generates a hardware configuration file for execution on FPGAs. We implemented this prototype system using our open-source OpenARC compiler; it performs source- to-source translation and optimization of the input OpenACC program into an OpenCL code, which is further compiled into a FPGA program by the backend Altera Offline OpenCL compiler. In addition, we show that our proposed FPGA-specific compiler optimizations and novel OpenACC pragma extensions assist the compiler in generating more efficient FPGA program. Evaluation on porting eight OpenACC benchmarks to an Altera FPGA and other accelerators, including NVIDIA/AMD GPUs and Intel Xeon Phis demonstrates the benefits of our strategy.

Seyong Lee  
Oak Ridge National Laboratory  
lees2@ornl.gov

#### MS119

##### Arm in HPC: Compilers and Tools

Arm and its partners are entering the HPC market with new CPUs and tools. The A64 instruction set *Armv8* has been extended to target HPC workloads with a new *Scalable Vector Extension (SVE)*, a set of SIMD vector instructions that enables *Vector Length Agnostic (VLA)* programming, which is used for writing portable binaries that work on multiple vector lengths. SVE has a bold impact on an auto-vectorizing compiler. Standard vectorization techniques rely on the knowledge of a fixed register length at compile time, an assumption that is not valid anymore for SVE VLA. The compiler provided in our HPC suite has been developed to support HPC users to produce auto-vectorized SVE code, and also to allow programmers to develop highly tuned computational kernels with the *Arm C Language Extensions for SVE*. The auto-vectorizer of the compiler has also been improved to extend the amount of loops that the compiler is able to vectorize with fixed-width vectorization targeting AArch64 *Advanced SIMD* vector extension. The HPC compiler can compile C, C++ and Fortran code. The HPC suite is also provided with a set of highly optimized HPC libraries, debuggers, and a code analysis tool that combines runtime performance data and static analysis to suggest improvements to the user code. An SVE user-space simulator is currently used for architectural exploration and performance analysis.

Francesco Petrogalli

ARM Ltd.  
francesco.petrogalli@arm.com

#### MS120

##### Parallel Preconditioning Techniques for Electromagnetic Problems

Firstly, we will introduce a PDE(Partial Differential Equation) based preconditioner for solving eddy current problems. Computational results of some large scale simulations with up to 1 billion unknowns and using up to 12280 CPU cores are presented to demonstrate that our toolbox is robust and scalable. And then the parallel source transfer domain decomposition method will be proposed to solve the scattering problem with high wave number.

Tao Cui  
LSEC, the Institute of Computational Mathematics  
Chinese Academy of Sciences  
tcui@lsec.cc.ac.cn

#### MS120

##### Efficient Low-rank Solver for Integral Equations on Distributed Memory Systems

Hierarchical Low-Rank Approximation (HLRA) methods, such as hierarchical matrices, can be used as approximation techniques for dense matrices coming from integral equations in scientific applications to reduce computational costs and memory usage. We consider the use of distributed memory computer systems to solve a linear system with a HLRA matrix as a coefficient matrix. To efficiently use a massive number of MPI processes, we have to balance the load and construct an efficient communication pattern. The complicated structure of HLRA-matrices prevents us from satisfying these requirements. Simplification of the structure can solve this problem. However, as a trade-off, memory usage of the approximated matrix is expected to increase. In this presentation, we discuss the trade-off problem.

Akihiro Ida  
The University of Tokyo  
ida@cc.u-tokyo.ac.jp

#### MS120

##### HiCMA: Hierarchical Computations on Manycore Architectures

FMM and H-matrices share a rare combination of optimally low  $O(N)$  arithmetic complexity and high arithmetic intensity (flops/Byte), with important phases being nearly compute-bound. This stands in contrast to workhorse solvers that have either low arithmetic complexity with low arithmetic intensity (e.g., Fast Fourier Transform and multigrid), or high arithmetic intensity with high arithmetic complexity (e.g., dense linear algebra and direct N-body summation). In short, fast multipole and H-matrix methods are mathematically efficient algorithms that are and likely will remain computationally efficient, in the sense of being compute-bound, on future architectures. Furthermore, when it comes to distributed memory applications, these methods have a communication complexity of  $O(\log P)$  for  $P$  processors, and permit high asynchronicity in their communication. They are therefore amenable to asynchronous programming models, which have been gaining popularity as purely bulk synchronous models appear stressed by energy-austere exascale hardware trends. Much

research exists on the mathematics of hierarchically low-rank methods and many high-performance implementations are available for traditional full rank methods. Their intersection on accelerators, where the hierarchy needs flattening, is HiCMA's target. Some modules of this open source project have already been adopted in the software libraries of major vendors. We describe currently available modules of this work in progress.

David E. Keyes  
KAUST  
david.keyes@kaust.edu.sa

Hatem Ltaief  
King Abdullah University of Science & Technology (KAUST)  
hatem.Ltaief@kaust.edu.sa

George Turkiyyah  
American University of Beirut  
gt02@aub.edu.lb

Rio Yokota  
Tokyo Institute of Technology  
rioyokota@gsic.titech.ac.jp

## MS120

### AutoMG: An Automatic Tuning Multigrid Procedure for the Sequences of Linear Systems

Many PDE based numerical simulations require to solve a sequence of linear systems, which involves lots of related systems of linear equations. For such sequences arising from the large-scale real world simulations, however, it is hard for the existing solvers to keep the tradeoff between robustness and parallelism due to the matrices properties in sequence change dynamically. In this work, we presented an automatic tuning multigrid preconditioning procedure, named as AutoMG, to balance the robustness and parallel scalability. The main idea of AutoMG is the introduction of automatic tuning mechanism for the components of smoothing and coarsening. For a given sequence, AutoMG allows us to customize specific components that adapting to the change of matrices properties in it. As a result, the smoothing and coarsening strategies, including the relaxation numbers and the required coarse levels, may change dynamically during the solution of the linear systems sequence, so that the parallel scalability is expected to be improved without losing the robustness for the ensemble solution. This talk will introduce the designing details of AutoMG, including an adaptive nested DD-type smoother based on nested domain partitioning, and an adaptive setup coarsening strategy based on the combination of geometric and algebraic hierarchy. The numerical results on tens of thousands of cores for a typical multi-physics simulation are given to show the efficiency of AutoMG.

Xiaowen Xu, Ran Xu, Hengbin An  
Institute of Applied Physics and Computational Mathematics  
xwxu@iapcm.ac.cn, xu\_ran@iapcm.ac.cn, an\_hengbin@iapcm.ac.cn

Zeyao Mo  
CAEP Software Center  
for High Performance Numerical Simulations

zeyao\_mo@iapcm.ac.cn

## MS121

### Design Examples of Accelerator-in-Switch

Accelerator-in-Switch (AiS) is a framework for building an accelerator logic tightly coupled with a switching hub in a single FPGA for high performance computation with heterogeneous environment with CPUs and GPUs. AiS is implemented on a partial reconfigurable region of an FPGA whose permanent region is used for a switching hub. A port of the switching hub is connected to the registers and local memory of AiS directly. AiS has a standard interface for a standard bus (Avalon MM bus, here) to exchange data between on-board DDR SDRAM and the local memory, and various types of accelerators can be implemented just by providing such an interface. The data input and output are done with the DMA controller inside the switching hub with the shared memory model between host CPUs and GPUs. Two example accelerators: a reduction calculator for a radiation transfer equations solver (RED) and LET generator for N-body simulation (LET) were implemented as the AiS on PEACH3, a switching hub for a PCIe direct interconnection network with Altera's Stratix V. By making use of the partial reconfiguration, multiple accelerators can be switched without stopping the switching hub. The time for place&route of an accelerator can be reduced by 47% compared to the case of the design combining the accelerator into the switching hub. The RED and LET achieved 45 times and 5.3 times performance of CPU/GPU execution, respectively.

Hideharu Amano  
Keio University, Tokyo, Japan  
hunga@am.ics.keio.ac.jp

## MS121

### Using OmpSs to Run Applications in FPGAs

The OmpSs programming model is a directive-based approach targeting the programming of heterogeneous architectures. It is based on tasking, very similarly to the tasking approach of OpenMP. OmpSs has two main components: the Mercurium compiler and the Nanos++ runtime system. They support an extension to OpenMP consisting of the annotation of functions as tasks. In this case any call to such a function is actually spawned as a task. We have used this extension to incorporate the ability to spawn tasks to GPUs, by writing the tasks in CUDA or OpenCL. Currently, we use the same approach to write and be able to spawn tasks in IP cores synthesized in FPGAs. In this talk we will examine the proposal we do to generate code for FPGAs. We target the Xilinx Zynq 7000 and Ultra-scale+ platforms. The compiler transforms the code based on the directives inserted by the programmer. The target and task directives, both very similar to the OpenMP approach, allow the compiler to offload the code for compilation with the Xilinx Vivado HLS compiler. The resulting VHDL code is incorporated on a base Vivado project to be synthesized into a bitstream. The Nanos++ runtime system deals with the data movements and management of the IP cores execution. During the talk we will examine a few code examples, and we will present the performance we get out of the FPGA.

Xavier Martorell  
Barcelona Supercomputing Center, Spain  
Computer Sciences

xavier.martorell@bsc.es

## MS121

### System Software for Future Heterogeneous, Reconfigurable Architecture

CMOS transistor scaling will come to an end within a decade, and no alternative switching technology is expected to be ready for deployment in a timely manner. Heterogeneity and reconfigurability are keys to improving both performance and energy efficiency. However, they pose significant challenges to programmability and usability because existing programming languages and system software stacks are designed for non-reconfigurable computing architecture or instruction-set architecture. In addition, data flow in reconfigurable architectures needs to be extended to host programming models in order to maximize the potential of reconfigurable architecture, which cannot be expressed naturally by discrete accelerator models used for lightly heterogeneous systems (e.g., CPUs+GPUs). In this talk, I will first present a summary of currently available programming models and runtime systems for field-programmable gate array (FPGA) platforms, which are the most practical reconfigurable architecture today. I will then present performance and energy efficiency obtained from the current generation of FPGA platforms on numerical simulation, data analytics, and machine learning and deep learning applications. I also will outline key challenges on future heterogeneous, reconfigurable architecture, focusing on programming and system-level abstraction, and discuss potential approaches, considering efficient data movement and seamless integration between CPUs and FPGAs.

Kazutomo Yoshii

Argonne National Laboratory  
kazutomo@mcs.anl.gov

Franck Cappello  
ANL  
cappello@mcs.anl.gov

## MS121

### Accelerating Stencil Computations Using OpenCL on FPGAs

High-Level Synthesis (HLS) is increasing in popularity and maturity. Where traditional designs used time-inefficient Hardware Descriptive Languages (HDLs), designs today outsource the hardware generation to HLS tools such as Intel SDK for OpenCL, Xilinx Vivado HLS/SDAccel and Maxeler MaxCompiler. Developers comfortably describe their hardware in software languages such as C/C++, and the HLS tool converts them into working hardware. It has been shown that using such HLS tools, FPGAs can achieve better power efficiency than GPUs for some applications, but can they rival GPUs in absolute performance as well? In this talk, we describe a parameterized and generic implementation of a Stencil Accelerator, written in OpenCL. We will discuss and empirically quantify pitfalls and opportunities in leveraging Intel FPGA SDK for OpenCL to create this accelerator. Our accelerator targets 2D and 3D stencils alike and, unlike existing work, places no restriction on the input size of the stencil. Our Stencil Accelerator exploits both Temporal and Spatial Blocking, and our performance model drives the parameter tuning to maximize performance. We show that FPGAs can indeed compete with high-end GPUs in terms of absolute performance for stencil computation. Furthermore, we estimate

that the upcoming Stratix 10 FPGAs will likely achieve higher performance than their same-generation GPUs in stencil computation.

Hamid Reza Zohouri, Artur M. Podobas

Tokyo Institute of Technology  
zohouri.h.aa@m.titech.ac.jp, artur.podobas@gmail.com

## MS122

### Network Design and Comparison Studies Using TraceR Simulations

The TraceR-CODES framework provides us with an extremely powerful simulation toolkit that can be used for performance prediction of parallel applications on current and future supercomputers using real execution traces. This enables a variety of network design, network comparison and what-if analysis studies involving job placement, task mapping and routing algorithms. In this talk, we will present studies conducted using TraceR to evaluate HPC networks, network configurations and job placements. We use execution traces of production DOE applications that are a part of the workload on DOE supercomputing centers.

Abhinav Bhatele, Nikhil Jain, Todd Gamblin

Lawrence Livermore National Laboratory  
bhatele@llnl.gov, nikhil@llnl.gov, tgamblin@llnl.gov

Laxmikant Kale  
University of Illinois at Urbana-Champaign  
kale@illinois.edu

## MS122

### Exploring HPC Storage and Interconnect Design Space Using CODES

With production HPC systems offering limited opportunities for studying the design of HPC systems, modeling and simulation are becoming popular choices for HPC design space exploration. In particular, parallel discrete-event simulation (PDES) can be used in a number of ways to answer "what-if" design questions. In this talk, we have focused on applying PDES to HPC interconnects, which are a key determinant of application performance. We will describe our experiences of using CODES discrete-event simulation framework to answer two questions. First, what is the impact of application communication interference, job and data placement policies on application performance? Second, what insights does PDES give us into modern interconnect architectures such as dragonfly, fat tree, slim fly and torus networks? The talk will conclude by identifying opportunities through which simulation can assist application users and HPC centers to carry out effective design space exploration.

Misbah Mubarak, Rob Ross

Argonne National Laboratory  
mmubarak@anl.gov, rross@mcs.anl.gov

Christopher Carothers  
Rensselaer Polytechnic Institute  
chrisc@rpi.cs.edu

Philip Carns, Matthieu Dorier  
Argonne National Laboratory  
carns@mcs.anl.gov, mdorier@anl.gov

Nikhil Jain, Abhinav Bhatele

Lawrence Livermore National Laboratory  
nikhil@llnl.gov, bhatele@llnl.gov

Noah Wolfe  
Rensselaer Polytechnic Institute  
wolfe@rpi.edu

## MS122

### Extending A Network Simulator for Power/Performance Prediction of Large Scale Interconnection Networks

Reducing the power consumption of the interconnection network in the HPC system is an important issue. As a method of reducing the power consumption of the interconnection network, there is a method of transitioning to the low power mode during the period when the packet is not processed. In this talk, we present Trace RP, which is an interconnection simulator supporting low power mode.

Takatsugu Ono, Yuta Kakibuka  
Kyushu University  
takatsugu.ono@cpc.ait.kyushu-u.ac.jp,  
yuta.kakibuka@cpc.ait.kyushu-u.ac.jp

Nikhil Jain, Abhinav Bhatele  
Lawrence Livermore National Laboratory  
nikhil@llnl.gov, bhatele@llnl.gov

Shinobu Miwa  
The University of Electro-Communication  
miwa@hpc.is.uec.ac.jp

Koji Inoue  
Kyushu University  
inoue@ait.kyushu-u.ac.jp

## MS122

### Software Defined Networking Applied to High Performance Computing

The OpenFlow-style Software Defined Networking (SDN) technology has shown promising performance in data centers and campus networks; and the High Performance Computing (HPC) community is significantly interested in adopting the SDN technology. In this work, we look at two applications of SDN to HPC, first to improve internal interconnect routing and secondly to allow coupled management of campus network SDN resources and HPC resources. Through multiple simulation tools, including application modeling, system level modeling, flow level simulation and low-level flit level simulation, we evaluate how SDN control features can enhance routing for Infiniband interconnects. Specifically, we show the advantages of SDN enabled Infiniband compared to conventional Infiniband on a fattree topology, and SDN enabled routing compared to adaptive routing on a dragonfly topology. To address the integration of SDN network and HPC resources, the project investigates merging the functionality of the SDN controller (openflow) with an Infiniband subnet manager (OpenSM). One motivating example would be to schedule the SDN campus network resources, a data producing experimental facility, and the required HPC processing power to enable coupling of large scale experimental resources and HPC clusters. This requires coordination of the tradition SDN network resources to build pathways from the experimental facility to the HPC analytics cluster and to various

storage resources.

Michael Lang  
Los Alamos National Laboratory  
mlang@lanl.gov

Xin Yuan  
Florida State University  
xyuan@cs.fsu.edu

## MS123

### Development of An Architecture-Aware Highly Scalable Software Package for Solving PDEs

In this talk, we introduce a new software package designed for solving partial differential equations frequently found in scientific and engineering computations. By developing this object-oriented software package, we strive to isolate the low-level architecture-related features and the high-level simulation functionalities by exploiting techniques such as automatic code generations and automatic performance optimizations. We will show with some applications in computational fluid dynamics that the presented software package can help take better advantage of the increasingly complicated supercomputer hardwares and at the same time ease the process of designing highly scalable and efficient simulation codes.

Yulong Ao  
Chinese Academy of Sciences  
yangchao@iscas.ac.cn

Chao Yang  
Software Institute, Chinese Academy of Sciences  
P. R. China  
yangchao@iscas.ac.cn

## MS123

### Solving Large Graphs in Parallel with Multigrid

In this talk we will discuss solving the Graph Laplacian on large social network graphs. The Laplacian matrix of a graph has variety of applications including sparse graph cuts and clustering. Social network graphs arising from large online communities have a couple of features that make solving their associated Laplacian challenging. We examine state of the art serial solvers and their limitations. We present our parallel solver which allows for solving very large Graph Laplacians on distributed memory machines.

Tristan Konolige, Jed Brown  
University of Colorado Boulder  
tristan.konolige@colorado.edu, jedbrown@colorado.edu

## MS123

### High-Performance Multiphysics Simulations for Ion Channels

The Poisson-Nernst-Planck (PNP) equation is a well-known model of ion transport describing many physical and biological phenomena. Due to ionic sizes, steric repulsion may appear in crowded ions of several biological systems, so the classical PNP equations become unreliable since the PNP equations represent ions as point particles of size zero. We consider a modified PNP model that can potentially include the steric repulsion ion-size effect and the selectivity of important types of potassium ion channels. I'll present spectral element schemes and implementation

for solving the PNP-steric models, and discuss implementation of unsteady and steady-state PNP solvers in the Argonne-developed scalable high-order software package, NekCEM. Discussion includes the treatments of Dirichlet, Neumann, non-flux, and Robin boundary conditions. I'll also demonstrate convergence studies for validating our schemes, provided with some preliminary results of the dynamics of charge concentrations in a real protein structure of KcsA potassium channel. Simulations are performed on more than thousands of CPU cores on the Argonne Leadership Computing Facility (ALCF).

Yu-Hsiang Lan  
National Taiwan University  
r05246001@ntu.edu.tw

MiSun Min  
Argonne National Laboratory  
Mathematics and Computer Science Division  
mmin@mcs.anl.gov

Paul Fischer  
University of Illinois at Urbana-Champaign  
fischerp@illinois.edu

Tai-Chia Lin  
National Taiwan University, Taiwan, ROC  
tclin@math.ntu.edu.tw

#### MS123

##### Simulation of Coupled Power-Grid and Water Networks on Extreme Scale Computers

A scientific framework for simulations of large-scale networks, such as is required for the analysis of critical infrastructure interaction and interdependencies, is needed for applications on exascale computers. Such a framework must be able to manage heterogeneous physics and unstructured topology, and must be reusable. We have developed DMNetwork, a class in PETSc software library that provides data and topology management and migration for network problems, along with multiphysics solvers to exploit the problem structure. It eases the application development cycle by providing the necessary infrastructure through simple abstractions to define and query the network. In this talk, we will present the design of the DMNetwork, illustrate its user interface, and demonstrate its ability to solve large coupled power-grid and water networks on extreme-scale computers.

Hong Zhang  
MCS, Argonne National Laboratory  
hzhang@mcs.anl.gov

Shrirang G. Abhyankar  
Argonne National Laboratory  
abhyshr@anl.gov

#### MS124

##### Performance of FDPs on Sunway TaihuLight and Other Modern HPC Systems

In this talk, we'll overview the performance of applications written using FDPS (Framework for Developing Particle Simulator) on various large HPC systems including Sunway TaihuLight. FDPS is based on the idea to separate the parallelization and most of performance from the description of the physical problem. Application developers define the

particle data structure and functions for particle-particle interaction, and then FDPS provide library functions for highly efficient parallelization. Thus, simple application programs run efficiently on large HPC machines.

Jun Makino  
Kobe University/  
Riken AICS  
jmakino@riken.jp

Masaki Iwasawa, Daisuke Namekata, Keigo Nitadori,  
Miyuki Tsubouchi, Long Wang  
RIKEN AICS  
cc67803@gmail.com, daisuke.namekata@riken.jp,  
keigo@riken.jp, miyuki.tsubouchi@riken.jp,  
long.wang@riken.jp

#### MS124

##### Parallelism and Performance Prediction in Molecular Dynamics and Molecular-Continuum Simulations

I introduce the topic of the minisymposium and describe (1) performance prediction for molecular dynamics (MD), and (2) parallelization of MD and ensembles. Concerning (1), I discuss regression of performance data on sparse grids. Concerning (2), I outline optimizations for MD on heterogeneous and accelerator architectures before diving into the distribution of MD ensembles on massively parallel systems in molecular-continuum simulations. The latter enable flow investigations across a wide range of spatial and time scales.

Philipp Neumann  
University of Hamburg, Germany  
Scientific Computing  
philipp.neumann@uni-hamburg.de

#### MS124

##### Fully Heterogeneous Load Balancing in Ls1 Mardyn

In molecular dynamics (MD) simulations, load balancing algorithms are essential to take heterogeneous particle systems, such as liquid-vapor systems, and heterogeneous architectures into account. However, few implementations exploit heterogeneous architectures, and basically none take both aspects into account. We present the *k*-d tree-based load balancing methods of the MD framework ls1 mardyn which automatically and dynamically resolve imbalances generated by heterogeneous architectures and heterogeneous particle distributions. Evaluations indicate performance improvements of up to 50%.

Steffen Seckler, Simon Griebel, Nikola P. Tchihev  
Technical University of Munich  
Department of Informatics  
steffen.seckler@tum.de, s.griebel@tum.de,  
tchihev@in.tum.de

Philipp Neumann  
University of Hamburg, Germany  
Scientific Computing  
philipp.neumann@uni-hamburg.de

Hans-Joachim Bungartz  
Technical University of Munich, Department of Informatics  
Chair of Scientific Computing in Computer Science



bungartz@in.tum.de

#### MS124

##### How Particle Methods Can Help Simulate the Structural Plasticity of the Brain at Scale

The neural network in the brain is not hard-wired. Even in the mature brain, new connections between neurons are formed and existing ones are deleted, which is called structural plasticity. The dynamics of the connectome is key to understanding how learning, memory, and healing after lesions such as stroke work. However, with current experimental techniques even the creation of an exact static connectivity map, which is required for various brain simulations, is very difficult. One alternative is to use simulation based on network models to predict the evolution of synapses between neurons, based on their specified activity targets. The Model of Structural Plasticity (MSP) by Butz et al. is an example of this approach. However, to predict which neurons connect to each other, the current MSP model computes probabilities for all pairs of neurons, resulting in a complexity  $O(n^2)$ . To simulate millions of neurons and beyond, this quadratic term is prohibitive. Inspired by hierarchical methods for solving n-body problems in particle physics, we propose a scalable approximation algorithm for MSP that reduces the complexity to  $O(n \log^2 n)$  without any notable impact on the quality of the results. We show that an MPI-based parallel implementation of our scalable algorithm can simulate the structural plasticity of up to  $10^9$  neurons - four orders of magnitude more than the naive  $O(n^2)$  version.

Sebastian Rinke  
Technische Universität Darmstadt  
rinke@cs.tu-darmstadt.de

Markus Butz-Ostendorf  
Biomax Informatics AG  
markus.butz-ostendorf@biomax.com

Marc-Andre Hermanns  
Juelich Aachen Research Alliance  
Forschungszentrum Juelich  
m.a.hermanns@fz-juelich.de

Mikael Naveau  
Simulation Laboratory Neuroscience  
Jülich Aachen Research Alliance, Forschungszentrum Jülich  
naveau@cyceron.fr

Felix Wolf  
Technische Universität Darmstadt  
wolf@cs.tu-darmstadt.de

#### MS125

##### Asynchronous Dual Decomposition for Stochastic Mixed-Integer Programs

We present an asynchronous bundle-trust-region (BTR) method for Lagrangian dual decomposition of stochastic mixed-integer programs. The asynchronous BTR method solves the Lagrangian master problem by using a bundle method with a trust-region constraint. The method enables asynchronous computations and can thus help mitigate load imbalance issues (associated with the solution of mixed-integer scenario subproblems) and improve parallel efficiency. We briefly describe the convergence proof

of the asynchronous method. Using the open-source parallel solver DSP, we present extensive numerical results on eighty instances of a large-scale stochastic unit commitment problem, and we demonstrate that the asynchronous approach significantly reduces the solution time in most problem instances.

Kibaek Kim  
Mathematics and Computer Science Division  
Argonne National Laboratory  
kimk@anl.gov

Cosmin G. Petra  
Lawrence Livermore National  
Center for Advanced Scientific Computing  
petra1@llnl.gov

Victor M. Zavala  
University of Wisconsin-Madison  
victor.zavala@wisc.edu

#### MS125

##### Parallel Quasi-Newton Methods for the Optimization of Complex Systems Using High-Performance Computing

We address the problem of parallelizing state-of-the-art nonlinear programming optimization algorithms. In particular, we focus on parallelizing quasi-Newton interior-point methods that use limited-memory secant Hessian approximations. Such interior-point methods are known to have better convergence properties and to be more effective on large-scale problems than gradient-based and derivative-free optimization algorithms. We target nonlinear and potentially nonconvex optimization problems with an arbitrary number of bound constraints and a small number of general equality and inequality constraints on the optimization variables. These problems occur for example in the form of optimal control, optimal design, and inverse problems governed by ordinary or partial differential equations, whenever they are expressed in a “reduced-space” optimization approach. We introduce and analyze the time and space complexity of a decomposition method for solving the quasi-Newton linear systems that leverages the fact that the quasi-Newton Hessian matrix has a small number of dense blocks that border a low-rank update of a diagonal matrix. This enables an efficient parallelization on memory-distributed computers of the iterations of the optimization algorithm, a state-of-the-art filter line-search interior-point algorithm. We illustrate the efficiency of the proposed method by solving structural topology optimization problems on up to 4,608 cores on a parallel machine.

Cosmin G. Petra  
Lawrence Livermore National  
Center for Advanced Scientific Computing  
petra1@llnl.gov

#### MS125

##### Assessing Performance of Parallel Branch-and-Bound Algorithms

Abstract: This talk focuses on assessing the performance of solvers employing a branch-and-bound algorithm. Performance measurement for such solvers presents many challenges, especially when the algorithms in question are parallelized. Parallel solvers can be assessed both on their raw performance and on their scalability. We describe the

challenges associated with performing "traditional" analyses for both scalability and performance, including that of choosing an appropriate test set. This may be exceedingly difficult both because solvers may have much different strengths and because we typically require instances to be solvable on a single core, which may be unrealistic. We focus on methods by which solvers with disparate capabilities can still be compared in terms of scalability and performance, as well as on methods by which unsolvable instances can still be incorporated into test sets.

Ted Ralphs  
Department of Industrial and Systems Engineering  
Lehigh University  
ted@lehigh.edu

Stephen Maher  
Lancaster University  
s.maher3@lancaster.ac.uk

Yuji Shinanno  
Zuse Institute Berlin  
shinano@zib.de

#### MS125

##### **Scalable and Provably Convergent Asynchronous Progressive Hedging for Scenario-Based Decomposition of Stochastic Convex Programs**

Abstract Not Available At Time Of Publication.

Jean-Paul Watson  
Sandia National Laboratories  
Discrete Math and Complex Systems  
jwatson@sandia.gov

#### MS126

##### **Parallel Multigrid Reduction in Time (MGRIT) with Application to Power Grid Simulations**

Since clock speeds are no longer increasing, time integration is becoming a sequential bottleneck. The multigrid reduction in time (MGRIT) algorithm is an approach for creating concurrency in the time dimension that can be exploited to overcome this bottleneck and is designed to build on existing codes and time integration techniques. Although MGRIT was originally designed for one-step methods, in this talk we present an approach for multi-step backward difference formula (BDF) integration of fully implicit differential algebraic equations (DAE) on adaptive variable timestep grids. Results from power grid applications, with and without discontinuities, will be presented. An overview of MGRIT and the corresponding open source XBraid library will also be given. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Robert D. Falgout  
Lawrence Livermore National Laboratory  
rfalgout@llnl.gov

Matthieu Lecouvez  
CEA  
matthieu.lecouvez@cea.fr

Jacob B. Schroder, Carol S. Woodward  
Lawrence Livermore National Laboratory

schroder2@llnl.gov, woodward6@llnl.gov

#### MS126

##### **A Spectral Coarse Solver for the Parallel-in-time Solution of the Time-periodic Eddy Current Problem**

A standard approach, commonly used for numerical simulation of electric machines, is based on the method of lines, i.e., spatial discretization of Maxwells equations, followed by a sequential time-stepping procedure of the resulting differential algebraic equations. During initial design stages, engineers only consider steady-state operating characteristics of electromagnetic devices. However, when an analyzed system features long settling time, one typically needs to perform a large number of (possibly nonlinear) solutions, which results in a highly time-consuming simulation. For this reason, the goal of our work is to accelerate convergence to the steady state, thereby shortening the computational time. The parareal algorithm allows parallelization of the time-marching procedure by employing discretization on coarse and fine grids. It was recently employed for the eddy current problem by the authors in [Schöps, "Parallel-in-time simulation of the eddy current problem using parareal", 2017]. In the present contribution, adapting to the time-periodicity of the desired steady-state solution, we will apply the parareal approach [Gander, "Analysis of two parareal algorithms for time-periodic problems", 2013] to the periodic problem of the semi-discrete eddy current model. Furthermore, we consider usage of spectral basis to be a suitable coarse-grid solution approach. Performance of the method is illustrated via application to a model of an electric machine.

Iryna Kulchytska-Ruchka  
Graduate School of Computational Engineering  
TU Darmstadt  
kulchytska@gsc.tu-darmstadt.de

Herbert De Gersem  
Institut Theorie Elektromagnetischer Felder, TU Darmstadt  
Graduate School Computational Engineering, TU Darmstadt  
degersem@temf.tu-darmstadt.de

Sebastian Schoeps  
Theorie Elektromagnetischer Felder (TEMF) and GSCE  
Technische Universität Darmstadt  
schoeps@gsc.tu-darmstadt.de

#### MS126

##### **PinT Computation of Swarm Behavior**

Among various group behaviors of animals, one of the most fascinating collective motion is observed in a flock of starlings, where we can see a sudden and synchronized change of motion. Some properties of the group can be simulated by a so-called BOID model with simple rules of local interactions. In this presentation, the parallel-in-time method is applied to swarm systems and self-propelled particle systems.

Toshiya Takami  
Oita University

takami-toshiya@oita-u.ac.jp

## MS126

### A Space-Time Multigrid Method for Electrophysiology

We present a parallel and efficient multilevel solution strategy for solving computational problems arising from electrophysiology. In particular we solve with a semi-geometric multigrid method a transient diffusion equation discretised with space-time finite elements. For the construction of coarse spaces in the spatial domain we use L2-projections. This approach allows representing the coarse levels of the multigrid hierarchy independently from the fine level mesh. Hence the simulation can be performed on arbitrary geometries. The multilevel hierarchy for the time domain is instead constructed with geometrically conforming 1D grids. While we use continuous finite elements in space, for stability reasons we adopt discontinuous elements in time (discontinuous Galerkin). Alternatively to discontinuous Galerkin we also employ artificial time diffusion, which allows us to reduce the number of degrees of freedom with the disadvantages of having to deal with user tuned parameters. By means of numerical experiments we illustrate the scalability and the convergence of our multilevel solution strategy and we compare the two time stabilisation techniques. We perform a spectral analysis of the discontinuous Galerkin approach in the space-time context. Finally, we discuss the applications of our method in electrophysiology and we describe our software framework.

Patrick Zulian

Università della Svizzera italiana, Switzerland  
patrick.zulian@usi.ch

Pietro Benedusi

Institute of Computational Science  
USI Lugano  
pietro.benedusi@usi.ch

Alessio Quaglino

Institute of Computational Science  
Università della Svizzera italiana  
alessio.quaglino@usi.ch

Rolf Krause

Università della Svizzera italiana, Switzerland  
rolf.krause@usi.ch

## PP1

### Dataset Characterization of Data Mining Algorithms

There are many applications which do not benefit from current computer architectures. Data mining is one such example. One reason is that the application often changes its behavior according to the input dataset. Similarly, it is not clear how the behavior changes as a function of the algorithm choice and the input dataset. In this talk, we will discuss the methodology used to characterize datasets in the context of program behavior for data mining applications.

Sayaka Akioka

Meiji University  
akioka@meiji.ac.jp

Suzanne M. Shontz

University of Kansas

shontz@ku.edu

## PP1

### Adoption of Less Synchronous Modes of Computation in Nonlinear Solvers

Additive-Schwarz Preconditioned Inexact Newton (ASPIN) is a domain decomposition method for solving large, sparse nonlinear systems of equations arising from the discretization of nonlinear partial differential equations. While ASPIN offers a robust reduced communication preconditioner choice, it exacerbates issues of load imbalance. We explore and analyze the performance of hybrid parallelization of ASPIN and load balancing using dynamic runtime system on petascale architecture.

Amani Alonazi

KAUST  
amani.alonazi@kaust.edu.sa

Lulu Liu

King Abdullah University of Science and Technology  
lulu.liu@kaust.edu.sa

Hatem Ltaief

King Abdullah University of Science & Technology (KAUST)  
hatem.Ltaief@kaust.edu.sa

David E. Keyes

KAUST  
david.keyes@kaust.edu.sa

## PP1

### A Proposal for a Nonlinear Semi-NMF Based Method with Bias Vectors and Regularization for Deep Neural Networks

Backpropagation has been widely used as a de-facto standard algorithm to compute weights for deep neural networks (DNNs). As another approach, we recently proposed a nonlinear semi-NMF method without bias vectors and regularization. In this poster, we propose an improvement of the nonlinear semi-NMF based method by considering bias vectors and regularization. Experimental results indicate that the proposed method shows higher recognition performance than the nonlinear semi-NMF based method.

Ryosuke Arai

University of Tsukuba  
arai@mma.cs.tsukuba.ac.jp

Akira Imakura, Tetsuya Sakurai

Department of Computer Science  
University of Tsukuba  
imakura@cs.tsukuba.ac.jp, sakurai@cs.tsukuba.ac.jp

## PP1

### Dynamic Load Balancing in Electric Propulsion Plume Simulations

Thermophysics Universal Research Framework (TURF) is a object-oriented framework that unifies the operators such that different numerical models and methods can be bridged easily. This framework is capable of simulating plasma and gas under a wide range of time and length

scales relevant to the spacecraft electric propulsion systems. One important application of TURF is the electric propulsion plume simulation to assess important spacecraft components, in which a plasma expands into a vacuum. In this simulation, heavy species are solved via particle-in-cell (PIC) method and electrons are solved as a fluid. Within a plume, the density drops by several orders of magnitude as moving away from the thruster; with constant weight particles, there is a large variation of simulation particle count per grid cell. TURF can be run in parallel by assigning different sub-domains to MPI processes during initialization; however, load balance is not guaranteed to be good, as the simulation runtime scales with the number of particles rather than the number of grid cells. The objective of this study is to accelerate the multi-core plume simulations by shuffling the sub-domains across different processes based on particle counts and dynamically balancing the loads on MPI processes as simulation progresses. TURF is designed to enable such operations since multiple sub-domains can belong to each MPI process. Load balancing will be accomplished using ParMETIS developed at University of Minnesota.

Samuel Araki  
ERC Inc. / Air Force Research Laboratory  
samuel.j.araki@gmail.com

## PP1

### Machine Learning Assisted Multiscale Modeling

The bulk rheology of a complex fluid is dictated by its dynamics of underlying microstructure. A recent work proposed that the bulk rheology can be actively learned from meso-scopic simulations where the dynamics of microstructure is explicitly computed. For example, a continuum model of polymeric fluids can be linked to a microscopic constitutive relation that is computed on-the-fly. The continuum solver provides transient flow fields to initiate meso-scopic simulations that can resolve individual polymers while the meso-scopic dynamics of polymers returns an effective constitutive relation to close the continuum equations. A multiscale modeling procedure is thus constructed with the cyclic process. An active learning scheme based on Gaussian Process Regression is used to narrow the simulation scope, thus minimize the necessity to perform expensive meso-scopic simulation. A demonstration of this framework that promotes on-the-fly micro- and meso-scopic coupling is presented.

Ansel Blumers  
Department of Physics  
Brown University  
ansel\_blumers@brown.edu

## PP1

### Sparse Iterative Solvers Review on Large Seismic Tomography

Seismic tomography is one of the main inversion tools providing approximated velocity model of the subsurface. The tomography workflow, based on ray shootings, leads to a sparse matrix system to solve. On recent seismic study, the matrix size easily reaches hundreds million lines by tens million columns. In this work, we performed a benchmark over the main highly parallel iterative solvers such as PETSc or HYPRE. Performance have been measured with industrial realistic data on the Total Pangea super-

computer.

Lionel Boillot  
TOTAL  
lionel.boillot@total.com

## PP1

### Task-Parallel Factorizations of Hierarchical Matrices Using OmpSs and Openmp

Large-scale systems of linear equations appear in many fundamental numerical simulations and also in recent methods for data analytics. Hierarchical matrices lie in-between dense and sparse scenarios. Therefore, it is natural to target linear algebra problems that involve these type of issues via task-parallelism, as this approach has reported successful results when solving dense and sparse linear systems. In our research, we pursue the development of task-parallel algorithms for hierarchical LU and Cholesky factorizations.

Rocío Carratalá-Sáez, José I. Aliaga  
UNIVERSITAT JAUME I  
rcarrata@uji.es, aliaga@uji.es

Enrique S. Quintana-Ortí  
Universidad Jaume I  
Spain  
quintana@icc.uji.es

## PP1

### A Performance Spectrum for Parallel Computational Frameworks that Solve PDEs

Important computational physics problems are often large-scale in nature, and it is highly desirable to have robust and high performing computational frameworks that can quickly address these problems. However, it is no trivial task to determine whether a computational framework is performing efficiently or is scalable. The aim of this work is to present various strategies for better understanding the performance of any parallel computational frameworks for solving PDEs. Important performance issues that negatively impact time-to-solution are discussed, and we propose a performance spectrum analysis that can enhance ones understanding of critical aforementioned performance issues. As proof of concept, we examine commonly used finite element simulation packages and software and apply the performance spectrum to quickly analyze the performance and scalability across various hardware platforms, software implementations, and numerical discretizations. It is shown that the proposed performance spectrum is a versatile performance model that is not only extendable to more complex PDEs such as hydrostatic ice sheet flow equations, but also useful for understanding hardware performance in a massively parallel computing environment. Potential applications and future extensions of this work are also discussed.

Justin Chang, Kalyana Nakshatrala  
University of Houston  
jychang48@gmail.com, knakshatrala@uh.edu

Matthew G. Knepley  
Rice University  
knepley@gmail.com

Lennart Johnsson  
University of Houston, USA  
KTH, Stockholm, Sweden

johnsson@cs.uh.edu

## PP1

### A User-Oriented Parallel Wave Propagation Solver for Geophysical Applications

Parallel computation is still often reserved to specialists of numerical computation. We describe in this poster a framework to allow non-specialists to perform very large scale simulations. The main issue is the quasi-automatic creation of the mesh, following both the demands of the user and numerical restrictions. Automation is made possible by the availability of topographical and material information in dedicated online databases. Scalability and versatility of the entire computation chain will be demonstrated.

Regis Cottereau, Lucio De Abreu Corrêa  
CNRS, CentraleSupélec, Université Paris-Saclay  
regis.cottereau@centralesupelec.fr,  
lucio.de-abreu-correa@centralesupelec.fr

José Camata  
NACAD - High Performance Computing Center  
COPPE/Federal University of Rio de Janeiro  
camata@nacad.ufrj.br

Alvaro Coutinho  
Dept. of Civil Engineering  
COPPE/Federal University of Rio de Janeiro  
alvaro@nacad.ufrj.br

## PP1

### Avoiding Communication in Proximal Methods for Convex Optimization

The iterative soft thresholding algorithm (ISTA) is used to solve convex regularized optimization problems in machine learning. Distributed implementations of the algorithm have become popular since they enable the analysis of large datasets. However, existing formulations of the algorithm communicate data at every iteration and thus have high communication overheads. The communication costs are closely tied to the mathematical formulation of the algorithm. In this work, we reformulate ISTA to communicate data in every “k” iterations and thus reduce the cost of communication in the algorithm when operating on large data sets. We formulate the algorithm for three different optimization methods on the Lasso problem and show that the latency cost is reduced by a factor of k while the bandwidth cost remains the same. The algorithm is re-organized so that to reduce synchronization steps while maintaining similar convergence rates and stability properties. The performance and scalability of the novel formulations for ISTA, FISTA (Fast Iterative Thresholding Algorithm), and proximal newton type methods are demonstrated on a distributed hardware platform.

Saeed Soori, Rachit Shah, Maryam Dehnavi  
Department of Electrical Engineering  
Rutgers University  
saeed.soori@rutgers.edu, rrs135@scarletmail.rutgers.edu,  
maryam.mehri@rutgers.edu

## PP1

### Large-Scale Estimation of Sparse Diagonally Dom-

### inant Inverse Covariance Matrices

We present a parallelized routine which utilizes sparse matrix technologies for estimation of large ( $>1$  million) sparse diagonally dominant inverse covariance matrices. This is achieved by eliminating the memory bottleneck of large datasets and leveraging the sparsity of the problem for performant matrix operations. Numerical examples indicate our implementation provides orders of magnitude speed-up with increased accuracy in comparison to state-of-the-art methods.

Aryan Eftekhari  
USI - Università della Svizzera italiana  
aryan.eftekhari@usi.ch

Olaf Schenk  
USI - Università della Svizzera italiana  
Institute of Computational Science  
olaf.schenk@usi.ch

## PP1

### Batched Factorization and Inversion for Iterative Solvers

Preconditioning is a crucial task for the efficient solution of large-scale sparse linear systems via iterative methods. The challenge is to find a preconditioner that accelerates their convergence. Preconditioners based on block-Jacobi are inherently-parallel and, therefore, highly appealing for massively-parallel systems. This poster summarizes the main results from our recent work towards the development of variable-size batched factorization, triangular solve, inversion and matrix-vector multiply routines on GPUs and their application in block-Jacobi preconditioning.

Goran Flegar  
Universitat Jaume I  
flegar@uji.es

Hartwig Anzt  
Innovative Computing Lab  
University of Tennessee  
hartwig.anzt@kit.edu

Jack J. Dongarra  
University of Tennessee, Oak Ridge National Laboratory,  
USA  
dongarra@icl.utk.edu

Enrique S. Quintana-Ortí  
Universidad Jaume I  
Spain  
quintana@icc.uji.es

## PP1

### Iterative Construction of Non-Symmetric Factored Sparse Approximate Preconditioners

Factored sparse approximate inverses (FSAI) play a key role in the efficient algebraic preconditioning of sparse linear systems of equations. For SPD problems remarkable results are obtained by building the FSAI non-zero pattern iteratively during its computation. Unfortunately, an equivalent algorithm still is missing in the non-symmetric case. In the present contribution we explore the possibility of iteratively computing FSAI for non-symmetric matrices by using an incomplete Krylov subspace bi-orthogonalization procedure. This technique is also compared with the idea

of directly minimizing the two norm of the off-diagonal row(/column) of the preconditioned matrix. This algorithm is intrinsically parallel as it computes the approximate inverse row(/column)-wisely, with each row independently computed from the others. In this preliminary work, we show the effectiveness of this preconditioner for Krylov subspace iterative methods, like BiCGstab and GMRES.

Andrea Franceschini, Carlo Janna, Massimiliano Ferronato  
University of Padova  
andrea.franceschini@dicea.unipd.it, janna@dmsa.unipd.it, ferronat@dmsa.unipd.it

## PP1

### Recent Progress on Low-Order Unstructured Implicit Finite-Element Earthquake Simulation on Oakforest-Pacs

Attaining high performance for low-order unstructured implicit finite-element methods used in earthquake simulations are challenging due to the randomness in data structure as well as the large memory transfer required. To circumvent this problem, we have been developing solver algorithms that reduce random access and memory transfer, which lead to high performance on the K computer. In this poster, we will show results porting this solver to second-generation Xeon-Phi based Oakforest-PACS in JC-AHPC.

Kohei Fujita, Tsuyoshi Ichimura  
Earthquake Research Institute  
The University of Tokyo  
fujita@eri.u-tokyo.ac.jp, ichimura@eri.u-tokyo.ac.jp

Masashi Horikoshi  
Software and Solutions Group  
Intel K.K.  
masashi.horikoshi@intel.com

Larry Meadows  
Data Center Group  
Intel Corporation  
lawrence.f.meadows@intel.com

Kengo Nakajima  
The University of Tokyo  
Information Technology Center  
nakajima@cc.u-tokyo.ac.jp

Muneo Hori, Lalith Madgededara  
Earthquake Research Institute  
The University of Tokyo  
hori@eri.u-tokyo.ac.jp, lalith@eri.u-tokyo.ac.jp

## PP1

### pFEM-CRAFT: A Library for Application-Level Fault-Resilience Based on the CRAFT Framework

Increasing number of compute nodes of the super-computer systems results in higher rate of failure. In the present work, we developed a library for application-level fault-resilience on supercomputers using the CRAFT(Checkpoint/ Restart and Automatic Fault Tolerance), which is a framework for fault recovery. The CRAFT framework provides recovery of communications by ULFM-MPI, and recovery of data by application-level checkpointing. Our library (pFEM-CRAFT) is for devel-

opment of fault-resilient FEM application using CRAFT framework. In parallel FEM codes using pFEM-CRAFT library, survived nodes read the data files, such as mesh files, and check-pointed files for results, which are originally owned by failed nodes. The data files read by the survived nodes are merged with files on the survived nodes. Finally, capability of repartitioning in the pFEM-CRAFT using ParMETIS provides load balance among compute nodes. Thus, we can continue the computation, even if the number of available compute nodes has been reduced after failure. In this poster, we will present the fundamental idea of pFEM-CRAFT, implementations on the Reebush-U system at the University of Tokyo, and experimental results by real applications.

Tatsuya Fukasawa  
Department of Electrical Engineering & Information Systems  
The University of Tokyo  
fukasawa@cspp.cc.u-tokyo.ac.jp

Faisal Shahzad  
Friedrich-Alexander Universitaet Erlangen-Nuernberg  
faisal.shahzad@fau.de

Kengo Nakajima  
The University of Tokyo  
Information Technology Center  
nakajima@cc.u-tokyo.ac.jp

Gerhard Wellein  
Department of Computer Science  
University of Erlangen-Nuremberg  
gerhard.wellein@fau.de

## PP1

### Principal Component Analysis with Large-Scale Electronic State Calculation for Ultra-Flexible Device Materials

The principal component analysis was carried out with large-scale electronic state calculation, so as to design ultra-flexible opt-electronics device materials. The descriptors were chosen to be the participation ratio of electronic wavefunctions, a measure of quantum localization. A classification problem was solved for 40,000 disordered organic polymer samples with 1,200 atoms or 3,597 descriptor elements on the K computer and Oakforest-PACS. The principal components indicate an insight on material design principle.

Takeo Hoshi, Hiroto Imachi, Yukiya Abe, Kentaro Oohira  
Tottori University  
hoshi@damp.tottori-u.ac.jp, lennmars@gmail.com, m16t8003b@edu.tottori-u.ac.jp, m16t8005x@edu.tottori-u.ac.jp

Koji Hukusima  
The University of Tokyo  
hukusima@phys.c.u-tokyo.ac.jp

## PP1

### Using the Moment to Reduce Linear System Size

We have seamlessly integrated Sakurai-Sugiura-type methods into our BEAST framework of parallel subspace iteration-based eigensolvers. Using moments with a constrained subspace size reduces the number of costly linear

solves that are necessary for updating the subspace. Together with adaptive schemes, this reduces the overall cost while achieving accuracy comparable to unconstrained subspace size or FEAST. Numerical results show the parallel performance and the robustness of the proposed approach.

Sarah Huber  
University of Wuppertal  
shuber@math.uni-wuppertal.de

Yasunori Futamura  
University of Tsukuba  
futamura@cs.tsukuba.ac.jp

Martin Galgon  
University of Wuppertal  
galgon@math.uni-wuppertal.de

Bruno Lang  
Bergische Universitat Wuppertal  
lang@uni-wuppertal.de

Tetsuya Sakurai  
Department of Computer Science  
University of Tsukuba  
sakurai@cs.tsukuba.ac.jp

## PP1

### Efficient Parallel Implementation of Fill-Reducing Ordering Using Spectral Nested Dissection for Large-Scale Sparse Linear Systems

We consider solving large-scale sparse linear systems in a variety of scientific calculations. Fill-reducing ordering is used to reduce the memory requirement and computational complexity in sparse direct solvers. However, it is difficult for implementing parallel methods of fill-reducing ordering efficiently. In this presentation, we show a parallel implementation of the spectral nested dissection that can be accelerated using an efficient sparse matrix-vector multiplication. We compare the performance of our implementation with the widely used METIS.

Yuta Inagawa, Yasunori Futamura  
University of Tsukuba  
inagawa@mma.cs.tsukuba.ac.jp,  
futamura@cs.tsukuba.ac.jp

Tetsuya Sakurai  
Department of Computer Science  
University of Tsukuba  
sakurai@cs.tsukuba.ac.jp

## PP1

### Large-Scale Dynamic Graph Processing on Hpc Systems

In many graph applications, the structure of the graph changes dynamically over time and may require real time analysis. However, most prior work for processing large graphs on HPC has not focused on dynamic graphs, rather static only. To address this issue, we have developed data structures and infrastructure management necessary to support dynamic graph analysis at large scale on distributed HPC platforms. We demonstrate its performance scaling out to manage large scale-free graphs.

Keita Iwabuchi, Roger Pearce

Lawrence Livermore National Laboratory  
iwabuchi1@llnl.gov, rpearce@llnl.gov

## PP1

### Fast Anomaly Diagnosis in HPC Systems Using Abnormal Subspace Sparse PCA Models

Understanding performance and resiliency at both system and application levels is essential and challenging in extreme-scale HPC environments. Towards this goal, logging systems and analytic tools based on them have been proposed and deployed in HPC systems. We propose an anomaly detection and diagnosis model, Abnormal Subspace Sparse Principle Component Analysis (ASPCA) model, which can serve as a means of locating root causes of performance degradation at the system level based on system logs. ASPCA model is self-explanatory and discriminative when detecting performance degradation with diversified root causes, and can be trained within minutes using the alternating direction method of multipliers (ADMM) method when applied to thousands of features extracted from system logs. As a case study, we applied our ASPCA model to Sunway TaihuLight supercomputer to diagnose I/O performance degradation on I/O proxy nodes. Based on the features extracted from Lustre system logs, our model successfully identified the object storage targets (OSTs) that were responsible for I/O performance degradation, and suggested that losing object storage serve (OSS) and cache malfunction were the root causes.

Wenming Jiang, Ying Zhao  
Tsinghua University  
CST  
jiangwm13@163.com, yingz@mail.tsinghua.edu.cn

## PP1

### Parallel Symmetric Tridiagonal Eigensolver Based on Bisection and Inverse Iteration Algorithms with the Blocked Classical Gram-Schmidt Reorthogonalization and the Shifted CholeskyQR Decomposition

We propose an OpenMP-based parallel symmetric tridiagonal eigensolver based on bisection and inverse iteration algorithms with the blocked classical Gram-Schmidt\*2 reorthogonalization and the shifted CholeskyQR\*2 decomposition. We can easily parallelize the bisection method. However, it is difficult to parallelize the inverse iteration algorithm. Thus, we use the blocked inverse iteration algorithm. The algorithm requires the reorthogonalization procedure and the QR decomposition. For the purpose of the reorthogonalization procedure, we adopt the blocked classical Gram-Schmidt\*2 reorthogonalization. In order to compute the QR decomposition, we employ the QR decomposition based on classical Gram-Schmidt\*2 algorithm or the shifted CholeskyQR\*2 decomposition. In the shifted CholeskyQR\*2 decomposition, we can choose two kinds of shift strategies. One is based on the numerical validation method. The other is based on diagonal elements of the Cholesky decomposition for symmetric matrices. We evaluate the QR decomposition based on classical Gram-Schmidt\*2 algorithm and the shifted CholeskyQR\*2 decomposition with these two kinds of shift strategies in the blocked inverse iteration algorithm.

Kinji Kimura, Masayuki Osawa  
Department of Applied Mathematics and Physics,  
Graduate School of Informatics, Kyoto University  
kimura.kinji.7z@kyoto-u.ac.jp,



osawa.masayuki.36z@st.kyoto-u.ac.jp

Yoshimasa Nakamura  
Graduate School of Informatics, Kyoto University  
ynaka@i.kyoto-u.ac.jp

## PP1

### A Fast and Efficient Preconditioning Method for Solving Ill-Conditioned Linear Systems by Partly Using LU Factors

We propose a preconditioning method for solving ill-conditioned linear systems with relatively large singular values, which partly utilizes a result of LU factorization efficiently. Our proposed method requires less computational cost than the previous preconditioning methods. In particular, if the coefficient matrix has a few relatively large or small singular values, the cost can be significantly reduced. We show some numerical results showing the effectiveness of the proposed method.

Yuka Kobayashi, Takeshi Ogita  
Tokyo Woman's Christian University  
yukako.109b@gmail.com, ogita@lab.twcu.ac.jp

## PP1

### Research of Finding Formula That Express Data Without Any Background Information

The purpose of this research is to find useful means to estimate the expression formula that can express observation data without background information. If all the form of functions included in the observed data can be identified, it is possible to interpret the phenomena that the data express. We are currently seeking good approach in the deep learning framework and to improve the discrimination capability for the learning process of equations.

Issei Koga  
The Graduate School and Faculty of Information Science and E  
21E17031W@s.kyushu-u.ac.jp

Kenji Ono  
Kyushu University  
keno@cc.kyushu-u.ac.jp

## PP1

### Locally Recursive Non-Locally Asynchronous Algorithms for FDTD Method on New GPGPU Cluster Architectures

Locally Recursive non-Locally Asynchronous algorithms were developed to maximize the performance for a given numerical scheme on a given computer system. It is achieved by space-time decomposition of the dependency graph of the problem into the polyhedral shapes of optimal data locality. The method includes a theory, that provides performance estimates for the given environment, gives optimal algorithm parameters and implementation guidelines. It includes an extension of the roofline model. A number of codes use this method. For example, there are working solutions for the elastic wave propagation, plasma physics, gas dynamics. DTmaxwell4 code for modeling of electromagnetic wave propagation was previously tested on the TSUBAME2.5 supercomputer [A. Zakirov, V. Levchenko, A. Perepelkina, Y. Zempo, High performance FDTD algorithm for GPGPU supercomputers, J. Phys.:

Conf. Seri., 759, 012100 (pp1-6), 2016] and showed good performance and scalability. The LRnLA theory estimates were verified. New computer systems based on GPGPU emerge, so in the current work, we analyze how our solution adapts to other supercomputer architectures. Namely, if the existing solution is ported to TSUBAME3.0, the expected acceleration is 3 time per one GPU, 5 times per one node. The linear scaling is expected not only for the weak scaling, but also for the strong scaling up to 500 nodes.

Vadim Levchenko, Anastasia Perepelkina  
Keldysh Institute of Applied Mathematics  
lev@keldysh.ru, mogmi@ya.ru

Anastasia Perepelkina  
Keldysh Institute of Applied Mathematics RAS  
mogmi@ya.ru

Andrey Zakirov  
Kintech Lab Ltd.  
zakirov@kintechlab.com

Yasunari Zempo  
Hosei University  
zempo@hosei.ac.jp

## PP1

### Effect of Algebraic Block Multi-Color Ordering for Multi-Threaded ILU-GMRES Solver

Algebraic block multi-color ordering is known as a parallelization method for a sparse triangular solver. In the previous work, we confirmed the effectiveness of the method in a multi-threaded ICCG solver for a linear system with a symmetric coefficient matrix. In this study, we enhance the method so as to deal with an unsymmetric coefficient matrix. We develop a multi-threaded ILU-GMRES solver based on the enhanced method and evaluate its performance in terms of both the runtime and the number of iterations.

Senxi Li  
Graduate School of Information Science and Technology  
Hokkaido University, Japan  
lisenxi@eis.hokudai.ac.jp

Takeshi Fukaya, Takeshi Iwashita  
Hokkaido University  
fukaya@iic.hokudai.ac.jp, iwashita@iic.hokudai.ac.jp

## PP1

### Numerical Simulation of Density-Driven Groundwater Flows with a Free Surface in Complicated Geometries

Presence of unsaturated zones in aquifers influences the density-driven haline flow essentially. Macroscopically, this interface between the saturated and the unsaturated parts can be modeled as a free surface which is tracked with the flow. We consider it as a moving boundary and use the ghost-fluid method to impose the boundary conditions for the main flow model. This technique is applied to the real hydrogeological formations. Because of the complicated, anisotropic, layered geometry and curved boundaries, unstructured grids consisting mainly of prisms are used. The very detailed structure of the domain assumes a high resolution in the simulation of the flow, so that the parallelization is necessary. The model is discretized by a finite

volume method. For the solution of the linearized systems, the geometrical multigrid method with ILU smoothing is used. In the talk, we present results of the parallel computations.

Dmitry Logashenko  
Extreme Computing Research Center  
King Abdullah University of Science and Technology  
dmitry.logashenko@kaust.edu.sa

Sebastian Reiter  
Goethe-Center for Scientific Computing  
Goethe-Universität Frankfurt  
sebastian.reiter@gcsc.uni-frankfurt.de

Gabriel Wittum  
Extreme Computing Research Center  
King Abdullah University of Science and Technology  
gabriel.wittum@kaust.edu.sa

### PP1

#### Efficient Reduction of the Generalized HPD Eigenproblems

We consider the reduction of the generalized hpd eigenproblem  $Ax = \lambda Bx$  to a standard one. Following the approach taken in the highly efficient ELPA library, two multiplications with the explicitly computed inverse of  $B$ 's Cholesky factor are come out. We present variants for these multiplications featuring significantly better scalability and higher performance than the current ELPA and ScaLAPACK reductions. The corresponding backtransformation of the eigenvectors is addressed as well.

Valeriy Manin, Bruno Lang  
Bergische Universität Wuppertal  
manin@math.uni-wuppertal.de, lang@uni-wuppertal.de

### PP1

#### Geodynamo Simulations and Visualizations Using Calypso

It is widely accepted that the geomagnetic field is generated by the convection of the liquid iron alloy in the Earth's outer core, so called dynamo process. Numerical simulations have had a large role to understand the dynamo process in the outer core. Geodynamo simulations require extremely fine spatial resolution on a massively parallel computer because the flow in the outer is expected to have a vast length scale. We have developed Calypso, which is for magnetohydrodynamics (MHD) simulation in a rotating spherical shell. Calypso uses spherical harmonics expansion and finite difference methods are used for the discretization, and is parallelized in two directions with changing the direction of the parallelization during the spherical harmonics transform. To visualize simulation results by Calypso, we implement two modules for visualization based on the visualization tools in GeoFEM. One is the sectioning and isosurfacing. The sectioning module can generate data on arbitrary quadratic surface. The sectioned data have a capability to perform data analysis on the sections. The volume rendering is also implemented in Calypso. The volume rendering routine generates one bitmap image for one scalar field. Consequently, the volume rendering data is not suitable for the quantitative analysis, but suitable to look the evolution of the fields. The volume rendering keeps good scaling to 4096 cores on TACC stamped.

Hiroaki Matsui

University of California, Davis  
hrmatsui@ucdavis.edu

### PP1

#### Dynamical Interaction Between Earth's Core and Mantle with Massively Parallel Computing

Simulating over 4 billion years of evolution of Earth's deep interior, the mantle convection is dominated for its time-scale because the time-scale of core dynamics is much shorter than that of mantle dynamics. However, a time-scale issue on dynamical interaction between numerical geodynamo and mantle convection simulations should be resolved in the future. Here, as a first step, I develop a static-coupling model between geodynamo and mantle convection simulations, which means that heterogeneous boundary conditions computed from mantle convection simulations with plate reconstruction data (200 Ma) with each a few 100s Ma are incorporated into numerical geodynamo simulations. Because of the time-scale difference between geodynamo and mantle convection, the geodynamo simulations are computed until reaching numerically stable solutions, which can create geomagnetic evolution data up to 200 Ma to present and be consistent with history of mantle dynamics. More details will be given in the presentation, which will include the correlation between geomagnetic secular variations and heat flow across the core-mantle boundary or plate tectonics history as well as computational efficiency analysis of geodynamo and mantle convection simulations in between CPUs and many integrated core architectures. In addition, I will discuss the future availability of many integrated core architectures such as Xeon Phi processors, K-computer and so on.

Takashi Nakagawa  
Japan Agency for Marine-Earth Science and Technology  
tnakashi@jamstec.go.jp

### PP1

#### Performance Evaluation of Hierarchical Matrix Computation on Various Modern Architectures

H-matrices is receiving a lot of attention as a computational method which can reduce the memory amount of dense matrix computation. However, because H-matrices is more complex computation than both dense and sparse matrix computation, acceleration of H-matrices is needed. Therefore, we are engaged in optimizing H-matrices on various modern computer environment. In this poster, we focus on the implementation of H-matrices on one computational node and show the performance of H-matrix computation on multi-core CPU, many-core architecture, and GPU and compare the performance aspect of them.

Satoshi Ohshima  
The University of Tokyo  
Information Technology Center  
ohshima@cc.kyushu-u.ac.jp

Ichitaro Yamazaki  
University of Tennessee, Knoxville  
iyamazak@icl.utk.edu

Akihiro Ida  
The University of Tokyo  
ida@cc.u-tokyo.ac.jp

Rio Yokota  
Tokyo Institute of Technology

rioyokota@gsic.titech.ac.jp

## PP1

### Accelerating Simulations of Cerebrovascular Blood Flow Through Parallelization in Time

Computer simulations of blood flow in arteries can help to understand and treat a range of different cardiovascular diseases. The Lattice-Boltzmann code HemeLB has a long history of successful use in computational hemodynamics, but simulations on complex geometries like the Circle of Willis still take several days to run. The poster presents ongoing work to incorporate parallel-in-time capabilities into HemeLB to extend parallel scaling and bring simulation times down to a day or less.

Daniel Ruprecht  
School of Mechanical Engineering  
University of Leeds  
d.ruprecht@leeds.ac.uk

Derek Groen  
Brunel University London  
djgroennl@gmail.com

Rupert Nash, David Scott  
EPCC, University of Edinburgh  
r.nash@epcc.ed.ac.uk, d.scott@epcc.ed.ac.uk

## PP1

### Efficient Robust Multi-Shift Triangular Solves

Eigenvectors for a matrix in real Schur form can be computed by solving many quasi-triangular linear systems of the form  $(T - \lambda I)x = b$ . If eigenvalues are clustered, then overflow may occur. Previously proposed algorithms update off-diagonal blocks with xGEMM, but the diagonal blocks may become a bottleneck. Our improvements include (1) minimal scaling behaviour, (2) underflow protection of scaling factors, (3) efficient overflow protection logic. We present performance results for KNL and Broadwell.

Angelika Schwarz  
Umeå University  
angies@cs.umu.se

Mirko Myllykoski  
Umeå University  
mirkom@cs.umu.se

Carl Christian Kjelgaard Mikkelsen, Lars Karlsson  
Umeå University, Dept. of Computing Science  
spock@cs.umu.se, larsk@cs.umu.se

## PP1

### Application of Game Theory to Correct Economics Decision

The object of this article is to demonstrate the possibilities of game theory as an instrument for study of economy. The approach to be used describe elementary games theoretic models as on integral part of economy with a collection of example to understand subject better. This paper addressed to theoreticians and practitioners of economy not particularly versed in games theory, rather than to those who are fluent in its mathematical language and intricacies. And we will give example that how much game

theory is more accurate than economy science.

Saeed Seyed Agha Banihashemi  
school of international relations  
ihusaied2001@yahoo.com

## PP1

### Development of Multi-Scale, Multi-Physics Methods for Multi-Fluid Plasma Simulations with Application to In-Space Propulsion

A framework for advanced multi-scale and multi-fidelity modeling and simulation of plasmas is presented. The multi-fluid plasma model is used due to its computational efficiency and capability to accurately represent multi-fluid effects when the spatial scales of the problem are comparable with the Larmor radius, and the temporal scale are in the order of the inverse of the ion cyclotron frequency. Propulsion devices, particularly the Electrodeless Lorentz Force (ELF) thruster relies on the formation and acceleration of a Field Reversed Configuration (FRC). The challenges associated with modeling such plasmas are considerable. Multi-physics aspects are present due to coupling of the electromagnetic fields, fluid variables, ionization/recombination physics, radiative losses, and kinetic effects all need to be accounted for in order to properly produce results that can be used as comparison to experimental data. In the framework employs a discontinuous Galerkin numerical method, and to tackle the multi-scale nature of the problem IMEX schemes are implemented in order to address the several orders of magnitude associated with the different physics. Results studying the formation or the FRC are presented, and compared to experimental data. Distribution A: Approved for public release; distribution unlimited; Clearance No. 17575.

Eder M. Sousa  
ERC Inc.  
Air Force Research Lab  
eder.sousa.ctr@us.af.mil

Robert Martin  
Air Force Research Lab  
robert.martin.101@us.af.mil

## PP1

### Toward Fault-Tolerant Parallel-in-Time Integration with Pfasst

Resilience and concurrency are important properties for numerical algorithms to possess in order to ensure sustained efficiency on current and future HPC architectures. Parallel-in-time integration methods can improve scaling of numerical codes for time-dependent partial differential equations beyond their traditional spatial scaling limits. They also possess certain characteristics that should allow them to provide algorithm-based fault tolerance. The poster will present results of an analysis studying how the PFAST algorithm can be augmented to allow recovery of data lost due to node failure.

Robert Speck  
Juelich Supercomputing Centre  
Forschungszentrum Juelich GmbH  
r.speck@fz-juelich.de

Daniel Ruprecht  
School of Mechanical Engineering  
University of Leeds

d.ruprecht@leeds.ac.uk

## PP1

### Convergence Property and Accuracy Improvement of Block Bicgstab Class Solvers for Linear Systems with Many Right-Hand Sides

Block Krylov subspace methods are efficient methods for solving linear systems with multiple right-hand sides. However, when the number of right-hand sides is large, convergence property and accuracy of approximates solutions of Block Krylov subspace methods often deteriorate. In this presentation, we will show that these drawbacks of Block BiCGSTAB class solvers can be improved by using QR factorization appropriately. Moreover, parallel performance of the methods will also be shown in the presentation.

Hiroto Tadano

Department of Computer Science  
University of Tsukuba  
tadano@cs.tsukuba.ac.jp

## PP1

### Dynamic Task Scheduling Implementation of Tile QR Decomposition on CPU/GPU Heterogeneous Cluster System

The tile algorithm for matrix decomposition can generate a large number of fine-grained tasks that can be executed in parallel. Because the sufficient number of tasks can resolve load imbalances and reduce the number of idle cores, the tile algorithm is suitable for modern multi-core architectures. Although the small amount of calculation of small tasks could not fully utilize the performance of GPUs, the recent GPUs can perform parallel batch execution of small tasks. We implement the tile QR decomposition on the CPU/GPU hybrid cluster system with the dynamic task scheduling using OpenMP 4.0. In this poster, we show our implementation in detail and the performance result.

Masatoshi Takayanagi, Tomohiro Suzuki

University of Yamanashi  
g16dm002@yamanashi.ac.jp, stomo@yamanashi.ac.jp

## PP1

### Contour Integral-Based Verified Computing for Partial Eigenvalues

We propose a contour integral method using the block SS-Hankel method for verified partial eigenvalue computations of a real symmetric matrix. The method requires solutions of the linear systems at quadrature points, which are computed in parallel. Upper bounds of truncated and rounding errors of a numerical quadrature give the rigorous enclosure of the complex moments. By rigorously enclosing eigenvalues of Hankel matrices, whose entries are the complex moments, our method yields verified partial eigenvalues.

Akitoshi Takayasu

Waseda University  
takitoshi@risk.tsukuba.ac.jp

Akira Imakura, Keiichi Morikuni  
Department of Computer Science  
University of Tsukuba

imakura@cs.tsukuba.ac.jp, morikuni@cs.tsukuba.ac.jp

## PP1

### Performance Evaluation of Large-Scale Deep Learning Framework ChainerMN on Oakforest-PACS

Deep Learning is emerging technology for wide field of application, and demand of computation resources for deep learning increases exponentially. In order to shorten a huge-amount training for deep learning, HPC systems are eligible in terms of total computation performance, the bandwidth of interconnect, and the IO performance for the storage. In this presentation, performance evaluation of a huge-amount training deep learning framework ChainerMN is demonstrated on Oakforest-PACS system operated by JCAHPC.

Kohei Tamura

Department of Electrical Engineering and Information Systems  
Graduate School of Engineering, The University of Tokyo  
tamura@cspp.cc.u-tokyo.ac.jp

Toshihiro Hanawa

Information Technology Center  
The University of Tokyo  
hanawa@cc.u-tokyo.ac.jp

## PP1

### The Schwarz Alternating Method for Concurrent Multiscale Coupling in Solid Mechanics

This poster describes the Schwarz alternating method for concurrent multi-scale coupling in solid mechanics. We describe the approachs variation formulation and convergence properties for quasi-static problems. We describe extensions of the method to dynamics problems, which enable usage of different time integrators in different subdomains. We describe the methods implementation in the open-source multiphysics code, Albany. We provide perspectives towards introducing pervasive multithreading into this implementation. Performance is studied on multi-scale quasi-static and dynamic examples.

Irina K. Tezaur

Sandia National Laboratories  
ikalash@sandia.gov

## PP1

### Parallel Implementation of Hall-effect Thruster Current-driven Instabilities Simulations

Plasma turbulence and current-driven instabilities in Hall-effect thrusters are a promising candidate for anomalous electron cross-field transport theory and pose a challenging multiscale problem. Previous investigations show this instability is associated with high-frequency (of the order of MHz) and short-wavelength (of the order of mm) fluctuations in both the electric field and the plasma density. The behavior of the instability at saturation is also observed to fluctuate in magnitude (on the order of kHz) as well. To investigate these instabilities, we make use of particle-in-cell simulations where positions and velocities of the plasma species are tracked in 3D phase space while the solution of Poisson equation is solved on a cartesian mesh in 2D with a finite difference scheme. Capturing fluctuations of interest on length scales comparable to a Hall-effect thruster chan-

nel becomes computationally expensive due to the requirement of resolving the electron Debye length and plasma frequency. As a result, a parallel implementation is essential for reducing the large computational cost required. The instability behavior and electron cross-field mobility is then compared to numerical and experimental values. This work is supported by the Air Force Office of Scientific Research award FA9550-17RQCOR465. Distribution A: Approved for Public Release; Distribution Unlimited. PA # 17576.

Jonathan Tran

University of California, Los Angeles  
ERC Incorporated  
jonathanktran@ucla.edu

## PP1

### Pseudo-Assembly Programming for Batched Matrix Factorization

While CUDA and OpenCL provide software layers and interfaces for GPU programming, the optimizations from compilers are not able to be controlled by the programmer and sometimes degrade the performance. However, there are lower level intermediate or pseudo-assembly languages (LLVM, PTX, HSAIL) which could give us better management of register/memory allocation and flow control. In this poster, we will show some preliminary result of batched matrix factorizations written in PTX for NVIDIA GPUs, to demonstrate the advantage of low-level programming over CUDA.

Yaohung M. Tsai

Department of Matheamtics  
National Taiwan University  
ev.hunter@gmail.com

Piotr Luszczek

Innovative Computing Laboratory  
University of Tennessee  
luszczek@icl.utk.edu

Jakub Kurzak

University of Tennessee, Knoxville  
kurzak@icl.utk.edu

Jack Dongarra

Innovative Computing Laboratory  
University of Tennessee  
dongarra@cs.utk.edu

## PP1

### Open Distributive Interoperable Executive Library (opendiel) for Multi-Systems Simulation and Analysis

We present an interoperable workflow platform the open Distributive Interoperable Executive Library (openDIEL) - to facilitate the collaboration, exploration, and execution of multi-disciplinary modeling projects suited for a diversified research community on emergent large-scale parallel computing platforms. It does so by providing a managing executive, a workflow configuration input file, and two sets of native communication protocols. OpenDIEL allows users to plug in their individual science codes (modules), prescribe the interactions between those modules, and schedule communications between them. OpenDIEL encapsulates user codes in a single MPI executable and

executes them as arranged in the configuration file.

Kwai L. Wong

Joint Institute for Computational Science  
University of Tennessee/ORNL  
kwong@utk.edu

## PP1

### A Parallel Implementation Technique of HOTRG for the 3D Cubic Lattice Ising Model

Higher order tensor renormalization group (HOTRG) is an algorithm to approximate a partition function in elementary particle physics field with a tensor network. For the three-dimensional cubic lattice Ising model, the computational costs increase  $O(D^{11})$  with the bond dimension  $D$ , which is desired to be 30 to 40. In this talk, we propose a parallel implementation technique of HOTRG for this model and evaluate its performance.

Haruka Yamada

University of Tsukuba  
yamada@mma.cs.tsukuba.ac.jp

Akira Imakura

Department of Computer Science  
University of Tsukuba  
imakura@cs.tsukuba.ac.jp

Toshiyuki Imamura

RIKEN Advance Institute for Computational Science  
imamura.toshiyuki@riken.jp

Tetsuya Sakurai

Department of Computer Science  
University of Tsukuba  
sakurai@cs.tsukuba.ac.jp

## PP1

### Fast Crustal Deformation Computation Method Using Openacc for Stochastic Inversion Analysis

Crustal deformation computation using 3-D high-fidelity models has been in heavy demand due to accumulation of observational data. This approach is computationally expensive and much repetitive computations are required for various application including stochastic inverse analysis. To handle the massive computation cost, we develop a fast finite element analysis method using OpenACC. We use algorithms appropriate for GPUs and accelerate calculations such as Element-by-Element method, which is one kind of sparse matrix-vector multiplications.

Takuma Yamaguchi, Kohei Fujita, Tsuyoshi Ichimura, Muneo Hori

Earthquake Research Institute  
The University of Tokyo  
yamaguchi@eri.u-tokyo.ac.jp, fujita@eri.u-tokyo.ac.jp, ichimura@eri.u-tokyo.ac.jp, hori@eri.u-tokyo.ac.jp

Lalith Wijerathne

University of Tokyo  
Earthquake Research Institute  
lalith@eri.u-tokyo.ac.jp

## PP1

### Consideration on the Optimal Balance Between

### Space Parallelism and Time Parallelism

This research considers the balance between space and time direction parallelism in time evolution simulation. Generally, it is difficult to find a good point at which the parallelism of space and time direction balances and minimize the execution time, because the optimal balance depends on the space size, the time step size, and degree of parallelism computing environment offers. This poster specifies the computing environment as FX10 SuperComputer system and considers the optimal balanced point with various sized problems.

Ryo Yoda, Akihiro Fujii, Teruo Tanaka  
Kogakuin University  
j114126@ns.kogakuin.ac.jp, fujii@cc.kogakuin.ac.jp,  
teru@cc.kogakuin.ac.jp

Kengo Nakajima  
The University of Tokyo  
Information Technology Center  
nakajima@cc.u-tokyo.ac.jp

### PP1

#### A Rational Function Algorithm of Blind Source Separation Based on ICA

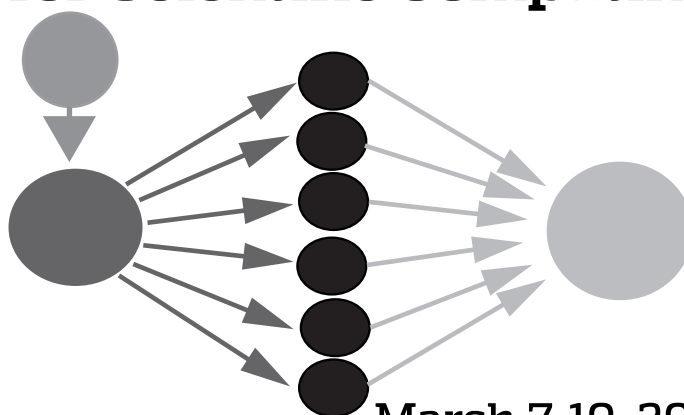
Blind Source Separation (BSS) is a typical and difficult problem in signal processing. The Extended Infomax algorithm of ICA (Independent Component Analysis) can separate the elements from the mixture of super-gaussian and sub-gaussian signals effectively. However, Switch function of the Extended Infomax algorithm needs to choose various nonlinear functions which makes it ineffective. In this paper, we propose the rational function algorithm, which is achieved by applying a rational function directly to a nonlinear function. The rational function algorithm avoids the selection process of switching functions. Numerical results demonstrate that the rational function Algorithm is more effective with a high validity.

Lei Du  
Dalian University of Technology  
China  
dulei@dlut.edu.cn

Jintao Zhang, Jinwei Xu  
School of Mathematics and Physics Science  
Dalian University of Technology at Panjin  
zjt@dlut.edu.cn, xujinwei@mail.dlut.edu.cn

**PP18 Organizer and Speaker Index**

SIAM Conference on  
**Parallel Processing  
for Scientific Computing**



March 7-10, 2018  
Waseda University  
Tokyo, Japan



**A**

Abdelrahman, Tarek S., MS75, 12:00 Fri  
 Abduljabbar, Mustafa, CP12, 9:25 Sat  
*Abhyankar, Shrirang G., MS123, 2:30 Sat*  
*Adams, Mark, MS32, 10:50 Thu*  
*Adams, Mark, MS44, 2:40 Thu*  
 Adams, Mark, MS44, 3:05 Thu  
*Agullo, Emmanuel, MS7, 1:00 Wed*  
*Agullo, Emmanuel, MS18, 3:10 Wed*  
*Agullo, Emmanuel, MS30, 10:50 Thu*  
*Agullo, Emmanuel, MS42, 2:40 Thu*  
 Aihara, Kensuke, MS106, 12:25 Sat  
 Aishima, Kensuke, MS95, 10:40 Sat  
 Akiba, Takuya, IP6, 8:30 Sat  
 Akioka, Sayaka, PP1, 7:15 Thu  
 Aktulga, H. Metin, CP1, 1:00 Wed  
 Al Daas, Hussam, MS28, 11:40 Thu  
 Al Farhan, Mohammed A., CP6, 2:40 Thu  
*Alexeev, Alexander, MS50, 4:50 Thu*  
 Alexeev, Alexander, MS50, 4:50 Thu  
 Alonazi, Amani, CP2, 1:00 Wed  
 Alonazi, Amani, PP1, 7:15 Thu  
 Amano, Hideharu, MS121, 3:45 Sat  
 An, Hengbin, MS106, 12:50 Sat  
 Anandkumar, Anima, MS20, 3:10 Wed  
 Angriman, Eugenio, MS58, 5:40 Thu  
 Antonopoulos, Christos D., MS16, 4:25 Wed  
 Anzt, Hartwig, MS87, 3:20 Fri  
 Ao, Yulong, MS123, 2:55 Sat  
 Aoki, Takayuki, MS4, 1:00 Wed  
*Aoki, Takayuki, MS27, 10:50 Thu*  
*Aoki, Takayuki, MS39, 2:40 Thu*  
 Apra, Edoardo, MS37, 2:40 Thu  
 Arai, Ryosuke, PP1, 7:15 Thu  
 Araki, Samuel, PP1, 7:15 Thu  
*Arora, Ritu, MS85, 2:30 Fri*

Arora, Ritu, MS85, 2:55 Fri  
 Artemov, Anton, CP8, 9:25 Fri  
*Aseeri, Samar A., MS2, 1:00 Wed*  
*Aseeri, Samar A., MS13, 3:10 Wed*  
 Aseeri, Samar A., MS13, 3:10 Wed  
 Aubert, Pierre, MS78, 12:00 Fri  
 Aumage, Olivier, MS35, 3:05 Thu  
 Azad, Ariful, MS79, 12:50 Fri

**B**

*Baboulin, Marc, MS87, 2:30 Fri*  
 Baboulin, Marc, MS87, 2:30 Fri  
*Baboulin, Marc, MS98, 9:25 Sat*  
 Baden, Scott B., MS62, 9:50 Fri  
*Bader, David A., MS58, 4:50 Thu*  
 Bader, David A., MS58, 5:15 Thu  
 Bader, Michael, MS12, 4:00 Wed  
*Bader, Michael, MS68, 9:25 Fri*  
*Bader, Michael, MS80, 11:35 Fri*  
*Bader, Michael, MS91, 2:30 Fri*  
*Badia, Santiago, MS96, 9:25 Sat*  
 Badia, Santiago, MS96, 9:25 Sat  
*Badia, Santiago, MS107, 11:35 Sat*  
*Badia, Santiago, MS118, 2:30 Sat*  
*Bagherpour, Negin, MS14, 3:10 Wed*  
 Bai, Zhaojun, MS78, 12:25 Fri  
 Balaji, Pavan, MS94, 9:50 Sat  
 Balaprakash, Prasanna, MS38, 2:40 Thu  
 Bale, Rahul, MS67, 10:40 Fri  
 Bangerth, Wolfgang, MS113, 11:35 Sat  
 Bar-Gera, Hillel, MS77, 12:00 Fri  
*Bartuschat, Dominik, MS102, 9:25 Sat*  
 Bartuschat, Dominik, MS102, 10:40 Sat  
*Bartuschat, Dominik, MS113, 11:35 Sat*  
 Basu, Protonu, MS75, 12:25 Fri  
 Battaglin, Casey, MS73, 12:25 Fri  
 Beamer, Scott, MS46, 2:40 Thu  
*Bekas, Costas, MS5, 1:00 Wed*  
*Bekas, Costas, MS16, 3:10 Wed*

Belonosov, Mikhail, MS63, 10:15 Fri  
 Ben-Nun, Tal, MS38, 3:55 Thu  
 Bergen, Ben, MS23, 11:15 Thu  
*Bergen, Ben, MS116, 2:30 Sat*  
*Berzins, Martin, MS68, 9:25 Fri*  
*Berzins, Martin, MS80, 11:35 Fri*  
*Berzins, Martin, MS91, 2:30 Fri*  
 Berzins, Martin, MS116, 3:45 Sat  
*Bhatele, Abhinav, MS100, 9:25 Sat*  
*Bhatele, Abhinav, MS111, 11:35 Sat*  
*Bhatele, Abhinav, MS122, 2:30 Sat*  
 Bhatele, Abhinav, MS122, 2:30 Sat  
 Bientinesi, Paolo, MS56, 6:05 Thu  
 Bihari, Barna, CP5, 10:50 Thu  
 Bilardi, Gianfranco, MS73, 12:50 Fri  
 Biros, George, MS33, 11:40 Thu  
*Biswas, Rupak, MS52, 4:50 Thu*  
*Biswas, Rupak, MS64, 9:25 Fri*  
 Biswas, Rupak, MS64, 9:50 Fri  
*Biswas, Rupak, MS76, 11:35 Fri*  
 Blum, Volker, MS25, 10:50 Thu  
 Blumers, Ansel, PP1, 7:15 Thu  
 Boillot, Lionel, PP1, 7:15 Thu  
 Boixo, Sergio, MS64, 10:15 Fri  
 Boku, Taisuke, MS74, 12:00 Fri  
*Boku, Taisuke, MS110, 11:35 Sat*  
*Boku, Taisuke, MS121, 2:30 Sat*  
*Bolten, Matthias, MS8, 1:00 Wed*  
 Bolten, Matthias, MS8, 1:25 Wed  
*Bolten, Matthias, MS19, 3:10 Wed*  
 Boman, Erik G., MS33, 12:05 Thu  
 Bordner, James, MS116, 2:55 Sat  
 Borrell, Ricard, MS68, 10:15 Fri  
*Bosilca, George, MS7, 1:00 Wed*  
*Bosilca, George, MS18, 3:10 Wed*  
*Bosilca, George, MS30, 10:50 Thu*  
*Bosilca, George, MS42, 2:40 Thu*  
 Bosilca, George, MS42, 2:40 Thu

Bozdag, Ebru, MS88, 2:55 Fri  
*Burstedde, Carsten, MS68, 9:25 Fri*  
 Burstedde, Carsten, MS68, 9:25 Fri  
*Burstedde, Carsten, MS80, 11:35 Fri*  
*Burstedde, Carsten, MS91, 2:30 Fri*  
*Burstedde, Carsten, MS117, 2:30 Sat*  
 Buurlage, Jan-Willem, MS61, 10:40 Fri

## C

Calhoun, Donna, MS117, 2:55 Sat  
 Calotoiu, Alexandru, MS41, 3:55 Thu  
 Calvin, Christophe, MS93, 3:20 Fri  
 Canning, Andrew M., CP12, 9:50 Sat  
 Cappello, Frank, MS7, 2:15 Wed  
 Carratalá-Sáez, Rocío, PP1, 7:15 Thu  
 Carribault, Patrick, MS74, 12:25 Fri  
*Carson, Erin C., MS28, 10:50 Thu*  
 Carson, Erin C., MS28, 10:50 Thu  
*Carson, Erin C., MS40, 2:40 Thu*  
*Carter, Jonathan, MS52, 4:50 Thu*  
*Carter, Jonathan, MS64, 9:25 Fri*  
 Carter, Jonathan, MS64, 10:40 Fri  
*Carter, Jonathan, MS76, 11:35 Fri*  
 Casas, Marc, MS5, 1:25 Wed  
 Catalyurek, Umit V., MS34, 12:05 Thu  
 Cecka, Cris, MS3, 1:25 Wed  
 Chac`on, Luis, MS107, 12:00 Sat  
*Chandrasekaran, Sunita, MS1, 1:00 Wed*  
 Chandrasekaran, Sunita, MS1, 1:00 Wed  
 Chang, Justin, PP1, 7:15 Thu  
 Charara, Ali M., MS14, 4:00 Wed  
 Charest, Marc, MS72, 11:35 Fri  
 Chatterjee, Sanjay, MS30, 11:40 Thu  
 Chavez, Gustavo, MS57, 5:40 Thu  
 Che, Shuai, MS108, 12:25 Sat  
*Chen, Chao, MS33, 10:50 Thu*  
*Chen, Chao, MS45, 2:40 Thu*  
*Chen, Chao, MS57, 4:50 Thu*

*Chen, Chao, MS69, 9:25 Fri*  
*Chen, Chao, MS81, 11:35 Fri*  
*Chen, Chao, MS92, 2:30 Fri*  
*Chen, Chao, MS109, 11:35 Sat*  
 Chen, Chao, MS109, 11:35 Sat  
*Chen, Chao, MS120, 2:30 Sat*  
 Chevalier, Cédric, MS93, 2:30 Fri  
 Chiba, Shigeru, MS15, 4:25 Wed  
*Choi, Jee, MS9, 1:00 Wed*  
 Choi, Jee, MS9, 1:00 Wed  
*Choi, Jee, MS20, 3:10 Wed*  
*Choi, Jee, MS56, 4:50 Thu*  
 Claus, Lisa, MS8, 1:50 Wed  
 Conejero, Javier, MS35, 3:30 Thu  
*Cools, Siegfried, MS28, 10:50 Thu*  
*Cools, Siegfried, MS40, 2:40 Thu*  
 Coskun, Ayse, MS10, 1:25 Wed  
 Cottereau, Regis, PP1, 7:15 Thu  
 Coveney, Peter, MS39, 3:30 Thu  
*Cui, Tao, MS32, 10:50 Thu*  
*Cui, Tao, MS44, 2:40 Thu*  
 Cui, Tao, MS120, 2:55 Sat

## D

Dahm, Johann, MS3, 2:15 Wed  
 Dai, Xiaoying, MS25, 11:15 Thu  
 Darve, Eric, MS33, 10:50 Thu  
*Darve, Eric F., MS33, 10:50 Thu*  
*Darve, Eric F., MS45, 2:40 Thu*  
*Darve, Eric F., MS57, 4:50 Thu*  
*Darve, Eric F., MS69, 9:25 Fri*  
*Darve, Eric F., MS81, 11:35 Fri*  
*Darve, Eric F., MS92, 2:30 Fri*  
 Davies, Mike, MS76, 11:35 Fri  
 Dawson, William, MS37, 3:55 Thu  
 D'Azevedo, Eduardo F., MS14, 4:25 Wed  
 De Falco, Carlo, CP1, 1:25 Wed  
 DeBardeleben, Nathan A., MS7, 1:25 Wed

Deelman, Ewa, MS94, 10:15 Sat  
 Dehnavi, Maryam, PP1, 7:15 Thu  
 Deiterding, Ralf, MS91, 2:55 Fri  
 Dekura, Haruhiko, MS99, 10:40 Sat  
*Demeshko, Irina, MS116, 2:30 Sat*  
 Demeshko, Irina, MS116, 2:30 Sat  
 Devarakonda, Aditya, MS61, 10:15 Fri  
*Deveci, Mehmet, MS34, 10:50 Thu*  
*Deveci, Mehmet, MS46, 2:40 Thu*  
 Deveci, Mehmet, MS90, 3:20 Fri  
 Doi, Jun, MS2, 1:50 Wed  
*Domke, Jens, MS71, 11:35 Fri*  
 Domke, Jens, MS71, 12:25 Fri  
 Dongarra, Jack J., MS43, 2:40 Thu  
 Donofrio, David, MS52, 4:50 Thu  
*Drozdz, Aleksandr, MS26, 10:50 Thu*  
*Drozdz, Aleksandr, MS38, 2:40 Thu*  
 Drzisga, Daniel, MS113, 12:00 Sat  
*Du, Lei, MS95, 9:25 Sat*  
*Du, Lei, MS106, 11:35 Sat*  
 Du, Lei, MS106, 11:35 Sat  
 Dubey, Anshu, MS70, 9:50 Fri  
 Dueben, Peter D., MS5, 2:15 Wed  
*Duff, Iain, MS43, 2:40 Thu*  
*Duff, Iain, MS55, 4:50 Thu*  
 Duff, Iain, MS55, 4:50 Thu

## E

Eastep, Jonathan, MS10, 1:50 Wed  
 Eaton, Joe, MS89, 3:45 Fri  
 Eckert, Yasuko, MS52, 5:15 Thu  
*Edwards, H. Carter, MS36, 2:40 Thu*  
 Edwards, H. Carter, MS36, 2:40 Thu  
*Edwards, H. Carter, MS48, 4:50 Thu*  
 Eftekhari, Aryan, PP1, 7:15 Thu  
*Egawa, Ryusuke, MS101, 9:25 Sat*  
 Egawa, Ryusuke, MS101, 9:25 Sat  
 Einkemmer, Lukas, MS24, 11:40 Thu  
*Emad, Nahid, MS78, 11:35 Fri*  
 Emad, Nahid, MS82, 12:00 Fri

*Emad, Nahid, MS89, 2:30 Fri*

Emani, Murali, CP14, 11:35 Sat

Endo, Toshio, MS114, 12:00 Sat

Endo, Wataru, CP8, 9:50 Fri

*Engelmann, Christian, MS7, 1:00 Wed*

*Engelmann, Christian, MS18, 3:10 Wed*

Engelmann, Christian, MS18, 4:00 Wed

*Engelmann, Christian, MS30, 10:50 Thu*

*Engelmann, Christian, MS42, 2:40 Thu*

Engwer, Christian, MS19, 3:35 Wed

*Engwer, Christian, MS105, 11:35 Sat*

Erhel, Jocelyne, MS28, 11:15 Thu

Ertl, Christoph M., MS12, 3:35 Wed

Eswar, Srinivas, MS97, 10:15 Sat

*Evans, Katherine J., MS47, 4:50 Thu*

*Evans, Noah, MS31, 10:50 Thu*

Evans, Noah, MS31, 10:50 Thu

## F

Falgout, Robert D., MS126, 3:20 Sat

Favino, Marco, MS117, 3:45 Sat

Featherstone, Nicholas A., MS88, 3:45 Fri

Feo, John T., MS90, 2:30 Fri

Ferreira da Silva, Rafael, MS10, 2:15 Wed

Flegar, Goran, PP1, 7:15 Thu

Forster, Christopher J., MS56, 4:50 Thu

Franceschini, Andrea, PP1, 7:15 Thu

*Franchetti, Franz, MS2, 1:00 Wed*

Franchetti, Franz, MS2, 1:25 Wed

*Franchetti, Franz, MS13, 3:10 Wed*

Friedhoff, Stephanie, MS104, 9:50 Sat

*Frisch, Jérôme, MS12, 3:10 Wed*

Frisch, Jérôme, MS12, 4:25 Wed

Fu, Haohuan, MS74, 12:50 Fri

Fuerlinger, Karl, MS62, 10:15 Fri

Fujisawa, Katsuki, MS77, 12:25 Fri

*Fujisawa, Katsuki, MS103, 9:25 Sat*

*Fujisawa, Katsuki, MS114, 11:35 Sat*

*Fujisawa, Katsuki, MS125, 2:30 Sat*

Fujita, Kohei, PP1, 7:15 Thu

Fukasawa, Tatsuya, PP1, 7:15 Thu

*Fukaya, Takeshi, MS87, 2:30 Fri*

*Fukaya, Takeshi, MS98, 9:25 Sat*

Fukaya, Takeshi, MS98, 9:25 Sat

Fukuda, Keisuke, MS38, 3:05 Thu

Futamura, Yasunori, MS18, 3:35 Wed

## G

Gamblin, Todd, MS1, 1:50 Wed

Gamell, Marc, MS42, 3:05 Thu

Gao, Xiang, CP6, 3:05 Thu

Gardner, David J., CP14, 12:25 Sat

Gassmoeller, Rene, MS88, 3:20 Fri

Gebremedhin, Assefaw, MS90, 2:55 Fri

Gerofi, Balazs, MS31, 12:05 Thu

*Ghattas, Omar, MS102, 9:25 Sat*

*Ghattas, Omar, MS113, 11:35 Sat*

Ghysels, Pieter, MS40, 3:05 Thu

*Gibou, Frederic, MS117, 2:30 Sat*

Gibou, Frederic, MS117, 2:30 Sat

Gibson, Thomas H., MS105, 12:25 Sat

*Giraud, Luc, MS7, 1:00 Wed*

*Giraud, Luc, MS18, 3:10 Wed*

*Giraud, Luc, MS30, 10:50 Thu*

*Giraud, Luc, MS42, 2:40 Thu*

*Godenschwager, Christian, MS27, 10:50 Thu*

Godenschwager, Christian, MS27, 10:50 Thu

*Godenschwager, Christian, MS39, 2:40 Thu*

Goh, Eng-Lim, MS52, 5:40 Thu

Gomi, Ayumu, CP3, 3:10 Wed

Gong, Shihua, MS118, 3:20 Sat

*Gracia, Jose, MS101, 9:25 Sat*

Gracia, Jose, MS101, 10:40 Sat

*Graillat, Stef, MS54, 4:50 Thu*

*Graillat, Stef, MS66, 9:25 Fri*

Graillat, Stef, MS66, 9:25 Fri

Grasedyck, Lars, MS6, 1:00 Wed

*Green, Oded, MS58, 4:50 Thu*

Grigori, Laura, MS55, 5:40 Thu

Gropp, William D., MS40, 3:30 Thu

Gryazin, Yury A., CP10, 2:30 Fri

Guo, Haoqiang, MS97, 10:40 Sat

*Guo, Xiaohu, MS32, 10:50 Thu*

Guo, Xiaohu, MS32, 11:40 Thu

*Guo, Xiaohu, MS44, 2:40 Thu*

Guo, Zhipeng, MS112, 12:00 Sat

## H

*Hager, Georg, MS29, 10:50 Thu*

Hager, Georg, MS29, 10:50 Thu

*Hager, Georg, MS41, 2:40 Thu*

Hahne, Jan, CP9, 11:35 Fri

*Haidar, Azzam, MS3, 1:00 Wed*

*Haidar, Azzam, MS14, 3:10 Wed*

Haidar, Azzam, MS14, 3:10 Wed

Hamada, Shoji, MS81, 12:25 Fri

Hammarling, Sven J., MS43, 3:05 Thu

*Hanawa, Toshihiro, MS110, 11:35 Sat*

Hanawa, Toshihiro, MS110, 12:25 Sat

*Hanawa, Toshihiro, MS121, 2:30 Sat*

Hancock, David, MS1, 2:15 Wed

Hasegawa, Yuta, MS27, 11:15 Thu

Haus, Utz-Uwe, MS114, 11:35 Sat

Hayashi, Koby B., MS20, 4:25 Wed

He, Xin, MS108, 12:00 Sat

Heavner, Nathan, CP10, 3:55 Fri

Helluy, Philippe, MS15, 3:35 Wed

Helm, Christian R., CP3, 3:35 Wed

Herbordt, Martin, MS110, 12:50 Sat

*Heroux, Michael, MS47, 4:50 Thu*

*Heroux, Michael, MS70, 9:25 Fri*

Heroux, Michael, MS70, 9:25 Fri

*Heroux, Michael, MS82, 11:35 Fri*

*Heroux, Michael, MS93, 2:30 Fri*

Higham, Nicholas J., MS54, 5:15 Thu

Hiraishi, Tasuku, MS57, 4:50 Thu

*Hirota, Yusuke, MS49, 4:50 Thu*

Hishinuma, Toshiaki, MS49, 6:05 Thu

Hoefler, Torsten, MS46, 3:05 Thu

Hoekstra, Rob, MS72, 12:50 Fri

Hoemmen, Mark, MS82, 12:25 Fri

Hoffer, Michael, MS17, 4:25 Wed

Holke, Johannes, MS68, 9:50 Fri

Holladay, Daniel A., MS48, 6:05 Thu

*Hollman, David S., MS23, 10:50 Thu*

Hollman, David S., MS23, 10:50 Thu

*Hollman, David S., MS35, 2:40 Thu*

*Hori, Takane, MS65, 9:25 Fri*

Hori, Takane, MS65, 9:25 Fri

*Hori, Takane, MS77, 11:35 Fri*

Hoshi, Takeo, MS37, 3:05 Thu

*Hoshi, Takeo, MS54, 4:50 Thu*

Hoshi, Takeo, PP1, 7:15 Thu

*Hoshi, Takeo, MS66, 9:25 Fri*

Hoshino, Tetsuya, MS45, 3:05 Thu

Hotta, Hideyuki, MS99, 10:15 Sat

Hötzer, Johannes, MS112, 12:25 Sat

*Hu, Xiangyu, MS32, 10:50 Thu*

Hu, Xiangyu, MS32, 10:50 Thu

*Hu, Xiangyu, MS44, 2:40 Thu*

Hu, Yao, MS71, 12:50 Fri

Huang, Jizu, MS83, 3:20 Fri

Huang, Su-Yun, MS86, 2:55 Fri

Huber, Sarah, PP1, 7:15 Thu

*Huckle, Thomas K., MS8, 1:00 Wed*

*Huckle, Thomas K., MS19, 3:10 Wed*

Huckle, Thomas K., MS19, 4:00 Wed

Huebl, Axel, MS1, 1:25 Wed

Hughes, Clayton, MS111, 12:50 Sat

Hukerikar, Saurabh, MS7, 1:00 Wed

Hutter, Edward, MS73, 12:00 Fri

*Hwang, Feng-Nan, MS83, 2:30 Fri*

Hwang, Feng-Nan, MS83, 2:30 Fri

## I

Iakymchuk, Roman, MS66, 9:50 Fri

Iancu, Costin, MS53, 5:15 Thu

Ibanez, Daniel, MS48, 5:15 Thu

Ichimura, Tsuyoshi, IP5, 4:40 Fri

*Ida, Akihiro, MS33, 10:50 Thu*

*Ida, Akihiro, MS45, 2:40 Thu*

*Ida, Akihiro, MS57, 4:50 Thu*

*Ida, Akihiro, MS69, 9:25 Fri*

*Ida, Akihiro, MS81, 11:35 Fri*

*Ida, Akihiro, MS92, 2:30 Fri*

Ida, Akihiro, MS120, 3:20 Sat

Idomura, Yasuhiro, CP5, 11:15 Thu

*Iizuka, Mikio, MS104, 9:25 Sat*

*Iizuka, Mikio, MS115, 11:35 Sat*

Iizuka, Mikio, MS115, 11:35 Sat

*Iizuka, Mikio, MS126, 2:30 Sat*

Ikuno, Soichiro, MS22, 3:10 Wed

*Imakura, Akira, MS95, 9:25 Sat*

Imakura, Akira, MS95, 9:25 Sat

*Imakura, Akira, MS106, 11:35 Sat*

*Imamura, Toshiyuki, MS75, 11:35 Fri*

*Imamura, Toshiyuki, MS86, 2:30 Fri*

Imamura, Toshiyuki, MS86, 2:30 Fri

Inagaki, Atsushi, MS12, 3:10 Wed

Inagawa, Yuta, PP1, 7:15 Thu

Incardona, Pietro, MS32, 12:05 Thu

*Iryo, Takamasa, MS77, 11:35 Fri*

Iryo, Takamasa, MS77, 11:35 Fri

Isaac, Tobin, MS68, 10:40 Fri

Isaev, Mikhail, MS38, 3:30 Thu

Ishida, Takashi, MS39, 2:40 Thu

Iwabuchi, Keita, PP1, 7:15 Thu

Iwase, Shigeru, MS78, 12:50 Fri

*Iwashita, Takeshi, MS11, 1:00 Wed*

*Iwashita, Takeshi, MS22, 3:10 Wed*

Iwashita, Takeshi, MS22, 4:00 Wed

*Iwashita, Takeshi, MS87, 2:30 Fri*

*Iwashita, Takeshi, MS98, 9:25 Sat*

*Iyo, Takamasa, MS65, 9:25 Fri*

## J

*Jacquelin, Mathias, MS84, 2:30 Fri*

Jacquelin, Mathias, MS84, 3:20 Fri

Jaeger, Julien, MS72, 12:00 Fri

Jain, Nikhil, MS100, 9:50 Sat

Janna, Carlo, CP10, 3:20 Fri

*Jansson, Niclas, MS67, 9:25 Fri*

Jansson, Niclas, MS67, 9:25 Fri

Javier Conejero, Francisco, MS21, 4:00 Wed

Jia, Weile, MS37, 3:30 Thu

Jiang, Wenming, PP1, 7:15 Thu

Jing, Yanfei, MS106, 12:00 Sat

Johansen, Hans, MS117, 3:20 Sat

*Jolivet, Pierre, MS60, 9:25 Fri*

Jolivet, Pierre, MS60, 10:40 Fri

Joo, Balint, MS36, 3:30 Thu

## K

Kågström, Bo T., MS43, 3:30 Thu

Kaiser, Hartmut, MS116, 3:20 Sat

Kalyanaraman, Ananth, MS79, 12:25 Fri

Kameyama, Masanori, MS102, 9:25 Sat

Kardos, Juraj, MS103, 10:15 Sat

*Karlin, Ian, MS72, 11:35 Fri*

*Katagiri, Takahiro, MS75, 11:35 Fri*

Katagiri, Takahiro, MS75, 11:35 Fri

*Katagiri, Takahiro, MS86, 2:30 Fri*

*Kawai, Hiroshi, MS96, 9:25 Sat*

Kawai, Hiroshi, MS96, 10:40 Sat

*Kawai, Hiroshi, MS107, 11:35 Sat*

*Kawai, Hiroshi, MS118, 2:30 Sat*

Kawai, Masatoshi, MS29, 11:40 Thu

Kaya, Oguz, MS9, 1:25 Wed

Kellogg, Louise, MS88, 2:30 Fri

Keyes, David E., MS120, 2:30 Sat

*Kim, Kibaek, MS103, 9:25 Sat*

*Kim, Kibaek, MS114, 11:35 Sat*

*Kim, Kibaek, MS125, 2:30 Sat*

Kim, Kibaek, MS125, 2:55 Sat

*Italicized names indicate session organizers*



Kim, Kyungjoo, MS109, 12:50 Sat  
 Kimura, Kinji, PP1, 7:15 Thu  
 Kjelgaard Mikkelsen, Carl Christian, MS43, 3:55 Thu  
 Kjolstad, Fredrik, MS9, 1:50 Wed  
 Klawonn, Axel, MS96, 9:50 Sat  
*Klingenberg, Christian F., MS4, 1:00 Wed*  
 Klingenberg, Christian F., MS4, 1:50 Wed  
*Klingenberg, Christian F., MS15, 3:10 Wed*  
 Knepley, Matthew G., MS102, 9:50 Sat  
 Knepper, Sarah, MS3, 1:00 Wed  
 Knight, Nicholas, MS61, 9:50 Fri  
 Knyazev, Andrew, CP11, 2:30 Fri  
 Kobayashi, Ryohei, MS110, 12:00 Sat  
 Kobayashi, Yuka, PP1, 7:15 Thu  
 Koch, Marcel, MS105, 12:50 Sat  
 Koch, Thorsten, MS103, 9:25 Sat  
*Koestler, Harald, MS4, 1:00 Wed*  
*Koestler, Harald, MS15, 3:10 Wed*  
 Koga, Issei, PP1, 7:15 Thu  
 Kolla, Hemanth, CP14, 12:00 Sat  
 Komatsu, Kazuhiko, MS4, 2:15 Wed  
*Kondo, Masaaki, MS10, 1:00 Wed*  
*Kondo, Masaaki, MS21, 3:10 Wed*  
 Konduri, Aditya, CP13, 9:25 Sat  
 Kong, Martin, MS119, 2:55 Sat  
 Konolige, Tristan, MS123, 3:20 Sat  
 Kopanicakova, Alena, MS19, 3:10 Wed  
*Koumoutsakos, Petros, MS32, 10:50 Thu*  
*Koumoutsakos, Petros, MS44, 2:40 Thu*  
 Koumoutsakos, Petros, MS44, 3:30 Thu  
 Kourounis, Drosos, MS103, 10:40 Sat  
 Kouya, Tomonori, CP8, 10:15 Fri  
 Krause, Mathias J., MS39, 3:05 Thu  
 Krause, Rolf, MS6, 1:25 Wed  
 Kriemann, Ronald, MS45, 3:55 Thu  
 Kronbichler, Martin, MS29, 11:15 Thu  
 Kuckuk, Sebastian, MS15, 4:00 Wed  
 Kulchytska-Ruchka, Iryna, MS126, 2:55 Sat

**L**  
 Labarta, Jesus, MS62, 10:40 Fri  
 Laguna, Ignacio, MS7, 1:50 Wed  
 Lahnert, Michael, MS27, 12:05 Thu  
 Lan, Yu-Hsiang, MS123, 3:45 Sat  
 Lanser, Martin, CP7, 4:50 Thu  
 Lee, Dongjin, MS98, 10:40 Sat  
*Lee, Jinpil, MS119, 2:30 Sat*  
 Lee, Jinpil, MS119, 2:30 Sat  
 Lee, Seyong, MS119, 3:20 Sat  
 Lelbach, Bryce, MS35, 2:40 Thu  
 Levchenko, Vadim, PP1, 7:15 Thu  
 Lewis, Katie, MS76, 12:00 Fri  
 Li, Jiajia, MS9, 2:15 Wed  
 Li, Jichun, CP15, 2:30 Sat  
 Li, Senxi, PP1, 7:15 Thu  
 Li, Shengguo, MS95, 10:15 Sat  
 Li, X. Sherry, MS93, 2:55 Fri  
 Li, Yingzhou, CP4, 3:10 Wed  
*Li, Yingzhou, MS109, 11:35 Sat*  
*Li, Yingzhou, MS120, 2:30 Sat*  
 Liang, Yun (Eric), MS97, 9:50 Sat  
 Liao, Li, MS59, 10:40 Fri  
*Lifflander, Jonathan J., MS23, 10:50 Thu*  
*Lifflander, Jonathan J., MS35, 2:40 Thu*  
 Lifflander, Jonathan J., MS35, 3:55 Thu  
 Lin, Matthew M., MS92, 2:55 Fri  
*Lin, Paul, MS96, 9:25 Sat*  
*Lin, Paul, MS107, 11:35 Sat*  
*Lin, Paul, MS118, 2:30 Sat*  
 Lin, Paul, MS118, 3:45 Sat  
 Lindner, Florian, CP2, 1:25 Wed  
*Lisitsa, Vadim, MS51, 4:50 Thu*  
 Lisitsa, Vadim, MS51, 4:50 Thu  
*Lisitsa, Vadim, MS63, 9:25 Fri*  
 Litvinenko, Alexander, CP12, 10:15 Sat  
*Liu, Weifeng, MS97, 9:25 Sat*  
*Liu, Weifeng, MS108, 11:35 Sat*  
 Liu, Weifeng, MS108, 12:50 Sat

Liu, Yang, MS92, 2:30 Fri  
 Loffeld, John, CP14, 12:50 Sat  
 Logashenko, Dmitry, PP1, 7:15 Thu  
 Lopez, Florent, MS55, 5:15 Thu  
 Losada, Nuria, MS42, 3:30 Thu  
 Low, Tze Meng, MS75, 12:50 Fri  
 Ltaief, Hatem, MS45, 3:30 Thu  
 Lucas, Bob, MS52, 6:05 Thu  
 Luszczek, Piotr, MS40, 3:55 Thu

**M**  
 Makino, Jun, MS124, 3:20 Sat  
*Malossi, Cristiano, MS5, 1:00 Wed*  
 Malossi, Cristiano, MS5, 1:00 Wed  
*Malossi, Cristiano, MS16, 3:10 Wed*  
 Manin, Valeriy, PP1, 7:15 Thu  
 Manne, Fredrik, MS79, 12:00 Fri  
 Marin Elena, Alin, MS44, 3:55 Thu  
 Marongiu, Andrea, MS5, 1:50 Wed  
*Marques, Osni A., MS75, 11:35 Fri*  
*Marques, Osni A., MS86, 2:30 Fri*  
 Marques, Osni A., MS86, 3:45 Fri  
 Martinsson, Gunnar, MS33, 11:15 Thu  
 Martorell, Xavier, MS121, 2:55 Sat  
*Maruyama, Naoya, MS26, 10:50 Thu*  
*Maruyama, Naoya, MS38, 2:40 Thu*  
 Mary, Theo, MS57, 6:05 Thu  
 Matsuda, Momo, CP9, 12:00 Fri  
 Matsui, Hiroaki, PP1, 7:15 Thu  
*Matsui, Hiroaki, MS88, 2:30 Fri*  
*Matsui, Hiroaki, MS99, 9:25 Sat*  
*Matsuo, Takayasu, MS24, 10:50 Thu*  
 Matsuo, Takayasu, MS24, 10:50 Thu  
 Mayumi, Akie, MS87, 2:55 Fri  
 McCabe, Arthur, MS31, 11:15 Thu  
 McColl, Bill, MS61, 9:25 Fri  
 McInnes, Lois Curfman, IP3, 1:45 Thu  
 McIntosh-Smith, Simon, MS45, 2:40 Thu  
*Mehl, Miriam, MS4, 1:00 Wed*  
*Mehl, Miriam, MS15, 3:10 Wed*  
 Mehri Dehnavi, Maryam M., MS56,

5:40 Thu

Meneses, Esteban, MS30, 10:50 Thu

Michelogiannakis, George, MS111,  
12:00 Sat

*Mifune, Takeshi, MS11, 1:00 Wed*

*Mifune, Takeshi, MS22, 3:10 Wed*

Mifune, Takeshi, MS22, 4:25 Wed

Miller, Phil, MS23, 11:40 Thu

Minami, Kazuo, MS101, 10:15 Sat

Miyagi, Mituhiro, MS115, 12:50 Sat

*Miyatake, Yuto, MS24, 10:50 Thu*

Moltó, Germán, CP13, 9:50 Sat

Moore, Shirley, MS111, 11:35 Sat

Moresi, Louis, MS113, 12:25 Sat

Morgan, Hannah, MS40, 2:40 Thu

Mubarak, Misbah, MS122, 2:55 Sat

*Muite, Benson K., MS2, 1:00 Wed*

*Muite, Benson K., MS13, 3:10 Wed*

Muite, Benson K., MS13, 3:35 Wed

*Mukunoki, Daichi, MS49, 4:50 Thu*

Mukunoki, Daichi, MS49, 4:50 Thu

*Mundani, Ralf-Peter, MS12, 3:10 Wed*

*Murai, Hitoshi, MS62, 9:25 Fri*

*Murai, Hitoshi, MS74, 11:35 Fri*

Murakami, Hiroshi, CP4, 3:35 Wed

*Müthing, Steffen, MS105, 11:35 Sat*

Müthing, Steffen, MS105, 11:35 Sat

## N

Nagasaka, Yusuke, MS93, 3:45 Fri

Nägel, Arne, MS17, 4:00 Wed

Nakagawa, Takashi, PP1, 7:15 Thu

*Nakagawa, Takashi, MS88, 2:30 Fri*

*Nakagawa, Takashi, MS99, 9:25 Sat*

*Nakajima, Kengo, MS70, 9:25 Fri*

*Nakajima, Kengo, MS82, 11:35 Fri*

Nakajima, Kengo, MS82, 11:35 Fri

*Nakajima, Kengo, MS93, 2:30 Fri*

Nakamura, Masanori, MS50, 5:40 Thu

Nakata, Ayako, MS25, 12:05 Thu

Naumann, Uwe, MS16, 3:35 Wed

*Neumann, Philipp, MS124, 2:30 Sat*

Neumann, Philipp, MS124, 2:30 Sat

*Ng, Esmond G., MS84, 2:30 Fri*

Ng, Esmond G., MS84, 2:30 Fri

Nicholaeff, David, CP2, 1:50 Wed

Nielsen, Allan, MS104, 10:40 Sat

*Nikolopoulos, Dimitrios S., MS5, 1:00 Wed*

*Nikolopoulos, Dimitrios S., MS16, 3:10 Wed*

Nikolopoulos, Dimitrios S., MS16, 3:10  
Wed

Nishimori, Hidetoshi, MS64, 9:25 Fri

Nishimura, Naoshi, MS81, 11:35 Fri

Noguchi, Hiroshi, MS50, 5:15 Thu

Nomura, Akihiro, MS71, 11:35 Fri

Nomura, Naoya, MS8, 2:15 Wed

Norman, Matthew R., MS47, 6:05 Thu

Nukada, Akira, MS2, 2:15 Wed

## O

Obersteiner, Michael, MS18, 4:25 Wed

Odajima, Tetsuya, MS49, 5:40 Thu

*Ogino, Masao, MS96, 9:25 Sat*

*Ogino, Masao, MS107, 11:35 Sat*

Ogino, Masao, MS107, 12:25 Sat

*Ogino, Masao, MS118, 2:30 Sat*

*Ogita, Takeshi, MS54, 4:50 Thu*

Ogita, Takeshi, MS54, 4:50 Thu

*Ogita, Takeshi, MS66, 9:25 Fri*

Ohi, Yoshiharu, MS104, 10:15 Sat

Ohshima, Satoshi, PP1, 7:15 Thu

Oishi, Shin'ichi, IP1, 5:15 Wed

Oliker, Leonid, MS29, 12:05 Thu

*Oliker, Leonid, MS52, 4:50 Thu*

*Oliker, Leonid, MS64, 9:25 Fri*

*Oliker, Leonid, MS76, 11:35 Fri*

Olson, Luke, MS8, 1:00 Wed

Onishi, Keiji, MS67, 9:50 Fri

*Ono, Kenji, MS104, 9:25 Sat*

*Ono, Kenji, MS115, 11:35 Sat*

*Ono, Kenji, MS126, 2:30 Sat*

Ono, Takatsugu, MS122, 3:20 Sat

Onodera, Naoyuki, MS27, 11:40 Thu

Osei-Kuffuor, Daniel, MS25, 11:40 Thu

Overfelt, James, MS36, 3:55 Thu

Ozaki, Katsuhisa, MS66, 10:15 Fri

## P

Paludetto Magri, Victor A., CP5, 11:40  
Thu

Parashar, Manish, MS94, 10:40 Sat

Parret-Fréaud, Augustin, CP7, 5:15 Thu

*Patki, Tapasya, MS10, 1:00 Wed*

Patki, Tapasya, MS10, 1:00 Wed

*Patki, Tapasya, MS21, 3:10 Wed*

Pawlowski, Roger, MS70, 10:15 Fri

Payne, Joshua, MS72, 12:25 Fri

Pearce, Roger, MS34, 11:15 Thu

Pedretti, Kevin, MS31, 11:40 Thu

Peng, Liqian, MS24, 11:15 Thu

Perego, Mauro, MS48, 4:50 Thu

Perepelkina, Anastasia, MS63, 9:25 Fri

Petersson, Anders, MS99, 9:25 Sat

*Petiton, Serge G., MS70, 9:25 Fri*

Petiton, Serge, MS74, 11:35 Fri

*Petiton, Serge G., MS82, 11:35 Fri*

*Petiton, Serge G., MS93, 2:30 Fri*

*Petra, Cosmin G., MS103, 9:25 Sat*

*Petra, Cosmin G., MS114, 11:35 Sat*

*Petra, Cosmin G., MS125, 2:30 Sat*

Petra, Cosmin G., MS125, 2:30 Sat

Petrogalli, Francesco, MS119, 3:45 Sat

Pfreundt, Franz-Josef, MS42, 3:55 Thu

Phipps, Eric, MS56, 5:15 Thu

*Pivkin, Igor V., MS50, 4:50 Thu*

Pivkin, Igor V., MS50, 6:05 Thu

*Podobas, Artur M., MS110, 11:35 Sat*

*Podobas, Artur M., MS121, 2:30 Sat*

*Pothen, Alex, MS79, 11:35 Fri*

Pothen, Alex, MS79, 11:35 Fri

*Pothen, Alex, MS90, 2:30 Fri*

*Prokopenko, Andrey, MS47, 4:50 Thu*

Prokopenko, Andrey, MS47, 4:50 Thu

Pueschel, Markus, MS41, 3:05 Thu

**Q**

Quintana-Ortí, Enrique S., MS5, 1:00 Wed  
 Quintana-Ortí, Enrique S., MS16, 3:10 Wed  
 Quintana-Ortí, Enrique S., MS16, 4:00 Wed

**R**

Rajamanickam, Siva, MS3, 1:00 Wed  
 Rajamanickam, Siva, MS3, 1:50 Wed  
 Rajamanickam, Siva, MS14, 3:10 Wed  
 Ralphs, Ted, MS125, 3:20 Sat  
 Ramanan, Paritosh P., MS109, 12:00 Sat  
 Ramanathan, Arvind, MS26, 11:40 Thu  
 Ramirez, Elizabeth C., CP9, 12:25 Fri  
 Rehfeldt, Daniel, MS114, 12:50 Sat  
 Reiter, Sebastian, MS17, 3:35 Wed  
 Remacle, Jean-François, CP6, 3:30 Thu  
 Reshetova, Galina, MS51, 4:50 Thu  
 Reshetova, Galina, MS63, 9:25 Fri  
 Reshetova, Galina, MS63, 9:50 Fri  
 Rettinger, Christoph, MS39, 3:55 Thu  
 Riedy, Jason, MS58, 4:50 Thu  
 Rippl, Michael, MS19, 4:25 Wed  
 Rogers, Tim, MS111, 12:25 Sat  
 Roman, Jose E., MS89, 2:55 Fri  
 Ronaghan, Elliot, MS53, 5:40 Thu  
 Ross, Rob, MS100, 9:25 Sat  
 Ross, Rob, MS111, 11:35 Sat  
 Ross, Rob, MS122, 2:30 Sat  
 Rouet, Francois-Henry, MS84, 2:55 Fri  
 Roux, François-Xavier, MS60, 10:15 Fri  
 Rüde, Ulrich, MS27, 10:50 Thu  
 Rüde, Ulrich, MS39, 2:40 Thu  
 Rudi, Johann, MS102, 10:15 Sat  
 Ruede, Ulrich J., MS59, 9:50 Fri  
 Ruede, Ulrich J., MS102, 9:25 Sat  
 Ruede, Ulrich J., MS113, 11:35 Sat  
 Ruede, Ulrich J., MS59, 9:25 Fri  
 Rump, Siegfried M., MS54, 5:40 Thu  
 Rupp, Karl, MS105, 12:00 Sat  
 Ruprecht, Daniel, PP1, 7:15 Thu  
 Ruprecht, Daniel, MS104, 9:25 Sat

Ruprecht, Daniel, MS115, 11:35 Sat

Ruprecht, Daniel, MS115, 12:25 Sat

Ruprecht, Daniel, MS126, 2:30 Sat

**S**

Saad, Tony, MS59, 10:15 Fri  
 Saad, Yousef, MS84, 3:45 Fri  
 Saka, Toshihide, CP12, 10:40 Sat  
 Sakurai, Tetsuya, MS78, 11:35 Fri  
 Sakurai, Tetsuya, MS89, 2:30 Fri  
 Samfass, Philipp, MS80, 11:35 Fri  
 Sanan, Patrick, MS113, 12:50 Sat  
 Sano, Kentaro, MS110, 11:35 Sat  
 Sano, Kentaro, MS110, 11:35 Sat  
 Sano, Kentaro, MS121, 2:30 Sat  
 Sarkar, Abhijit, CP7, 5:40 Thu  
 Sato, Daisuke, MS69, 10:40 Fri  
 Sato, Mitsuhisa, MS62, 9:25 Fri  
 Sattari, Sulimon, CP9, 12:50 Fri  
 Sbalzarini, Ivo F., MS32, 10:50 Thu  
 Schaller, Matthieu, MS32, 11:15 Thu  
 Schanen, Michel, MS114, 12:25 Sat  
 Schenk, Olaf, MS103, 9:25 Sat  
 Schenk, Olaf, MS114, 11:35 Sat  
 Schenk, Olaf, MS125, 2:30 Sat  
 Schlatter, Philipp, MS67, 10:15 Fri  
 Schöbel, Ruth, MS104, 9:25 Sat  
 Schornbaum, Florian, MS91, 2:30 Fri  
 Schreiber, Robert S., MS58, 4:50 Thu  
 Schulz, Martin, MS10, 1:00 Wed  
 Schulz, Martin, MS21, 3:10 Wed  
 Schulz, Volker H., MS6, 1:50 Wed  
 Schuman, Catherine, MS76, 12:25 Fri  
 Schwartz, Oded, MS61, 9:25 Fri  
 Schwartz, Oded, MS73, 11:35 Fri  
 Schwartz, Oded, MS73, 11:35 Fri  
 Schwarz, Angelika, PP1, 7:15 Thu  
 Seckler, Steffen, MS124, 3:45 Sat  
 Semba, Kazuki, MS11, 1:50 Wed  
 Seyed Agha Banihashemi, Saeed, PP1, 7:15 Thu  
 Shalf, John, IP2, 8:30 Thu

Shamis, Pavel, MS53, 6:05 Thu

Shende, Sameer, MS85, 3:45 Fri

Shimokawabe, Takashi, MS91, 3:45 Fri

Shin, Yoonsu, CP13, 10:15 Sat

Shinano, Yuji, MS103, 9:25 Sat

Shinano, Yuji, MS114, 11:35 Sat

Shinano, Yuji, MS125, 2:30 Sat

Shioya, Akemi, CP8, 10:40 Fri

Shioya, Ryuji, MS96, 9:25 Sat

Shioya, Ryuji, MS107, 11:35 Sat

Shioya, Ryuji, MS118, 2:30 Sat

Shioya, Ryuji, MS118, 2:30 Sat

Shipton, Jemma, MS115, 12:00 Sat

Shiu, Wen-Shin, MS83, 3:45 Fri

Shu, Raphael, MS26, 12:05 Thu

Sid-Lakhdar, Wissam M., CP5, 12:05 Thu

Siefert, Christopher, CP11, 2:55 Fri

Simpson, Toby, MS90, 3:45 Fri

Sistek, Jakub, CP4, 4:00 Wed

Slaughter, Elliott, MS23, 12:05 Thu

Slota, George M., MS34, 10:50 Thu

Slota, George M., MS34, 10:50 Thu

Slota, George M., MS46, 2:40 Thu

Smith, Cameron, MS46, 3:55 Thu

Smith, Shaden, MS20, 3:35 Wed

Solovyev, Sergey, MS51, 5:15 Thu

Song, Shuaiwen, MS97, 9:25 Sat

Sosonkina, Masha, MS98, 9:50 Sat

Sousa, Eder M., PP1, 7:15 Thu

Speck, Robert, PP1, 7:15 Thu

Speck, Robert, MS104, 9:25 Sat

Speck, Robert, MS115, 11:35 Sat

Speck, Robert, MS126, 2:30 Sat

Spillane, Nicole, MS98, 10:15 Sat

Stadler, Georg, MS65, 10:40 Fri

Stals, Linda, MS18, 3:10 Wed

Strohmaier, Erich, MS101, 9:50 Sat

Subasi, Omer, MS30, 11:15 Thu

Subedi, Pradeep, MS21, 4:25 Wed

Sugimoto, Shin-ichiro, MS11, 1:25 Wed



Susukita, Ryutaro, MS100, 10:40 Sat

Suter, Frédéric, MS100, 10:15 Sat

*Suzuki, Atsushi, MS60, 9:25 Fri*

Suzuki, Atsushi, MS60, 9:25 Fri

## T

Tackley, Paul J., MS99, 9:50 Sat

Tadano, Hiroto, PP1, 7:15 Thu

Tagami, Daisuke, MS118, 2:55 Sat

*Takahashi, Daisuke, MS2, 1:00 Wed*

Takahashi, Daisuke, MS2, 1:00 Wed

*Takahashi, Daisuke, MS13, 3:10 Wed*

Takahashi, Toru, MS92, 3:20 Fri

*Takahashi, Yasuhito, MS11, 1:00 Wed*

Takahashi, Yasuhito, MS11, 1:00 Wed

*Takahashi, Yasuhito, MS22, 3:10 Wed*

*Takaki, Tomohiro, MS112, 11:35 Sat*

Takaki, Tomohiro, MS112, 11:35 Sat

Takami, Toshiya, MS126, 2:30 Sat

Takayanagi, Masatoshi, PP1, 7:15 Thu

Takayasu, Akitoshi, PP1, 7:15 Thu

Takei, Amane, MS66, 10:40 Fri

*Takizawa, Hiroyuki, MS85, 2:30 Fri*

Takizawa, Hiroyuki, MS85, 2:30 Fri

Tamura, Kohei, PP1, 7:15 Thu

Tan, Guangming, MS58, 6:05 Thu

*Tan, Guangming, MS97, 9:25 Sat*

*Tan, Guangming, MS108, 11:35 Sat*

Tan, Yiyu, MS20, 4:00 Wed

Tang, William, MS26, 11:15 Thu

Tang, Yu-Hang, CP11, 3:20 Fri

Tao, Cui, MS44, 2:40 Thu

*Taufer, Michela, MS94, 9:25 Sat*

Taufer, Michela, MS94, 9:25 Sat

*Taura, Kenjiro, MS53, 4:50 Thu*

Taura, Kenjiro, MS53, 4:50 Thu

Tcheverda, Vladimir A., MS51, 5:40 Thu

*Teranishi, Keita, MS9, 1:00 Wed*

*Teranishi, Keita, MS7, 1:00 Wed*

*Teranishi, Keita, MS20, 3:10 Wed*

*Teranishi, Keita, MS18, 3:10 Wed*

*Teranishi, Keita, MS30, 10:50 Thu*

*Teranishi, Keita, MS42, 2:40 Thu*

*Teranishi, Keita, MS56, 4:50 Thu*

Terao, Takeshi, MS86, 3:20 Fri

Tezaur, Irina K., PP1, 7:15 Thu

Thoennes, Dominik, MS107, 11:35 Sat

Tissot, Olivier, MS55, 6:05 Thu

*Toledo, Sivan A., MS61, 9:25 Fri*

*Toledo, Sivan A., MS73, 11:35 Fri*

Tom, Ancy, MS46, 3:30 Thu

Toth, Alexander, CP1, 1:50 Wed

Tran, Jonathan, PP1, 7:15 Thu

Trinitis, Carsten, MS21, 3:35 Wed

Tromp, Jeroen, MS65, 9:50 Fri

Trott, Christian, MS36, 3:05 Thu

Truong, Nguyen, MS71, 12:00 Fri

Trzil, Zachary, CP13, 10:40 Sat

Tsai, Yaohung M., PP1, 7:15 Thu

*Tsubokura, Makoto, MS67, 9:25 Fri*

Tsuburaya, Tomonori, MS11, 2:15 Wed

Tsuji, Miwako, MS30, 12:05 Thu

Tuminaro, Raymond, MS107, 12:50 Sat

Tumolo, Giovanni, CP2, 2:15 Wed

Turkiyyah, George, MS57, 5:15 Thu

## U

Uekermann, Benjamin, MS4, 1:25 Wed

*Umemoto, Koichiro, MS88, 2:30 Fri*

*Umemoto, Koichiro, MS99, 9:25 Sat*

Unat, Didem, MS80, 12:50 Fri

Uno, Atsuya, MS21, 3:10 Wed

Uphoff, Carsten, MS65, 10:15 Fri

## V

Valero Lara, Pedro, MS14, 3:35 Wed

*Van Essen, Brian, MS26, 10:50 Thu*

Van Essen, Brian, MS26, 10:50 Thu

*Van Essen, Brian, MS38, 2:40 Thu*

Vanroose, Wim I., IP7, 4:20 Sat

Veerapaneni, Shravan, MS13, 4:00 Wed

Vetter, Jeffrey S., MS41, 3:30 Thu

Vineyard, Craig, MS76, 12:50 Fri

Vinh Truong Duy, Truong, MS13, 4:25 Wed

Vogel, Andreas, MS17, 3:10 Wed

Vogel, James, MS81, 12:00 Fri

*Vuduc, Richard, MS9, 1:00 Wed*

*Vuduc, Richard, MS20, 3:10 Wed*

Vuduc, Richard, MS41, 2:40 Thu

*Vuduc, Richard, MS56, 4:50 Thu*

## W

*Wahib, Mohamed, MS26, 10:50 Thu*

*Wahib, Mohamed, MS38, 2:40 Thu*

Wahib, Mohamed, MS91, 3:20 Fri

Walden, Aaron, MS47, 5:40 Thu

Wang, Hong, MS69, 9:25 Fri

Wang, Weichung, MS22, 3:35 Wed

Washio, Takumi, MS60, 9:50 Fri

Watkins, Jerry, MS48, 5:40 Thu

Watson, Jean-Paul, MS125, 3:45 Sat

Weide, Klaus, MS80, 12:25 Fri

Weinzierl, Tobias, MS80, 12:00 Fri

Weisbecker, Clement, CP4, 4:25 Wed

*Wellein, Gerhard, MS29, 10:50 Thu*

*Wellein, Gerhard, MS41, 2:40 Thu*

Wellein, Gerhard, MS85, 3:20 Fri

*Wilke, Jeremiah, MS100, 9:25 Sat*

Wilke, Jeremiah, MS100, 9:25 Sat

*Wilke, Jeremiah, MS111, 11:35 Sat*

*Wilke, Jeremiah, MS122, 2:30 Sat*

*Wittum, Gabriel, MS6, 1:00 Wed*

Wittum, Gabriel, MS6, 2:15 Wed

*Wittum, Gabriel, MS17, 3:10 Wed*

*Wolf, Felix, MS124, 2:30 Sat*

Wolf, Felix, MS124, 2:55 Sat

Wolf, Michael, MS34, 11:40 Thu

Womeldorff, Geoff, MS47, 5:15 Thu

*Womeldorff, Geoff, MS72, 11:35 Fri*

Wong, Kwai L., PP1, 7:15 Thu

Wong, Kwai L., CP15, 2:55 Sat

Woodward, Carol S., MS70, 10:40 Fri

Wu, Xinzhe, MS89, 2:30 Fri

## X

*Xu, Xiaowen, MS59, 9:25 Fri*

Xu, Xiaowen, MS120, 3:45 Sat

Xue, Wei, MS108, 11:35 Sat

Xuele, Lin, MS69, 9:50 Fri

**Z**

Zhao, Xuan, MS69, 10:15 Fri

Zheng, Ning, MS95, 9:50 Sat

Zhuang, Sicong, MS28, 12:05 Thu

Zohouri, Hamid Reza, MS121, 2:30  
Sat

*Zounon, Mawussi, MS3, 1:00 Wed*

*Zounon, Mawussi, MS14, 3:10 Wed*

Zulian, Patrick, MS126, 3:45 Sat

Notes

