6th JLESC Workshop

November 30 - December 2, 2016

WORKSHOP GUIDEBOOK

[1]

[1] Ver. 29 Nov. 2016

Welcome

Organizers

The workshop is organized by the JLESC partners:

Host Institute

RIKEN Advanced Institute for Computational Science (RIKEN AICS)

- Kimihiko HIRAO
- Akira UKAWA
- Shigeo OKAYA

Local Organizing Committee

- Mitsuhisa SATO
- Naoya MARUYAMA
- Emiko ADACHI
- Tadasu SATO
- Miwako TSUJI

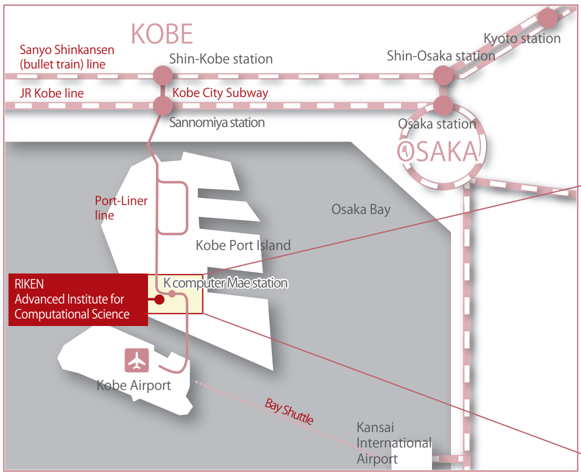
Local Staff

- Yasuhiro SAKAI
- Hisako SASAKI
- Yoko MIYATA
- Manabu YAGI
- Kenta SHIRAI
- Mio KAMEI
- Junko FUKUYOSHI
- Yoshihiro FUKUDA
- Izumi SAKIKAWA
- Mihoko TANAKA
- Mitsuhiro HIRATA
- Teruko KITAGAWA
- Chisato MATSUO
- Tomoko NAKASHIMA

Supporters

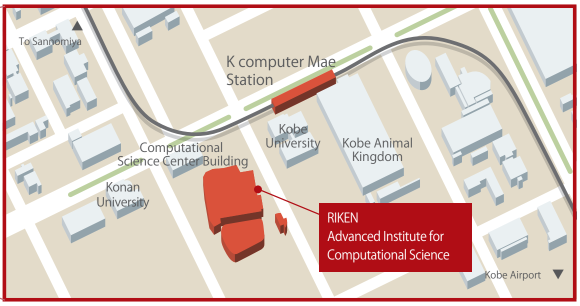
- Hyogo Prefecture
- Kobe City
- Foundation for Computational Science (FOCUS)

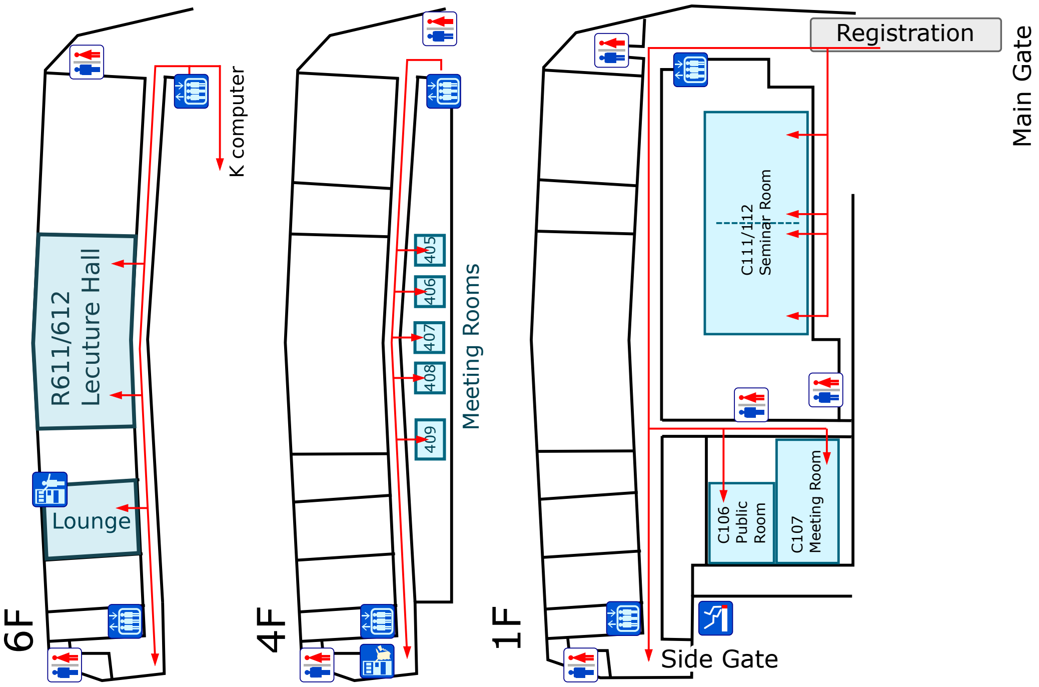
Map



RIKEN
Advanced Institute for
Computational Science

7-1-26 Minatojima-minami-machi, Chuo-ku,
Kobe, Hyogo 650-0047, Japan
http://www.aics.riken.jp/en/





Enter AICS Facility

- ID card should always be with you. You cannot enter AICS without the ID card.
- The participants with ID card can access only to the lecture hall and lounge in the 6th floor, meeting rooms in the 4th floor and rooms for open space from the 1st to 5th floor.
- Entrance (Main Gate) Open : AM 08:30-
- Entrance (Main Gate) Close : – 18:30 After 18h30, you can only exit through the Side Gate of AICS.
- ID card will be issued at the registration of the Main Gate when you come to AICS, and should be returned when you go out of AICS after the workshop at Main/Side gate of AICS.

WiFi Access

SSID	aics-guest
Network key 30. Nov.	9012345678901
Network key 01. Dec.	1098765432109
Network key 02. Dec.	1098765432109
Security option	WPA2-PSK(AES)

Then, open the web browser and follow the instructions.

Eduroam is also available .

Other Notes

- Please follow the instruction of AICS staff when you take pictures because there are some areas that are prohibit taking pictures without permission.
- Smoking is NOT allowed in AICS Facility. Smoking Area is only outside of the Side Gate.
- Eating in the lecture hall is NOT allowed. The lounge in the 6th floor and C111/C112, C107 in the 1st floor are the only place for eating and drinking.
- We have a healthcare center in the 3rd floor. If you feel so bad, please feel free to ask AICS staff.
- Do NOT touch any art works in the building.

Lunch & Dinner

Gala dinner

Place	The restaurant Shushinkan
Time and Date	19:00-21:00 30 Nov.
Note	Buses will departs from AICS at 17:45-18:00

Lunch and Dinner at AICS

Place	AICS Seminar Room (1F) and Lounge (6F)
Time and Date	13:00-14:00 01 Dec.
Time and Date	18:00-20:00 01 Dec.
Note	Buffet lunch/dinner will be served.

Lunch at "Kobe animal kingdom"

Place	Kobe animal kingdom
Time and Date	13:00-14:30 02 Dec.
Note	Buffet lunch will be served. The Kobe animal kingdom is just a few blocks form AICS.

Time Table

Wed. 30 Nov 2016

Time	Room 1 (Lecture Hall, 6F)	Room 2 (Seminar Room, 1F)
12:00-	Registration (Entrance Hall)	
14:00-15:30	Opening Session	
14:00-14:10	Welcome Address – Kimihiko Hirao (RIKEN) Opening – Franck Cappello (ANL)	
14:10-15:00	Keynote – William T.C. Kramer (UIUC)	
15:00-15:15	Opening talk 1 **”Update of FLAGSHIP2020 project and ARMv8 SVE”** Mitsuhisa Sato (RIKEN)	
15:15-15:30	Opening talk 2 **”MareNostrum4: general purpose cluster and emergeing technology clusters”** Jesus Labarta (BSC)	
15:30-16:00	Coffee Break **K computer Visitor's Hall is opened**	 ©RIKEN
16:00-17:20	Session 1: I/O Storage and In-Situ Processing	Session 2: Numerical Methods
16:00-16:20	Project Talk: **”Automatic I/O Scheduling algorithm selection for parallel file systems”** Ramon Nou (BSC) / Francieli Zanon (INRIA)	Project Talk: **”HPC libraries for solving dense symmetric eigenvalue problems”** Inge Gutheil (JSC)
16:20-16:40	Project Talk: **”Toward taming large and complex data flows in data-centric supercomputing”** Francois Tessier (ANL)	Project Talk: **”Shared Infrastructure for Source Transformation Automatic Differentiation”** Sri Hari Krishna Narayanan (ANL)
16:40-16:50	Individual Talk: **”Mochi: composable lightweight data services for HPC”** Philip Carns (ANL)	Individual Talk: **”Exploring eigensolvers for large sparse non-Hermitian matrices”** Hiroya Suno (RIKEN)
16:50-17:00	Individual Talk: **”NVRAM POSIX-like Filesystem with I/O hints support”** Alberto Miranda/Ramon Nou (BSC)	Individual Talk: **”Towards Automated Load Balancing via Spectrum Slicing for FEAST-like solvers”** Jan Winkelmann (JSC)
17:00-17:10	Individual Talk: **”An Argument for On-Demand and HPC Convergence”** Kate Keahey (ANL)	Individual Talk: **”Bidiagonalization with Parallel Tiled Algorithms”** Julien Langou (INRIA)
17:10-17:20	Individual Talk: **”Significantly Improving Lossy Compression for Scientific HPC Data based on Multi-dimensional Prediction and Error-controlled Quantization”** Sheng Di (ANL)	Individual Talk: **”Targeting Interface Problems at Scale with Coupled Elliptic Solvers”** Natalie Beams (UIUC)
17:45-18:00	Bus Departure @ Entrance	
19:00-	Dinner @ Shushinkan	

Thu. 01 Dec 2016

Time	Room 1 (Lecture Hall, 6F)	Room 2 (Seminar Room, 1F)
09:00-10:30	Session 3: Performance Tools & I/O Storage and In-Situ Processing	Session 4: I/O Storage and In-Situ Processing & Advanced Architectures
09:00-9:20	Project Talk: **”Developer tools for porting and tuning parallel applications on extreme-scale parallel systems”** Christian Feld (JSC)	Project Talk: **”Exploiting the Dragonfly Topology to Improve Communication Operations in MPI”** Nathanaël CHERIERE (INRIA)
09:20-9:40	Project Talk: **”SSSP: Simplified Sustained System Performance Metric”** Miwako Tsuji (RIKEN)	Individual Talk: **”In situ workflow tools: Technologies and opportunities”** Justin M Wozniak (ANL) Individual Talk: **”Early Work in the Application of Deep Learning for Scientific Visualization”** Rob Sisneros (NCSA)
09:40-10:00	Project Talk: **”Use of the Folding profiler to assist on data distribution for heterogeneous memory systems”** Antonio J. Peña (BSC)	Individual Talk: **”Performance of High level FPGA accelerators”** Carlos Alvarez (BSC) Individual Talk: **”Supporting extreme-scale real-time science workflows on exascale computers ”** Raj Kettimuthu (ANL)
10:00-10:10	Individual Talk: **”Multi objective optimization of HPC kernels for performance, power, and energy”** Prasanna Balaprakash (ANL)	Individual Talk: **”Towards efficient Big Data processing in HPC systems: Performance Analysis of Spark on HPC ”** Orcun Yildiz (INRIA)
10:10-10:20	Individual Talk: **”Some causes about performance fluctuations of applications”** Kiyoshi Kumahata (RIKEN)	Individual Talk: **”A Suite of Collaborative Heterogeneous Applications for Integrated Architectures”** Simon Garcia De Gonzalo (UIUC)
10:20-10:30	Individual Talk: **”HPC-Tools JUBE, LLview and SIONlib at JSC: Recent developments”** Wolfgang Frings (JSC), Sebastian Lehrs (JSC), Kay Thust (JSC)	Individual Talk: **”Exploring Memory Management and Performance on Deep-Memory Architectures”** Swann Perarnau (ANL)
10:30-11:00	Coffee Break	Coffee Break
11:00-12:30	Session 5: Performance Tools BOS	Session 6: I/O Storage and In-Situ Processing BOS
11:00-12:30	BOS **”Interfacing Task-based Runtimes with Performance Tools”** Bernd Mohr (JSC) [1h30min]	BOS **”Convergence of Cloud, BigData, and HPC ”** Kate Keahey (ANL), Gabriel Antoniu (INRIA), Rosa Badia (BSC) [1h30min]
12:30-13:00	Group Photo @ Visitor’s Hall	
13:00-14:00	Lunch	Lunch

Time	Room 1 (Lecture Hall, 6F)	Room 2 (Seminar Room, 1F)
14:00-15:30	Session 7: Numetical Methods & Applications	
14:00-14:20	Project Talk: **"Reducing Communication in Sparse Iterative and Direct Solvers"** William Gropp (UIUC)	
14:20-14:40	Project Talk: **"Comparison of Meshing and CFD Methods for Accurate Flow Simulations on HPC systems"** Andreas Lintermann (JSC), Keiji Onishi(RIKEN)	
14:40-14:50	Individual Talk: **"Handling Pointers and Dynamic Memory in Algorithmic Differentiation"** Sri Hari Krishna Narayanan (ANL)	
14:50-15:00	Individual Talk: **"Coupled multiphysics and parallel programming"** Mariano Vazquez (BSC)	
15:00-15:20	Project Talk: **"Optimizing ChASE eigensolver for Bethe-Salpeter computations on multi-GPUs"** Edoardo Di Napoli (JSC)	
15:20-15:30	Individual Talk: **"Extreme-scaling applications at JSC"** Brian Wylie (JSC), Dirk Broemmel (JSC), Wolfgang Frings (JSC)	
15:30-16:00	Coffee Break	Coffee Break
16:00-17:15	Session 8: Applications BOS	
16:00-17:15	BOS **"HPC Linear Algebra in Materials Science"** Takahito Nakajima (RIKEN) [1h15min]	
18:00-20:00	Dinner	Dinner

Fri. 02 Dec 2016

Time	Room 1 (Lecture Hall, 6F)	Room 2 (Seminar Room, 1F)
09:00-10:30	Session 9: Resillience	Session 10: Programming and Runtime
09:00-09:20	Project Talk: **”Optimization of Fault-Tolerance Strategies for Workflow Applications”** Aurélien Cavelan (INRIA)	Project Talk: **”Scalability Enhancements to FMM for Molecular Dynamics Simulations”** David Haensel (JSC)
09:20-09:30	Individual Talk: **”Failure detection”** Yves Robert (INRIA)	Individual Talk: **”Daino: A High-level AMR Framework on GPUs”** Mohamed Wahib (RIKEN)
09:30-09:40	Individual Talk: **”Identification and imapct of failure cascades”** Frederic Vivien (INRIA)	Individual Talk: **”Portable Asynchronous Progress Model for MPI on Multi- and Many-Core Systems”** Min Si (ANL)
09:40-10:00	Project Talk: **”New Techniques to Design Silent Data Corruption Detectors”** Leonardo Bautista (BSC)	Individual Talk: **”Efficient Composition of task-graph based Code”** Christian Perez (INRIA)
		Individual Talk: **”In C++ we trust – performance portability out of the box?”** Ivo Kabadshow (JSC)
10:00-10:10	Individual Talk: **”Runtime driven online estimation of memory vulnerability”** Luc Jaulmes (BSC)	Individual Talk: **”Feedback control for Autonomic High-Performance Computing”** Eric Rutten (INRIA)
10:10-10:20	Individual Talk: **”Fault Tolerance and Approximate Computing”** Osman Unsal (BSC)	Individual Talk: **”An Overview of the IHK/McKernel Lightweight Multikernel OS for Kernel-Assisted Communication Progression”** Balazs Gerofi (RIKEN)
10:20-10:30	Individual Talk: **”Fault-tolerance for FPGAs”** Osman Unsal (BSC)	Individual Talk: **”New Parallel Execution Model - User-Level Implementation”** Atsushi Hori (RIKEN)
10:30-11:00	Coffee Break	Coffee Break
11:00-12:45	Session 11: Programming and Runtime & Applications BOS	Session 12: Advanced Architectures & Resillience
11:00-11:10	Project Talk: **”Enhancing Asynchronous Parallelism in OmpSs with Argobots”** Jesus Labarta (BSC)	Individual Talk: **”Leveraging FPGA technology for next-generation high-performance computing systems”** Kazutomo Yoshii (ANL)
11:10-11:20		Individual Talk: **”Lessons learned from HPRC”** Volodymyr Kindratenko (UIUC)
11:20-11:30	Individual Talk: **”Exploring RDMA libraries for PGAS language”** Tetsuya Odajima (RIKEN)	BOS **”Memory Errors at Extreme Scale”** Leonardo Bautista Gomez (BSC) [1h]
11:30-12:30	BOS **”HPC challenge of LQCD and related”** Yoshifumi Nakamura (RIKEN) [1h15min]	
12:20-12:45		Discussion and move to Room 1
12:45-13:00	Closing Session	
12:45-13:00	Closing – Mitsuhisa Sato (RIKEN) and William T.C. Kramer (UIUC)	
13:00-14:30	Lunch @ Kobe Animal Kingdom	

For more details, please refer:
http://www.aics.riken.jp/workshop/jlesc/6th_jlesc_book.pdf

Opening Session

Welcome Address
Kimihiko Hirao (RIKEN)

Opening
Franck Cappello (ANL)

Keynote: **TBD**
Bill Kramer (UIUC)

Talk: **Update of FLAGSHIP2020 project and ARMv8 SVE**
Mitsuhisa Sato (RIKEN)
FLAGSHIP2020 project has been launched to develop and deploy the post-K computer since 2014, and now the basic design of the system has been finished. We have decided to use ARM v8 SVE (Scalable Vector Extension), defined by ARM, with Fujitsu's extensions as the CPU architecture of the computing node processor. I will report the update on the project and present the feature of new ARM SVE.

Talk: **MareNostrum4: general purpose cluster and emergeing technology clusters**
Jesus Labarta (BSC)

Session 1

1.1 Project Talk: **Automatic I/O Scheduling algorithm selection for parallel file systems**
"Automatic I/O Scheduling algorithm selection for parallel file systems - Integration and first results"
Ramon Nou (BSC) / Francieli Zanon (INRIA)
We introduced the ability to use intra-workload I/O scheduling changes in AGIOS (presented at 3rd JLESC Workshop), guided by two complementary methods: Armed Bandid, a probability-guided approach when we do not know the workload and markov chain using pattern matching to predict which is the next best scheduler for the next period. Results with AB are obtaining high performance over the standard OrangeFS. AGIOS is from UFGRS/INRIA and BSC used AB and pattern matching for automatic kernel I/O scheduler changes. Next steps will include the pattern matching incorporation to the system.

1.2 Project Talk: **Toward taming large and complex data flows in data-centric supercomputing**
"Topology-Aware Data Aggregation for Intensive I/O on Large-Scale Supercomputers"
Francois Tessier (ANL)
Reading and writing data efficiently from storage systems is critical for high performance data-centric applications. These I/O systems are being increasingly characterized by complex topologies and deeper memory hierarchies. Effective parallel I/O solutions are needed to scale applications on current and future supercomputers. Data aggregation is an efficient approach consisting of electing some processes in charge of aggregating data from a set of neighbors and writing the aggregated data into storage. Thus, the bandwidth use can be optimized while the contention is reduced. In this work, we take into account the network topology for mapping aggregators and we propose an optimized buffering system in order to reduce the aggregation cost. We validate our approach using micro-benchmarks and the I/O kernel of a large-scale cosmology simulation. We show improvements up to 15x faster for I/O operations compared to a standard implementation of MPI I/O.

1.3 Individual Talk: **Mochi: composable lightweight data services for HPC**
Philip Carns (ANL)
Parallel file systems have formed the basis for data management in HPC for decades, but specialized data services are increasingly prevalent in areas such as library management, fault tolerance, code coupling, burst buffer management, and in situ analytics. These specialized data services are effective because they can more readily provide semantics and functionality tailored to the task at hand than a one-size-fits-all storage system. Adoption of specialized data services is limited by their flexibility and portability, however. Our goal in Mochi is to develop a reusable collection of HPC micro-services that can be composed and customized to rapidly construct new domain-specific or even application-specific data services. We present updates from services under development that highlight this functionality and explore how to lower the barrier to entry for new HPC data services.

1.4 Individual Talk: **NVRAM POSIX-like Filesystem with I/O hints support**
Alberto Miranda/Ramon Nou (BSC)

As a work in progress for the NEXTGenIO european project, the filesystem will transparently work as a collaborative burst buffer between all the NVDIMMs (or other technologies) available on the compute nodes. It will support user I/O hints to specify distribution and data's lifecycle. Collaboration for applications, and testing environments.

1.5 Individual Talk: **An Argument for On-Demand and HPC Convergence**

Kate Keahey (ANL)

There is currently a divergence in the scientific community. On one hand, we have HPC resources, typically managed by batch schedulers, which offer very powerful capabilities but do not satisfy user QoS issues such as on-demand availability. On the other hand therefore, we have relatively small clusters, typically operated by experimental facilities that provide controlled availability for the users but are under-utilized and do not have the resources to scale. We propose an experiment that combines those resources in an approach based mostly on commodity software technologies and explores the consequences of such a merger. We evaluate our approach experimentally by examining two years' worth of traces from the experimental cluster at the Advance Photon Source at ANL and batch trace from the Lab Computing Resource Center (LCRC) at ANL and enacting selected scenarios on hundreds of nodes. Our results demonstrate significant benefits in cost, utilization, and availability.

1.6 Individual Talk: **Significantly Improving Lossy Compression for Scientific HPC Data based on Multidimensional Prediction and Error-controlled Quantization**

Sheng Di (ANL)

Today's HPC applications are producing extremely large amounts of data, such that data storage and its performance are becoming a serious problem for scientific research. In this work, we design a new error-controlled lossy compression algorithm for the large-scale high-entropy scientific data sets. Our key contribution is significantly improving the prediction hitting rate (or accuracy) for each data point based on its nearby data values along multiple dimensions. On the one hand, we carefully derive a series of multi-layer prediction formulas in the context of data compression. One serious challenging issue is that the data prediction has to be performed based on the decompressed values during the compression for guaranteeing the error bounds, which may degrade the prediction accuracy in turn. As such, we explore the best layer for the prediction by considering the impact of decompression errors on the prediction accuracy. On the other hand, we propose an adaptive error-controlled quantization encoder, which can further improve the prediction hitting rate a lot. The data size can be reduced significantly after performing the variable-length encoding because of the fairly uneven distribution produced by our quantization encoder. We evaluate the new compressor by production scientific data sets, and compare it to many other state-of-the-art compressors, including GZIP, FPZIP, ZFP, SZ, ISABELA, etc.. Experiments show that our compressor is the best in class, especially on compression factors (or bit-rates) and compression errors (including RMSE, NRMSE, PSNR, etc.). Our solution is better than the second-best solution ZFP by nearly 2.3x increase in compression factor and 5.4x reduction in normalized root mean squared error on average with reasonable error bounds and user-desired bit-rates.

Session 2

2.1 Project Talk: **HPC libraries for solving dense symmetric eigenvalue problems "Comparison of library eigensolvers for dense symmetric matrices on K-computer, JUQUEEN, and JURECA"**

Inge Gutheil (JSC)

In many applications for example in Density Functional Theory (DFT) used in physics, chemistry, and materials science the computation of eigenvalues and eigenvectors of dense symmetric matrices is an important issue. There are three modern libraries for the solution of this problem, EigenExa, ELPA, and Elemental. They behave different on different computer architectures and we will show which library should be preferred on the three different compters, K-computer, BlueGene/Q (JUQUEEN) and a cluster of Intel processors (JURECA).

2.2 Project Talk: **Shared Infrastructure for Source Transformation Automatic Differentiation "Handling Pointers and Dynamic Memory in Algorithmic Differentiation"**

Sri Hari Krishna Narayanan (ANL)

Proper handling of pointers and the (de)allocation of dynamic memory in the context of an adjoint computation via source transformation has so far had no established solution that is both comprehensive and efficient. This talk gives a categorization of the memory references involving pointers to heap and stack memory along with principal options to recover addresses in the reverse sweep. The main contributions are a code analysis algorithm to determine which remedy applies, memory mapping algorithms for the general case where one cannot assume invariant absolute addresses and an algorithm for the handling of pointers upon restoring checkpoints that reuses the memory mapping approach for the reverse sweep.

2.3 Individual Talk: **Exploring eigensolvers for large sparse non-Hermitian matrices**

Hiroya Suno (RIKEN)
We are exploring ways for computing eigenvalues and eigenvectors of large sparse non-Hermitian matrices, such as those arising in Lattice Quantum Chromodynamics (lattice QCD) simulation. We have been exploring so far the Sakurai-Sugiura (SS) method, a method based on a contour integral, which allows us to compute desired eigenvalues located inside a given contour of the complex plane, as well as the associated eigenvectors. We have tested the SS method with large sparse matrices with the matrix order being up to about one billion, and have been able to compute eigenvalues for several simple cases with a certain accuracy. We are now ready to explore some other eigensolvers, such as ARPACK (Arnoldi Package) and ChASE (Chebyshev Accelerated Subspace iteration Eigensolver).

2.4 Individual Talk: **Towards Automated Load Balancing via Spectrum Slicing for FEAST-like solvers**
Jan Winkelmann (JSC)
Subspace iteration algorithms accelerated by rational filtering, such as FEAST, have recently re-emerged as a research topic in solving for interior eigenvalue problems. FEAST-like solvers are Rayleigh-Ritz solvers with rational filter functions, and as a result require re-orthogonalization on long vectors only in rare cases. Application of the filter functions, the computationally most expensive part, offers three levels of parallelism: 1) multiple spectral slices, 2) multiple linear system solves per slice, and 3) multiple right-hand sides per system solves. While the second and third level of parallelism are currently exploited, the first level is often difficult to efficiently realize. An efficient algorithmic procedure to load-balance multiple independent spectral slices is not yet available. Currently, existing solvers must rely on the user's prior knowledge. An automatic procedure to split a user specific interval into multiple load-balanced slices would greatly improve the state of the art. We outline how, both the algorithmic selection of filter functions and the spectral slices, can be at the center of load-balancing issues. Additionally, we present the tools and heuristics developed in an effort to tackle the problems.

2.5 Individual Talk: **Bidiagonalization with Parallel Tiled Algorithms**
Julien Langou (INRIA)
"In a recent paper, we considered algorithms for going from a ""full"" matrix to a condensed ""band bidiagonal"" form using orthogonal transformations. We use the framework of ""algorithms by tiles"". We considered many reduction trees and obtained conclusive results on parallel distributed experiments on a cluster of multicore nodes. We will present these results. Based on these encouraging results, we will discuss the following five open problems. We believe the five open problems below are relevant to JLESC. (1) Applying the same techniques to symmetric tridiagonalization methods for the symmetric eigenvalue problem, (2) Impact (storage and computation time) on the computation of the singular vectors (and the eigenvectors in the symmetric eigenvalue problem case), (3) Performing experiments on very large scale machine to see the scalability of the methods, (4) Examining the trade-off between TS and TT kernels, (5) Experiment with scalable parallel distributed solution for going to band bidiagonal (or tridiagonal) to bidiagonal form. "

2.6 Individual Talk: **Targeting Interface Problems at Scale with Coupled Elliptic Solvers**
Natalie Beams (UIUC)
The creation, adaptation, and maintenance of volume meshes that conform to problem geometry is costly and represents a scaling challenge. Furthermore, semi-structured meshes offer distinct computational advantages over fully unstructured meshes. We present a family of methods that permits the enforcement of a broad class of boundary and interface conditions on surfaces that do not coincide with element boundaries while maintaining high-order accuracy. Our methods leverage the advantages of finite element and integral equation methods in order to solve elliptic problems on a (potentially) structured volume mesh with embedded domains. A benefit of this approach is that standard, un-modified finite element basis functions can be used, in contrast to immersed finite element methods. Additionally, the computational mechanics for the integral equation portion of the coupled solution remain unchanged. One limiting factor in our methodology (and in many other simulations) is the dependence on a scalable fast multipole method. We discuss implications in a parallel setting and potential directions for collaborations in the Joint Lab.

Session 3

3.1 Project Talk: **Developer tools for porting and tuning parallel applications on extreme-scale parallel systems**
"Developer tools project update"
Christian Feld (JSC)
Developments in the partners' tools will be reported, particularly the design and initial prototyping of XMPT modelled on the OMPT tools interface for OpenMP, which is expected to facilitate measurement of applications using XMP with Extrae, Score-P and other tools. We also provide an update on recent and planned training with our tools, and ongoing work to define a common performance analysis terminology, methodology, and efficiency metrics

for MPI and multithreaded applications.

3.2 Project Talk: **SSSP: Simplified Sustained System Performance Metric**
"Recent Update about SSSP"

Miwako Tsuji (RIKEN)

"The SSP (Sustained System Performance) metric is used to measure the performance of existing and future supercomputer systems at NERSC, NCSA, the Australian Bureau of Meteorology and other sites. The SSP metric takes into account the performance of various scientific applications and input data sets (aka a ""benchmark""), which represent some part of the sites' workload. In this collaboration, we propose the SSSP (Simplified Sustained System Performance) metric that makes performance projection using a set of simple existing benchmarks to the SPP metric for real applications. The benchmarks used as the ""simple"" may be existing simple benchmarks such as HPCC benchmark, HPL, parts of the SPECFP benchmark, and other simplified pseudo benchmarks which data already exist or easy to be measured. "

3.3 Project Talk: **Use of the Folding profiler to assist on data distribution for heterogeneous memory systems**
"Profiler-assisted data distribution for heterogeneous memory systems: getting close"

Antonio J. Peña (BSC)

In this project we aim at using the Extrae profiler along with its Folding capabilities to provide optimized data distributions for heterogeneous memory systems based on coarse-grained sampling of hardware counters. In this talk we present the latest progress on this project on both the profiling tool and the programming model sides.

3.4 Individual Talk: **Multi objective optimization of HPC kernels for performance, power, and energy**

Prasanna Balaprakash (ANL)

Code optimization in the high-performance computing realm has traditionally focused on reducing execution time. The problem, in mathematical terms, has been expressed as a single-objective optimization problem. The expected concerns of next-generation systems, however, demand a more detailed analysis of the interplay among execution time and other metrics. Metrics such as power, performance, energy, and resiliency may all be targeted together and traded against one another. We present a multi-objective formulation of the code optimization problem and a machine-learning-based search algorithm. Our proposed framework helps one explore potential tradeoffs among multiple objectives and provides a significantly richer analysis than can be achieved by treating additional metrics as hard constraints. We empirically examine a variety of metrics, architectures, and code optimization decisions and provide evidence that such tradeoffs exist in practice.

3.5 Individual Talk: **Some causes about performance fluctuations of applications**

Kiyoshi Kumahata (RIKEN)

"During the operation of the K computer, running time of an application occasionally becomes longer or shorter than previously measured time under the same conditions. We call this ""running time fluctuation"". Running time fluctuations disturb the efficient operation of a supercomputer. And it waste precious computer resources. Thus, we have been investigating and resolving such issue on the K computer. These issues may occur in many applications and supercomputers. In this talk, some causes of running time fluctuations that we have ever encountered in the past are introduced."

3.6 Individual Talk: **HPC-Tools JUBE, LLview and SIONlib at JSC: Recent developments**

Wolfgang Frings (JSC), Sebastian Lehrs (JSC), Kay Thust (JSC)

In this talk we will present the recent developments of the benchmarking environment JUBE, the batch system monitoring tool LLview, and the parallel I/O library SIONlib. In detail, we will present how the benchmarking environment JUBE is integrated into a performance evaluation workflow, which is applied in the EU-project EoCoE. For LLview we will show the recently implemented job-based monitoring of I/O metrics. For SIONlib we will focus on the work we did in the EU-project DEEP-ER to support the efficient use of node-local storage in parallel task-local I/O. For all tool we show our future plans and present possible topics for collaboration.

Session 4

4.1 Project Talk: **Exploiting the Dragonfly Topology to Improve Communication Operations in MPI**
"Topology-Aware Scatter and AllGather operations for Dragonfly Networks"

Nathanaël CHERIERE (INRIA)

"High-radix direct network topologies such as Dragonfly have been proposed for petascale and exascale supercomputers because they ensure fast interconnections and reduce the cost of the network compared with traditional network topologies. The design of new machines such as Theta with a Dragonfly network present an opportunity to further improve the performance of distributed applications by making the algorithms aware of the topology. Indeed, current algorithms do not consider the topology and thus lose numerous opportunities of optimization for performance that have been created by the topology. This talk describes optimized algorithms for two collective operations: AllGather and Scatter and presents the results of an evaluation using the CODES simulator. *Note: this talk is the result of a 5-month internship of Nathana¨el CHEERIER (INRIA) at ANL. The project emerged recently and will be added to the JLESC web site (topic: I/O, storage and in situ processing)."

4.2 Individual Talk: **In situ workflow tools: Technologies and opportunities**

Justin M Wozniak (ANL)

This talk will present recent advances in programming interfaces for Decaf-based in situ workflows.

4.3 Individual Talk: **Early Work in the Application of Deep Learning for Scientific Visualization**

Rob Sisneros (NCSA)

In recent years, deep learning has become a trusted path towards intelligent automation and learning in various fields. Large scale image classification, computer vision, text analysis, 3D model reconstruction, etc. are only some of the areas that have benefited from deep learning. Its applications in scientific visualization however, have seldom been studied. A reason for this may be differences between how visualization scientists and machine learning scientists perceive the user's role in data analysis; while machine learning tries to eliminate the user by means of intelligent automation, visualization benefits from unparalleled human intelligence. We believe it is possible to bridge this gap to the benefit of scientific visualization by automating complex and daunting tasks that result in visualizations. In this work we investigate deep learning transfer function design. We will discuss the elements of an automatic technique that utilizes state of the art deep learning and evolutionary optimization to create transfer functions based on sample target images. Even in this early stage the approach has shed light on the explorative process of transfer function design and shows promise to deliver impactful data insights and help free users to focus on analyzing over generating scientific visualization results.

4.4 Individual Talk: **Performance of High level FPGA accelerators**

Carlos Alvarez (BSC)

Using high level programming models to program FPGA accelerators through automatic tool-chains is a new field that seeks to drive the new heterogeneous platforms that are currently appearing in the market. However, as in any new field, many questions, as its efficiency and/or real productivity, remain unanswered. This talk will present the results obtained when analyzing the performance obtained using this new tools. The talk will describe how Paraver traces can help to analyze the performance of the resulting HDL accelerators obtained and highlight the problems discovered and the envisioned solutions.

4.5 Individual Talk: **Supporting extreme-scale real-time science workflows on exascale computers**

Raj Kettimuthu (ANL)

Activities at experimental facilities such as light sources and fusion tokamaks are tightly scheduled, with timing driven by factors ranging from the physical processes involved in an experiment to the travel schedules of on-site researchers. Thus, computing must often be available at a specific time, for a specific period, with a high degree of reliability. Such near-realtime requirements are hard to meet on current HPC systems, which are typically batch-scheduled under policies in which an arriving job is run immediately only if enough resources are available, and is queued otherwise. We are investigating the following aspects and are looking for collaboration opportunities: 1) What changes will be required to the scheduling algorithms, architecture, and implementation of exascale computers if they are to support real-time experimental science workloads effectively? 2) What are implications for other exascale workloads? 3) What system-level support can be beneficial? We would like to examine a wide range of design alternatives, develop new system models and simulation methods, and perform extensive simulation-based (and real-world) studies using various combinations of real-world batch job traces and synthetic real-time jobs (created based on the model that mimics the actual real-time jobs) to answer these questions.

4.6 Individual Talk: **Towards efficient Big Data processing in HPC systems: Performance Analysis of Spark on HPC**

Orcun Yildiz (INRIA)

On paving the way towards convergence of HPC and Big Data, adoption of Big Data processing frameworks into HPC systems remains a challenge. In this work, we conduct an experimental campaign to provide a clearer understanding of the performance of Big Data processing frameworks on HPC systems. In this talk, we present the results of our campaign together with the insights and open questions on how to design efficient Big Data processing solutions for HPC systems. We believe that our findings can interest participants who is working on the convergence between HPC and Big Data. Moreover, we hope that this talk will setup potential collaboration(s) towards efficient Big Data processing in HPC systems.

4.7 Individual Talk: **A Suite of Collaborative Heterogeneous Applications for Integrated Architectures**

Simon Garcia De Gonzalo (UIUC)

Heterogeneous systems are evolving into computing platforms with tighter integration between CPU and GPU. This is possible thanks to new features such as shared memory space, memory coherence, and system-wide atomic operations. Exponents of this trend are the Heterogeneous System Architecture (HSA) and the NVIDIA Pascal architecture. Programming frameworks such as OpenCL 2.0 and CUDA 8.0 allow programmers to exploit these platforms with fine-grain coordination of CPU and GPU threads. To evaluate these new architectures and programming languages, and to empower researchers to experiment with new ideas, a suite of benchmarks targeting these architectures with close CPU-GPU collaboration is needed. We present Chai (Collaborative Heterogeneous Applications for Integrated-architectures), a benchmarks suite that leverage the latest features of heterogeneous architectures. These benchmarks cover a wide range of collaboration patterns, exhibit great diversity within each pattern, and are each implemented in five different programming models: OpenCL 2.0, OpenCL 1.2, C++ AMP, CUDA 7.5, and CUDA-Sim, with CUDA 8.0 in progress.

4.8 Individual Talk: **Exploring Memory Management and Performance on Deep-Memory Architectures**

Swann Perarnau (ANL)

Hardware advances are enabling new memory technologies resulting in a deepening of the memory hierarchy. On-package DRAM, now available on Intel's Knights Landing architecture, provides higher bandwidth but limited capacity. Soon, byte-addressable NVRAM will also become available, deepening further the memory hierarchy. The Argo group at Argonne recently started exploring how this memory hierarchy should be exposed to HPC applications. Our approach follows several axes: 1) improve and augment current operating system interfaces to allow the available memory types to be explicitly managed efficiently and transparently. 2) develop tools to analyze the memory access patterns of HPC application, focusing on providing guidance on the use of the memory hierarchy and the above interfaces. 3) integrate automatic memory management facilities into parallel runtimes, so that applications can benefit from better usage of the memory hierarchy transparently and effortlessly. At the OS level, we are currently designing a low-level framework to perform asynchronous memory migration between nodes of the memory hierarchy. Early results indicate that the use of this framework along with out-of-core programming schemes can significantly improve the performance of application whose working set cannot fit in on-package memory. This summer, we prototyped a memory tracing and analysis toolset to extract from HPC application the locality of memory access to specific data structures. This can be then used to identify which data structures would benefit from being migrated across the memory hierarchy, either statically at allocation time or at runtime using our memory migration framework. We are also developing models and heuristics to guide automatic memory migration inside runtimes like StarPU, OmpSS or OpenMP 4. While we already collaborate with JLESC members (RIKEN and INRIA) for this work, we would like to extend these collaborations to other partners and formally establish (or join) JLESC projects to help those collaborations. During this presentation, we will outline the current state of this work and suggest possible collaboration points.

Session 5

5.1 Break Out session: **Interfacing Task-based Runtimes with Performance Tools**

Char: Bernd Mohr (JSC)

Speakers:
Bernd Mohr(JSC) "Overview/Problem Statement"
Hitoshi Murai(RIKEN) "The Performance Tool API of XMP"
Judit Gimenez(BSC) "Measurements with Extra-E via OMP-T"
Christian Feld (JSC) "Measurements with Score-E via OMP-T"
not-yet-confimed (BSC) "OMP-T support of the OmpSs runtime system"
not-yet-confirmed (INRIA) "Interfacing Star-PI with Performance Tools"

Providing performance measurement and analysis capabilities for a parallel programming system requires to implement a monitoring API in the parallel runtime and the necessary matching measurement component in each performance tool. While MPI had a standard monitoring API (PMPI, now also MPI-T) from the beginning, thread- or task-based programming systems either do not provide such a feature or only proprietary interfaces. Currently, a standard interface for OpenMP also covering tasks and accelerators is under discussion. The session intents to bring together researchers from the performance tool and parallel runtime system areas to discuss the feasibility to use OMP-T also as interface for other task-based systems like OmpSs, XMP, or Star-PU and what extensions or additions would be needed to support non-OpenMP models.

Session 6

6.1 Break Out session: **Convergence of Cloud, BigData, and HPC**
Char: Kate Keahey (ANL), Gabriel Antoniu (INRIA), Rosa Badia (BSC)

Speakers:
In addition to the chairs, we will have talks from Ramon Nou (BSC), Bruno Raffin (INRIA). We will also solicit additional speakers from these institutions as well as from NCSA.

Unprecedented growth of opportunities in experimental sciences, together with on-demand access, control over computing environment provided by containers and virtualization, and streamlined data processing techniques from the cloud world are revolutionizing the ways in which we can do science. They also however raise questions about how these new datacenter management and programming techniques relate to traditional HPC. Those questions introduce the potential for innovation across the stack including technology for node management (e.g., containers versus virtualization, integration of smart NICs and accelerators), storage (e.g., explicit QoS in storage, exploiting storage hierarchies), networking (SDN, network reservations), resource management (e.g., managing dynamicity, multi-aspect leases), models (improving predictability versus embracing unpredictability, ability to understand and express trade-offs), frameworks (emphasis on loosely coupled programming models, workflows, approaches to resilience, programmable platforms, malleable applications), data (the emergence of data-centric abstractions, data consistency models). We propose to organize a breakout session to organize and discuss specific ideas and formulate collaborations and partnerships on the convergence of cloud, BigData and HPC technologies across the software stack. The session will be structured around presentation foci in each of the areas described above and will seek to chart out both present and future collaboration on innovative topics.

Session 7

7.1 Project Talk: **Reducing Communication in Sparse Iterative and Direct Solvers**
"Reducing Communication in Sparse Iterative and Direct Solvers"
William Gropp (UIUC)
The focus of this project is on reducing the communication overhead in sparse kernels, with application to both iterative and direct solvers. Communication represents a significant challenge, particularly as architectures move to more layers of parallelism. In this talk we highlight several steps in addressing communication limitations, including the development of an improved performance model and a new approach to handling communication in the sparse matrix-vector multiplication operation by increasing locality. Our recent findings are highlighting several areas that are rich for collaborative opportunities through the JointLab and we discuss the next steps in continuing to develop these methods.

7.2 Project Talk: **Comparison of Meshing and CFD Methods for Accurate Flow Simulations on HPC systems**
"Recent Scaling-Enabling Developments of the CFD codes CUBE and ZFS"
Andreas Lintermann (JSC), Keiji Onishi(RIKEN)
The talk focuses on recent developments in the CFD-codes CUBE from AICS, Riken, and ZFS from JSC. It discusses efforts to increase the performance of both codes and to extend their applicability to various engineering problems, i.e., to the technical and biofluid-mechanical engineering realm. The methodological discussion will be complemented by examples from different applications.

7.3 Individual Talk: **Handling Pointers and Dynamic Memory in Algorithmic Differentiation**
Sri Hari Krishna Narayanan (ANL)
Proper handling of pointers and the (de)allocation of dynamic memory in the context of an adjoint computation via source transformation has so far had no established solution that is both comprehensive and efficient. This talk gives a categorization of the memory references involving pointers to heap and stack memory along with principal options to recover addresses in the reverse sweep. The main contributions are a code analysis algorithm to determine which remedy applies, memory mapping algorithms for the general case where one cannot assume invariant absolute addresses and an algorithm for the handling of pointers upon restoring checkpoints that reuses the memory mapping approach for the reverse sweep.

7.4 Individual Talk: **Coupled multiphysics and parallel programming**
Mariano Vazquez (BSC)
In this talk, BSC will address problems encountered to implement parallel multiphysics in Alya. Several issues appear at MPI tasks level and are related with point-to-point data interchange, where a correct communication pattern

must be establish to avoid bottlenecks. We have explored different solutions: different numerical schemes, solution strategies, runtime load balancing... We would like to share our experiences (good and bad ones) to seek for help and arise discussion.

7.5 Project Talk: **Optimizing ChASE eigensolver for Bethe-Salpeter computations on multi-GPUs**
"Efficient parallel implementation of the ChASE library on distributed CPU-GPU computing architectures"
Edoardo Di Napoli (JSC)
The Chebyshev Accelerated Subspace iteration Eigensolver (ChASE) is an iterative eigensolver developed at the JSC by the SimLab ab initio. The solver targets principally sequences of dense eigenvalue problems as they arise in Density functional Theory, but can also work on the single eigenproblem. ChASE leverages on the preponderant use of BLAS 3 subroutines to achieve close-to-peak performance. Currently, the library can be executed in parallel on many- and multi-core platforms. The latest development of this project dealt with the extension of the CUDA build to encompass multiple GPUs on distinct CPUs. As such, this hybrid parallelization will use MPI as well as CUDA interfaces effectively exploiting heterogeneous multi-GPU platforms. The extended library was tested on large and dense eigenproblems extracted from excitonic Hamiltonian. The ultimate goal is to integrate this new parallel implementation of ChASE with the VASP-BSE code.

7.6 Individual Talk: **Extreme-scaling applications at JSC**
Brian Wylie (JSC), Dirk Broemmel (JSC), Wolfgang Frings (JSC)
Since 2006 JSC has held a series of well-received Extreme Scaling Workshops for its Blue Gene systems, and with the High-Q Club has documented 28 application codes that successfully scaled to exploit the full 28 racks (with 458752 cores capable of running over 1.8 million threads) of its JUQUEEN Blue Gene/Q. We briefly review these activities and what might be lessons for future exascale computer systems.

Session 8

8.1 Break Out session: **HPC Linear Algebra in Materials Science**
Char: Takahito Nakajima (RIKEN)

Speakers:
William Dawson (RIKEN) "Large scale matrix polynomial computation for linear scaling quantum chemistry"
Edoardo Di Napoli (JSC) "High-Performance generation of Hamiltonian and overlap matrices in DFT methods based on linearized and augmented plane waves"
Eric Mikida (UICU) "The OpenAtom project and high performance GW software for excited-state computations"
Hiroshi Ueda (RIKEN) "Wavefunction predictions based on tensor network algorithms in quantum spin systems"

An atomic- and molecular-level understanding of the origin of properties and the mechanism of chemical reactions in materials will provide insight for developing new materials. Although a number of diverse experimental methods have been developed, it still remains difficult to investigate the mechanism of the chemical reaction and the origin of the functionality of complicated molecules and materials in details. Therefore, computational simulations that can predict the properties and functions of materials at the atomic and molecular levels is keenly awaited as a replacement for experiment. The scope of this session will cover novel computational approaches to quantum chemistry and condensed matter physics and their applications to molecules and materials, particularly on the topics:
1) New theoretical development with interface between chemistry and physics
2) Novel development of program and algorithm for high performance computing
3) Their material applications to solve grand challenges, for example,
alternative energy resources with solar cells and artificial photosynthesis, and so on.

Session 9

9.1 Project Talk: **Optimization of Fault-Tolerance Strategies for Workflow Applications**
"Optimization of Fault-Tolerance Strategies for Workflow Applications"
Aurélien Cavelan (INRIA)
Checkpointing is the traditional fault-tolerance technique when it comes to resilience for large-scale platforms. Unfortunately, as platform scale increases, checkpoints must become more frequent to accommodate with the increasing Mean Time Between Failure (MTBF). As such, it is expected that checkpoint-recovery will become a major bottleneck for applications running on post-petascale platforms. In this paper, we focus on replication as a way of mitigating the checkpointing-recovery overhead. In particular, we investigate the impact of replication on the execution of a single task. A task can be checkpointed and/or replicated, so that if only one replica fails, no recovery is needed. Replication can be done at different application level. We study process-replication, where each process

can be replicated several times, and we adopt a passive approach: waiting for the application to either succeed or fail. Finally, we consider both fail-stop and silent errors. We derive closed-form formulas for the expected execution time and first-order approximations for the overhead and the optimal checkpoint interval.

9.2 Individual Talk: **Failure detection**
Yves Robert (INRIA)
This talk briefly describes current methods for failure detection and outlines their shortcomings. it also presents a new algorithm that overcomes some (but not all) problems. Finally, itaddresses open questions.

9.3 Individual Talk: **Identification and imapct of failure cascades**
Frederic Vivien (INRIA)
Most studies assume that failures are independent and not time correlated. The question we address in this work is: can we identify time-correlated failures in supercomputer traces? could the potential cascade of failures have an impact on our usage of fault tolerance mechanisms?

9.4 Project Talk: **New Techniques to Design Silent Data Corruption Detectors**
"Recent Work on Detecting SDC"
Leonardo Bautista (BSC)
We will present the results obtained in the last 6 months on this topic.

9.5 Individual Talk: **Runtime driven online estimation of memory vulnerability**
Luc Jaulmes (BSC)
Memory reliability is measured as a fault rate, i.e. a probability over a given amount of time. The missing link to know the fault probability of any data stored in memory is its storage duration. By analyzing memory access patterns of an application, we can thus determine the vulnerability of data stored in memory, and thus the optimal amount of redundancy to keep fault probabilities below an acecptable threshold at all times. While such a data vulnerability metric has been approached offline [Luo et al., DSN'14] and with an analytical model [Yu et al., SC'14], we estimate it online using performance counters on real hardware. This allows to dynamically get fault probabilities for memory storage, and opens the door to runtime optimizations. The open problem remains the right set of actuators to use for a runtime system, in order to adapt the strength of memory protection. Some leads are to have different ECC strengths, either through an adaptable ECC scheme whose amount of redundancy can be adjusted, or through different chips with the option of migrating data under different protection requirements. Another lead is to allow strong ECC at all times, instead tuning parameters that impact resilience in order to save power and time, such as reducing DRAM refresh rates, wherever we know the uniform redundancy is superfluous.

9.6 Individual Talk: **Fault Tolerance and Approximate Computing**
Osman Unsal (BSC)
There is a correlation between fault-tolerance and approximate computing. Everything that is critical for fault-tolerance - be it code, data structures, threads or tasks - is not conductive to approximation. On the other hand, anything that is relatively less critical for fault-tolerance can be approximated. This means that if parts of an application is annotated for realiability-criticality; the same annotations could be leveraged for approximation without the need to further analyze and annotate the application for approximation. In particular, we have considered tasks for approximation that were unmarked as reliability-critical either by programmer or runtime; initial results are not conclusive, needing further exploration.

9.7 Individual Talk: **Fault-tolerance for FPGAs**
Osman Unsal (BSC)
Similar to GPUs, the first generations of FPGAs did not offer any hardware-based fault tolerance. However, their potential deployment in mission critical and HPC environments have led FPGA system integrators to include ECC-support for on-chip Block RAM and SRAM memory structures in state-of-the-art FPGAs. However, most of the area of an FPGA is reserved for programmable logic rather than memory - unlike CPUs who dedicate a majority of the chip area for cache memory structures -. Since fault rate is proportional to area, any reliability proposal for FPGAs should therefore target programmable logic structures as well. However, protecting every latch with ECC or any other redundancy technique is not effective. We think that an end-to-end integrated reliability solution involving ABFT, runtime and hardware is neccessary to select the most vulnerable programmable logic structures of an FPGA and protect them. In the initial step, we created a fault injection mechanism which injects faults to any desired latch of an FPGA. Our initial results support our intuition that certain latches are much more reliability critical than others.

Session 10

10.1 Project Talk: **Scalability Enhancements to FMM for Molecular Dynamics Simulations "Scratching The Millisecond - The Std::Way"**

David Haensel (JSC)

Modern HPC resources owe their peak performance to an increased node-level core count. In order to keep up scalability, a low-overhead threading approach for the user software is required. In our case, an FMM serves as a Coulomb solver for molecular dynamics simulations with the goal of milliseconds execution time for a single time step. To avoid parallelization overhead, e.g. synchronization points or load imbalance, algorithm-aware strategies have to be applied. Such measures will improve performance, especially for a tasking approach with dependency resolving and work scheduling. Implementing those specific strategies in a scheduler and dependency resolver of a third party library could be quite challenging. Also, relying solely on universal dynamic scheduling implementations could affect performance unfavorably. The current C++ language standard (C++11) offers several robust features for parallel intranode programming. With the help of those standardized C++ features, we added a tasking layer to our FMM library. In this talk we want to present, which C++11 features are most suited for tasking and how we apply and tailor such schemes for our purposes. Finally we will show an in-depth performance analysis of our current model utilizing dynamic work-stealing.

10.2 Individual Talk: **Daino: A High-level AMR Framework on GPUs**

Mohamed Wahib (RIKEN)

Adaptive Mesh Refinement methods reduce computational requirements of problems by increasing resolution for only areas of interest. However, in practice, efficient AMR implementations are difficult considering that the mesh hierarchy management must be optimized for the underlying hardware. Architecture complexity of GPUs can render efficient AMR to be particularity challenging in GPU-accelerated supercomputers. This talk presents a compiler-based high-level framework that can automatically transform serial uniform mesh code annotated by the user into parallel adaptive mesh code optimized for GPU-accelerated supercomputers. We show experimental results on three production applications. The speedups of code generated by our framework are comparable to hand-written AMR code while achieving good and weak scaling up to 3,600 GPUs.

10.3 Individual Talk: **Portable Asynchronous Progress Model for MPI on Multi- and Many-Core Systems**

Min Si (ANL)

Casper provides efficient process-based asynchronous progress for MPI one-sided communication on multi- and many-core architectures by dedicating a small number of cores to the asynchronous progress engine. It is designed as an external library of MPI through the PMPI interface thus ensuring portability among various MPI implementations and platforms. Although this model has successfully improved the performance of large quantum chemistry application NWChem by up to 30, it still lacks the solution for several issues, including the general approach to deliver optimal performance in multi-phase applications, supporting other MPI communication models such as two-sided and collectives, and cooperation with other PMPI based libraries such as the MPI performance tools. In this talk, we first propose an efficient dynamic adaptation mechanism to address the issue in multi-phase applications. It allows Casper to dynamically study the performance characteristics of application internal phases and adapt the configuration of asynchronous progress for each phase efficiently. Then we shortly introduce the ongoing work of Casper to improve the dynamic adaptation and resolve the second issue – supporting two-sided and collective modes, by integrating with the PVAS and ULP concepts contributed by RIKEN researchers. Finally, we discuss the potential collaborations to work on the challenge in the PMPI tools cooperation, and to seek the chance to integrate Casper within other MPI-based parallel runtime systems such as XMP on MPI.

10.4 Individual Talk: **Efficient Composition of task-graph based Code**

Christian Perez (INRIA)

Composition of code is still an open issue in HPC applications. The problem is even more difficult as task based programming models seem unavoidable. Moreover, applications may need to be specialized with respect to the objective function (max. perf, power caping, etc). This talk is about showing the possibilities and the open challenge that a component base approach brings using some examples.

10.5 Individual Talk: **In C++ we trust – performance portability out of the box?**

Ivo Kabadshow (JSC)

The HPC landscape is very diverse and performance on a single platform may come at a high price of lost portability. In this talk we like to outline our attempt to reach high resource utilization without losing portability. We will present our current C++11 software layout for a fast multipole method (FMM). We will focus on intranode performance, especially vectorization and multithreading. We will also show how abstraction and template-metaprogramming helps us to maintain clean and readable code without impacting performance.

10.6 Individual Talk: **Feedback control for Autonomic High-Performance Computing**
Eric Rutten (INRIA)
" Due to the increasing complexity, scale and heterogeneity in computing systems and applications, including hardware, software, communications and networks, there are growing needs for runtime management of resources, in an automated self-adaptation to the variations due to data or the environment. Such feedback loops are the object of Autonomic Computing (AC) and can involve the use of Control Theory. We have first results, in joint work with J.F. Mèhaut, N. Zhou, G. Delaval, and B. Robu, illustrating the topic, concerning the topic of the dynamical management of the trade-off between parallel computation and synchronization. Higher parallelism can potentially bring speedup, but also higher synchronization costs around shared data, which can even outgrow the gain. Additionally threads locality on different cores may impact on program performance as well. We have studied the problem in the particular framework of Software Transactional Memory (STM), and proposed solutions to dynamically adapt degree of parallelism and thread mapping policy, in order to diminish program execution time. We propose to address the perspectives of adopting such feedback-loop based approaches, which are very large, and the topic remains very novel. We propose to generalize the approach, on a first level, to other platforms and synchronization mechanisms (OpenMP, Charm++), and further, to consider objectives of different nature than response performance, but also of energy consumption, or dependability for example. Relevance to JLESC relates to the topic ""Programming Languages and Runtimes"". Preliminary exchanges have begun with S. Kale (UIUC) who works on related topics of load balancing regulation."

10.7 Individual Talk: **An Overview of the IHK/McKernel Lightweight Multikernel OS for Kernel-Assisted Communication Progression**
Balazs Gerofi (RIKEN)
"The ""Lightweight Kernel (LWK) Assisted Communication Progression"" JLESC project's goal is to propose mechanisms to obtain asynchronous progression of communications running on an LWK without disturbing application thread scheduling. The project combines the IHK/McKernel (RIKEN, Tokyo) lightweight multi-kernel with the MadMPI+Pioman (INRIA, Bordeaux) communication library. IHK/McKernel is a multi-kernel approach running Linux and LWK(s) side-by-side on compute nodes with the primary aim of exploiting the lightweight kernel's ability to provide scalable and consistent execution environment for large-scale parallel applications, but to retain the full POSIX/Linux APIs at the same time. In this talk, we provide an architectural overview of the IHK/McKernel software stack and report the current status of the collaboration."

10.8 Individual Talk: **New Parallel Execution Model - User-Level Implementation**
Atsushi Hori (RIKEN)
New execution model implemented at user-level to map multiple processes in one virtual address space will be introduced

Session 11

11.1 Project Talk: **Enhancing Asynchronous Parallelism in OmpSs with Argobots**
"Improving Hybrid MPI/OmpSs applications using Argobots"
Jesus Labarta (BSC)
"This talk will present the latest work done on the ""Enhancing Asynchronous Parallelism in OmpSs with Argobots"" project, where we have implemented the OmpSs tasking model on top of the Argobots threading library. In this talk we will explain our latest experiments that show how the integration between the Argobots threading library and the MPICH MPI library can improve both programmability and performance of hybrid MPI and OmpSs applications. Moreover, we also propose and compare this close integration of Argobots and MPICH with an alternative method based on MPI call interception."

11.2 Individual Talk: **Exploring RDMA libraries for PGAS language**
Tetsuya Odajima (RIKEN)
Recently, there are many communication libraries. Especially, an MPI-3 is widely used in many of HPC applications. Similarly, some PGAS languages employ MPI-3 as communication layer. However, there are not enough evidence that MPI-3 is optimal for runtime of PGAS language. In this project, we explore RDMA library which has high performance and high usability for developer of languages. In this talk, we evaluate preliminary performance of these libraries to find and optimal library for PGAS runtime system.

11.3 Break Out session: **HPC challenge of LQCD and related**
Char: Yoshifumi Nakamura (RIKEN)

Speakers:

Stefan Krieg (JSC) "QCD software development at JSC"
Hiroya Suno (RIKEN)"Algorithm and code development in lattice QCD using the K computer"
Giorgio Silvi (JSC)"Qlua optimization for many core architectures"
Yoshifumi Nakamura (RIKEN)"Towards high performance Lattice QCD simulations on Exascale computers"

We propose a break-out session for elementary particle physics applications with a focus on Lattice QCD. Lattice QCD (LQCD) is an approach to solving the quantum chromodynamics (QCD) which describs interaction for quark and gluon, is basic theorem of hadron physics, and application to understand particle collisions and state of matter in extreme environments like early universe, neutron stars and so on. LQCD is defined on 4 dimensional (space and time) lattices. The domain decomposition is used with MPI and Openmp. The hot spot is at solving a Dirac equation by using Krylov subspace methods. Theoretical required B/F in matrix vector multiplication is about 2 in double precision. When program is highly optimized on L2 cache, costs for global reduction and boundary exchange between processes become dominant. So this application is one of most challenging applications on extreme scale computers The aim of the breakout session is to reveal issues of elementary particle physics applications, mainly LQCD, on extreme scale computers and to explore the possibility of new collaboration to tackle them.

Session 12

12.1 Individual Talk: **Leveraging FPGA technology for next-generation high-performance computing systems**
Kazutomo Yoshii (ANL)
The performance progress of microprocessors has been driven by Moore's law, doubling the number of transistors every 18 to 24 months. Such fabulous scaling will end soon. In the post-Moore era, the energy efficiency of computing will be a major concern. FPGA technology could be a key to maximizing the energy efficiency as well as performance. We present a summary of the "Re-form" project and describe major challenges in the adoption of FPGA in high-performance computing.

12.2 Individual Talk: **Lessons learned from HPRC**
Volodymyr Kindratenko (UIUC)
As far as 2008, High-Performance Reconfigurable Computing (HPRC) has been shown to "achieve up to four orders of magnitude improvement in performance, up to three orders of magnitude reduction in power consumption, and two orders of magnitude saving in cost and size requirements compared with contemporary microprocessors when running compute-intensive applications based on integer arithmetic. " (IEEE Computer, February 2008). Yet, HPRC continues to remain an obscure technology that has yet to realize its promise in the field of HPC. The most significant obstacle in realizing its potential is due to the lack of a programming model capable of extracting performance from the reconfigurable hardware that is simple enough to be accepted by the scientific computing community without the highly specialized knowledge of hardware design principles. While recent work on OpenCL based design flow is an attempt to address this challenge, in my opinion, it fails short of both being easy to use by the scientific computing community and being capable of extracting full FPGA performance.

12.3 Break Out session: **Memory Errors at Extreme Scale**
Char: Leonardo Bautista Gomez (BSC)

Speakers:
Leonardo Bautista-Gomez (BSC) "Introduction and objectives of the session"
Prasanna Balaprakash (ANL) "Statistical analysis and predictive modeling for memory errors atexascale"
Osman Unsal (BSC) "Methodologies for Memory Error Logging"

Correlation analysis of memory errors across multiple JLESC machines.