

The International Conference for High Performance Computing,
Networking, Storage and Analysis



SC17

Denver, CO | hpc
connects.

Conference
Program

Sponsors:



sighpc



computer
society



TCHPC



Colorado Convention Center

Conference Dates: November 12-17, 2017

Exhibition Dates: November 13-16, 2017

Michael B. Hancock
Mayor



City and County of Denver

OFFICE OF THE MAYOR
CITY AND COUNTY BUILDING
DENVER, CO 80202-5390
TELEPHONE: (720) 865-9090 • FAX: (720) 865-8787
TTY/ TTD: (720) 865-9010

November 12, 2017

"SC17"
Colorado Convention Center
700 14th Street
Denver, CO 80202



Greetings:

Welcome to the "Mile High City". We are excited to welcome Supercomputing back to Denver for what is sure to be another thought provoking and informative conference.

Technological advancements continuously drive the way the public and private sectors operate, and "SC17" will once again bring together many of the world's brightest, most innovative minds to explore the latest in supercomputing and networking. And as Denver and others work to create "smart cities," your work will help to elevate the level of service we can provide to residents and visitors alike. Innovative high-performance computing will be a necessity to advance Denver and cities across the world as better and smarter places to live and work.

While you're here, I invite you to take the time to enjoy the wonderful attractions our city has to offer, including many activities and venues that provide world class entertainment, dining and shopping options. Downtown Denver is one of the nation's most walkable, and boasts a variety of attractions including the nation's second largest performing arts complex, three art museums, three sports stadiums, a U.S. Mint and more than 300 restaurants.

Enjoy your stay in Denver, and we look forward to seeing you all again in 2019!

Respectfully,

A handwritten signature in black ink, appearing to read "M. Hancock", followed by a long horizontal line.

Michael B. Hancock
Mayor

General Information

Registration and Conference Store

The registration area and conference store are located in Lobby F. They will be open during the following days and hours:

Day	Date	Hours
Saturday	November 11	1pm – 6pm
Sunday	November 12	7am – 6pm
Monday	November 13	7am – 9pm
Tuesday	November 14	7:30am – 6pm
Wednesday	November 15	7:30am – 6pm
Thursday	November 16	7:30am – 5pm
Friday	November 17	8am – 11am

Exhibition Hall Hours

Day	Date	Hours
Tuesday	November 14	10am-6pm
Wednesday	November 15	10am-6pm
Thursday	November 16	10am-3pm

SC17 Information Booths

Need up-to-the-minute information about what's happening at the conference. Need to know where to find your next session? What restaurants are close by? Where to get a document printed? These questions and more can be answered by a quick stop at one of the SC information booths. There are two booth locations for your convenience. The Main Booth is located in Lobby F near the conference store and registration. The Satellite Booth is located in the D concourse near the 600 series rooms.

	Main Booth	Satellite Booth
Saturday	1pm-5pm	Closed
Sunday	8am-6pm	8am-5pm
Monday	8am-7pm	8am-5pm
Tuesday	8am-6pm	10am-6pm
Wednesday	8am-6pm	8am-4pm
Thursday	8am-6pm	Closed
Friday	8:30am-12pm	Closed

SC18 Preview Booth

Members of next year's SC committee will be available in the SC18 preview booth (located in the A Atrium) to offer information and discuss next year's SC conference in Dallas, Texas. Stop by for a copy of next year's Call for Participation and pick up some free gifts! The booth will be open during the following hours:

Day	Date	Hours
Tuesday	November 14	10am-5pm
Wednesday	November 15	10am-5pm
Thursday	November 16	10am-3pm

Parenting Room

SC17 has provided a room with privacy areas for parents attending to infants, located in Room XXX. The room also features a refrigerator with lock boxes, changing tables, lockers, and a diaper disposal. The room will be open during the following dates and hours:

Day	Date	Hours
Saturday	November 11	noon – 5pm
Sunday	November 12	8am – 6pm
Monday	November 13	8am – 9pm
Tuesday	November 14	8am – 9pm
Wednesday	November 15	8am – 7pm
Thursday	November 16	8am – 6pm
Friday	November 17	8am – 1pm

Child Care

On-site child care services in the convention center is being provided by KiddieCorp for children between ages 6 months through 12 years old. The deadline for sign-up was Friday, October 21, 2017. All children using the SC17 childcare services need to have been registered in advance by a parent or legal guardian. There may be limited availability for drop-ins depending on the day, time and age. Contact KiddieCorp at 858-455-1718.

Day	Date	Hours
Sunday	November 12	8am – 6pm
Monday	November 13	8am – 9pm
Tuesday	November 14	8am – 7pm
Wednesday	November 15	8am – 7pm
Thursday	November 16	8am – 6pm

Coat and Bag Check

The coat and bag check station is located in Lobby A. The hours are as follows:

Day	Date	Hours
Saturday	November 11	noon – 8pm
Sunday	November 12	7am – 6pm
Monday	November 13	7am – 10pm
Tuesday	November 14	7am – 9pm
Wednesday	November 15	7am – 8pm
Thursday	November 16	7am – 6pm
Friday	November 17	7am – 1pm

Multi-faith Prayer and Quiet Room

A quiet room is a physical space where conversation and interaction are not allowed, where attendees can go for quiet, prayer, meditation, or similar activities. The Quiet Room is located in Mezzanine A, accessed from the elevators on the South-East corner of Lobby A.

Day	Date	Hours
Sunday	November 12	8am-6pm
Monday	November 13	8am-9pm
Tuesday	November 14	8am-7pm
Wednesday	November 15	8am-7pm
Thursday	November 16	8am-6pm
Friday	November 17	8am-Noon

Social Events

Exhibitors' Reception

Sunday, November 12

6pm-9pm

McNichols Civic Center
144 West Colfax Avenue

SC17 will host an Exhibitor Party at the McNichols Civic Center for registered exhibitors. The party is SC17's way of thanking exhibitors for their participation and support of the conference. The event will include entertainment along with food and drinks.

The McNichols Civic Center Building is located in the heart of downtown Denver at the northwest corner of Civic Center Park. In 1909, the cornerstone of the building was laid, setting the foundation for the then Carnegie Library that would become a center of learning in the park. That tradition continues, as the building was re-opened in 2012 as a contemporary hub for arts, culture and events for the people of Denver. This stunning Greek Revival building with its classic Corinthian columns and iconic colonnade across its front, offers new experiences in a classic space.

The McNichols Civic Center Building is within walking distance (@ .6 mile) from the Colorado Convention Center. Limited busing will be provided to/from the Convention Center. An Exhibitor badge is required to attend this event. Guest tickets may be purchased in advance at the Registration desk (no tickets will be available at the party).

Exhibits Gala Opening Reception

Monday, November 13

7pm-9pm

SC17 will host its annual Grand Opening Gala in the Exhibit Hall. This will be your first opportunity to see the latest high performance computing, networking, storage, analysis, and research products, services, and innovations. This event is open to all Technical Program, Exhibitors and Students@SC registrations.

Posters Reception

Tuesday, November 14

5:15pm-7pm

Four Seasons Ballroom

The Posters Reception is an opportunity for attendees to interact with poster presenters and includes research and ACM Student Research Competition posters, Doctoral Showcase posters as well as the Scientific Visualization & Data Analytics Showcase.

Technical Program Conference Reception

Thursday, November 16

6pm-9pm

EXDO Event Center

1399 35th St.

Denver, CO 80205

To thank our Technical Program attendees and to encourage continued global and inter-galactic interactions, SC17 is hosting a conference reception for all Technical Program attendees at the [EXDO Event Center](#), Denver's most versatile event space in the heart of the city's vibrant River North Art District.

This year's event features a sci-fi theme, with a virtual reality lounge, a green-screen photo booth, an Emulator Screen DJ, a live painting demonstration, and much more.

Even the food will be out of this world - with a "Mad Scientist" salad station, "Fire-in-the-Sky" flame wall, "Close Encounters of the Third" kind pasta station, "Lost-in-the-Pacific" station and of course "Alien Pod" s'mores. Guests are encouraged to come dressed as their favorite sci-fi character! So plan ahead.

A Tech Program badge, event ticket, and government-issued photo ID are required to attend this event. Busing will be provided to / from the Convention Center from the F Lobby from 5:30pm until 9:00pm. Note, you will need your badge and your event ticket to get on the shuttle.

Family Day

Family Day is Wednesday, November 15, 4pm-6pm. Adults and accompanied children are permitted on the floor during these hours when accompanied by a registered conference attendee.

Conference Services/ Convention Center Facilities

ATMs

ATMS are located in Lobby F, B and D.

Business Center

Last minute presentation preparations? Need copies or shipping service? The UPS Store located in the main concourse near Lobby F offers a variety of services tailored to help you make the best of event at the Colorado Convention Center. Wheelchair and scooter rentals are also available.

City and Dining Information

In the main lobby of the Colorado Convention Center across from the Blue Bear, VISIT DENVER offers a complimentary staffed city information and restaurant reservation service. Information specialists can assist with providing information on attractions, dining, transportation options, maps, brochures and a myriad of other helpful hints for navigating Denver.

Day	Date	Hours
Sunday	November 12	8am-6pm
Monday	November 13	8am-6pm
Tuesday	November 14	8am-6pm
Wednesday	November 15	8am-6pm
Thursday	November 16	8:30am-5pm
Friday	November 17	8am-Noon

SC17

Registration Pass Access

Each registration category provides access to a different set of conference activities listed below

	Technical Program	Technical Program + Workshop	Workshop Only	Tutorials	Exhibitor 24-hour Access	Exhibit Hall Only Tue-Thur
HPC Matters Plenary (Monday)	*	*	*	*	*	*
Keynote (Tuesday)	*	*	*	*	*	*
Panels (Friday Only)	*	*	*	*	*	*
Awards (Thursday)	*	*			*	*
Birds-of-a-Feather	*	*			*	*
Exhibit Floor	*	*			*	*
Exhibitor Forum	*	*			*	*
HPC Impact Showcase	*	*			*	*
Student Cluster Competition	*	*			*	*
Invited Talks (Plenary)	*	*			*	
Posters	*	*			*	
Exhibits Gala Opening (Monday)	*	*			*	
EXHIBITOR / TECH BADGE						
Exhibitors Reception (Sunday)					*	
TICKET REQUIRED						
Doctoral Showcase	*	*				
Invited Talks (Non-Plenary)	*	*				
Panels (Tue-Thur)	*	*				
Papers	*	*				
Poster Reception (Tuesday)	*	*				
Scientific Visualization Showcase	*	*				
Technical Program Reception (Thursday) TICKET REQUIRED	*	*				
Awards Luncheon (Thursday)						
TICKET REQUIRED						
Tutorial Lunch (Sun/Mon)				*		
TICKET REQUIRED						
Tutorial Sessions				*		
Workshops		*	*			

Table of Contents

ACM Gordon Bell Finalist
ACM Student Research Competition
Awards Presentation
Birds of a Feather
Doctoral Showcase
Early Career Program
Emerging Technologies
Exhibitor Forum
Exhibits
HPC Impact Showcase
HPC Matters Plenary
Inclusivity
Invited Talk

Keynote & Plenary Talk
Navigating SC17
Panel
Paper
Poster
Reception
Scientific Visualization & Data Analytics Showcase
Student Job/Opportunity Fair
Students@SC
Tech Program Breaks
Test of Time
Tutorial
Workshop

ACM Gordon Bell Finalist [\(back to top\)](#)

Thursday, November 16th

Room: 301-302-303
10:30 am - 12:00 pm

ACM Gordon Bell Finalists

Redesigning CAM-SE for Petascale Climate Modeling Performance on Sunway TaihuLight

Lin Gan (Tsinghua University)

We refactor and optimize the entire Community Atmosphere Model (CAM) to the full system of the Sunway TaihuLight, and provide a petascale climate modeling performance. We scale the CAM to 1.5 million cores with a simulation speed of 2.81 simulated years per day using OpenACC directives at the first stage. We then apply a more aggressive and challenging finer-grained redesign of the HOMME dynamical core, to achieve finer memory control, more efficient vectorization and overlap between computation and communication. Besides, a register communication based parallelism scheme is proposed to minimize the data dependencies in the modules. By doing so, our optimized kernels running on a 260-core Sunway processor outperform the established HOMME kernels on a platform with up to 184 Intel Xeon E5-2680V3 CPU cores. And our implementation has achieved a sustainable double-precision performance around 2.5 Pflops for a 0.75 km global simulation when using 8,519,680 cores.

15-Pflops Nonlinear Earthquake Simulation on Sunway TaihuLight: Enabling Depiction of Realistic 10 Hz Scenarios

Haohuan Fu (Tsinghua University)

This paper reports our work on building a highly efficient earthquake simulation platform on Sunway TaihuLight, with 125 Pflops computing power and over 10 million cores. With the platform originated from AWP-ODC and CG-FDM, a large part of our efforts focuses on redesigning the velocity, stress, and plasticity processing kernels for the completely different microarchitecture and significantly increased parallelism of Sunway TaihuLight. By a combined approach including (1) an optimized parallelization scheme, (2) the most suitable blocking configuration, (3) fusion of co-located arrays, (4) register communication with CPE ID remapping for halo exchanges, and (5) customized ROM-less evaluation of elementary function, we manage to achieve an efficient utilization of over 12.2% of the theoretical peak of the entire system. Our program provides a sustained performance of over 15 Pflops and enables the simulation of the Tangshan earthquake with a spatial resolution of 25 m and a frequency of 10 Hz.

Massively Parallel 3D Image Reconstruction

Xiao Wang (Purdue University)

Computed Tomography (CT) image reconstruction is an important technique used in a wide range of applications. Among reconstruction methods, Model-Based Iterative Reconstruction (MBIR) generally produces higher quality images. However, the irregular data access pattern, the difficulty of effective parallelization and slow algorithmic convergence have made MBIR impractical for many applications. This paper presents a new algorithm for MBIR, Non-Uniform Parallel Super-Voxel (NU-PSV), that regularizes the data access pattern, enables massive parallelism and ensures fast convergence. We compare the NU-PSV algorithm with two state-of-the-art implementations on a 69632-core distributed system. Results indicate that the NU-PSV algorithm has an average speedup of 1665 compared to the fastest state-of-the-art implementations.

ACM Student Research Competition [\(back to top\)](#)

Tuesday, November 14th

Room: Four Seasons Ballroom

8:30 am - 5:00 pm

Research Posters

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

8:30 am - 5:00 pm

ACM Student Research Competition

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

Room: Four Seasons Ballroom

5:15 pm - 7:00 pm

Poster Reception

A01: GEMM-Like Tensor-Tensor Contraction (GETT)

Authors: *Paul Springer (RWTH Aachen University), Paolo Bientinesi (RWTH Aachen University)*

Tensor contractions (TC) are a performance critical component in numerous scientific computations. Despite the close connection between matrix-matrix products (GEMM) and TCs, the performance of the latter is in general vastly inferior to that of an optimized GEMM. To close such a gap, we propose a novel approach: GEMM-like Tensor-Tensor multiplication (GETT). GETT mimics the design of a high-performance GEMM implementation; as such, it systematically reduces an arbitrary tensor contractions to a highly-optimized "macro-kernel". This macro-kernel operates on suitably "packed" sub-tensors that reside in specified levels of the cache hierarchy. GETT's decisive feature is its ability to pack subtensors via tensor transpositions, yielding efficient packing routines. In contrast to previous approaches to TCs, GETT attains the same I/O cost as an equally-sized GEMM, making GETT especially well-suited for bandwidth-bound TCs. GETT's excellent performance is highlighted across a wide range of random tensor contractions.

SF

A02: Accelerating the Higher Order Singular Value Decomposition Algorithm for Big Data with GPUs

Authors: *Yuhsiang M. Tsai (National Taiwan University)*

With the explosion of big data, finding ways of compressing large datasets with multi-way relationship - i.e., tensors - quickly and efficiently has become critical in HPC.

High-order singular value decomposition (HOSVD) method provides us with the means to attain both extremely high compression ratio and low error rate through low-rank approximation.

However, parallelizing HOSVD efficiently on GPUs remains a challenging problem, largely due to the lack of a fast SVD implementation that can stream data to the limited GPU memory through the PCIe bottleneck.

Our work studies, optimizes, and then contrasts four different methods for calculating singular vectors for performance, weak/strong scalability and accuracy in the context of HOSVD. We also discuss ways of load balancing the problem across multiple GPUs on a single node, and discuss the pros and cons of these different algorithms for GPU acceleration.

SF

A03: A High-Speed Algorithm for Genome-Wide Association Studies on Multi-GPU Systems

Authors: *Yen Chen Chen (National Taiwan University)*

We develop an algorithm as long as a CUDA code for GWAS (Genome-Wide Associate Studies). This algorithm can work efficiently on GPU and has high scalability. The core of the algorithm is an accurate and fast p-value integration reformation, which accelerates the most time-consuming part of the algorithm. With the algorithm, researchers can now deal with tens of billions of SNP to trait pair in only a few minutes. Even better, since this algorithm is highly scalable, you can increase the problem size as long as you have enough computing power.

A04: Optimization of the AIREBO Many-Body Potential for KNL**Authors:** Markus Höhnerbach (RWTH Aachen University)

Molecular dynamics simulations are an indispensable research tool for computational chemistry and material science. Empirical many-body potentials promise high-fidelity simulations that capture bonding and reaction behavior accurately, providing a level of detail in between more classical molecular dynamics and quantum methods.

The AIREBO potential is one such example that provides forces and energies for molecular dynamics (MD) simulations of carbon and carbohydrate structures. Allowing many-body potentials to profit from the recent architectural advances still poses a challenge due to deeply nested, short loops. We develop an optimized, vectorized AIREBO implementation for Intel's Xeon and Xeon Phi (co)processors and integrate it into the open-source LAMMPS molecular dynamics code. By both introducing improvements to the code and vectorization, we achieve a sustained real-world speedup of two on Broadwell, and a speedup of four on KNL. The optimized code will be distributed with each LAMMPS download as part of the USER-INTEL package.

SF

A05: Parallel Prefix Algorithms for the Registration of Arbitrarily Long Electron Micrograph Series**Authors:** Marcin Copik (RWTH Aachen University)

Recent advances in the technology of transmission electron microscopy have allowed for a more precise visualization of materials and physical processes, such as metal oxidation. Nevertheless, the quality of information is limited by the damage caused by an electron beam, movement of the specimen or other environmental factors. A novel registration method has been proposed to remove those limitations by acquiring a series of low dose microscopy frames and performing a computational registration on them to understand and visualize the sample. This process can be represented as a prefix sum with a complex and computationally intensive binary operator and a parallelization is necessary to enable processing long series of microscopy images. With our parallelization scheme, the time of registration of results from ten seconds of microscopy acquisition has been decreased from almost thirteen hours to less than seven minutes on 512 Intel IvyBridge cores.

A06: Accelerating Big Data Processing in the Cloud with Scalable Communication and I/O Schemes**Authors:** Shashank Gugnani (Ohio State University)

With the advent of cloud computing, the field of Big Data has seen rapid growth. Most cloud providers provide hardware resources such as NVMe SSDs, large memory nodes, and SR-IOV. This opens up the possibility of large-scale high-performance data analytics and provides opportunities to use these resources to develop new designs. Cloud computing provides flexibility, scalability, and reliability, which are important requirements of Big Data frameworks. However, several important requirements are missing, such as performance, scalability, fault-tolerance, and consistency. The focus of this research work revolves around developing communication and I/O designs and concepts which can provide these requirements to Big Data frameworks. Specifically, we explore new ways to provide fault-tolerance and consistency in cloud storage systems, providing scalable and high-performance communication frameworks, and co-designing with Big Data stacks to leverage these features.

A07: Scalable Parallel Scripting in the Cloud**Authors:** Benjamin H. Glick (Lewis & Clark College)

It's often complicated, time consuming, but frequently necessary to successfully port complex workflows to multiple high-performance environments. Parsl is a Python-based parallel scripting library that provides a simple model for describing and executing dataflow-based scripts over arbitrary execution resources such as clouds, campus clusters, and high-performance systems. Parsl's execution layer abstracts the differences between providers enabling provisioning and management of compute nodes for use with a pilot system. In this poster, we describe the development of a new execution provider designed to support Amazon Web Services (AWS) and Microsoft's Azure. This provider supports the transparent execution of implicitly parallel Python-based scripts using elastic cloud resources. We demonstrate that Parsl is capable of executing thousands of applications per second over this elastic execution fabric.

A08: Virtualized Big Data: Reproducing Simulation Output on Demand**Authors:** Salvatore Di Girolamo (ETH Zurich)

Scientific simulations are being pushed to the extreme in terms of size and complexity of the addressed problems, producing astonishing amount of data. If the data is stored on disk, analysis applications can randomly access simulation output. Yet, storing the massive amounts simulation data is challenging. This is primarily due to the high storage costs and the fact that compute capabilities grow faster than storage capacities and bandwidths. In-situ analysis removes the storage costs but applications lose random access.

We propose to not store the full simulation output data but to produce it on demand. Our system intercepts I/O requests of both analysis tools and simulators, enabling data virtualization. This new paradigm allows us to explore the computation-storage tradeoff, by trading computation power for storage space. Overall, SDAVi offers a viable path towards exa-scale scientific simulations, by exploiting the growing computing power and relaxing the storage capacity requirements.

SF

A09: Ring: Unifying Replication and Erasure Coding to Rule Resilience in KV-Stores

Authors: *Konstantin Taranov (ETH Zurich)*

There is a wide range of storage schemes employed by KV-stores to ensure reliability of stored keys. However, previous implementations do not allow choosing the storage scheme dynamically, thereby forcing developers to commit to a single scheme. Such inefficient data management wastes cluster resources such as memory usage, network load, latency, availability and many others.

To solve this problem, we have designed a strongly consistent key-value store Ring that empowers its users to explicitly manage storage parameters like other resources, such as memory or processor time. The key feature of Ring is that all keys live in the same strongly consistent namespace and a user does not need to specify the resilience when looking up a key or value.

Our poster demonstrates how future applications that manage resilience of key-value pairs consciously can reduce the overall operational cost and improve performance significantly.

A10: Revealing the Power of Neural Networks to Capture Accurate Job Resource Usage from Unparsed Job Scripts and Application Inputs

Authors: *Michael R. Wyatt (University of Delaware)*

Next generation HPC schedulers will rely heavily on accurate information about resource usage of submitted jobs. The information provided by users is often inaccurate and previous prediction models, which rely on parsed job script features, fail to accurately predict for all HPC jobs. We propose a new representation of job scripts and inclusion of application input decks for resource usage predictions with a neural network. Our contributions are a method for representing job scripts as image-like data, an automated method for predicting job resource usage from job script images and input deck features, and validation of our methods with real HPC data. We demonstrate that when job scripts for an application are very similar, our method performs better than other methods. We observe an average decrease in error of 2 node-hours compared to state of the art methods.

A11: Finding a Needle in a Field of Haystacks: Lightweight Metadata Search for Large-Scale Distributed Research Repositories

Authors: *Anna Blue Keleher (University of Maryland)*

Fast, scalable, and distributed search services are commonly available for single nodes, but lead to high infrastructure costs when scaled across tens of thousands of filesystems and repositories, as is the case with Globus. Endpoint-specific indexes may instead be stored on their respective nodes, but while this distributes storage costs between users, it also creates significant query overhead. Our solution provides a compromise by introducing two levels of indexes: a single centralized "second-level index" (SLI) that aggregates and summarizes terms from each endpoint; and many endpoint-level indexes that are referenced by the SLI and used only when needed. We show, via experiments on Globus-accessible filesystems, that the SLI reduces the amount of space needed on central servers by over 96% while also reducing the set of endpoints that need to execute user queries.

SF

A12: Applying Image Feature Extraction to Cluttered Scientific Repositories

Authors: *Emily Herron (Mercer University)*

Over time many scientific repositories and file systems become disorganized, containing poorly described and error-ridden data. As a result, it is often difficult for researchers to discover crucial data. In this poster, we present a collection of image processing modules that collectively extract metadata from a variety of image formats. We implement these modules in Skluma—a system designed to automatically extract metadata from structured and semi-structured scientific formats. Our modules apply several image metadata extraction techniques that include processing file system metadata, header information, color content statistics, extracted text, feature-based clusters, and predicting tags using a supervised learning model. Our goal is to collect a large number of metadata that may then be used to organize, understand, and analyze data stored in a repository.

SF

A13: Deep Learning with HPC Simulations for Extracting Hidden Signals: Detecting Gravitational Waves

Authors: *Daniel George (National Center for Supercomputing Applications, University of Illinois)*

We introduce Deep Filtering, new machine learning method for end-to-end time-series signal processing, which combines two deep one-dimensional convolutional neural networks for classification and regression to detect and characterize signals much weaker than the background noise. We trained this method with a novel curriculum learning scheme on data derived from HPC simulations and applied it for gravitational wave analysis specifically for mergers of black holes and demonstrated that it significantly outperforms conventional machine learning techniques, is far more efficient than matched-filtering, offering several orders-of-magnitude speed-up, allowing real-time processing of raw big data with minimal resources, and extends the range of detectable signals. This initiates a new paradigm for scientific research which employs massively-parallel numerical simulations to train artificial intelligence algorithms that exploit emerging hardware architectures such as deep-learning-optimized GPUs. Our approach offers a unique framework to enable coincident detection campaigns of gravitational wave sources and their electromagnetic counterparts.

SF

A14: Analysis of Synthetic Graph Generation Methods for Directed Network Graphs

Authors: *Spencer Callicott (Mississippi State University)*

Historically, scientific experiments have been conducted to generate scale-free network graphs based on structure. Metrics used to measure veracity ensure the integrity of a scale-free algorithm given a seed. However, studies do not explore the performance benefits or drawbacks of specific algorithms running on Apache Spark and GraphX. Recognizing the lack of performance benchmarks demands ensuring accuracy through experimenting. This study will utilize the Stochastic Kronecker Graph model to synthetically generate graphs given a seed graph.

A15: Quantifying Compiler Effects on Code Performance and Reproducibility Using FLiT

Authors: *Michael Bentley (University of Utah)*

A busy application developer likes to focus on doing science, but instead is often distracted by the sheer variety of available hardware platforms, their compilers, and associated optimization flags. Exclusive pursuit of speed may jeopardize the reproducibility of scientific experiments. On the other hand, performance is central to HPC. Our previous work provided a unique testing framework called FLiT that helps developers exploit performance without jeopardizing reproducibility. To verify that FLiT is useful for real-world libraries and applications, it was applied to MFEM, a finite element library used in various HPC applications. I show that the compilation with the fastest average runtime for the converted MFEM examples is also bitwise reproducible. For these examples, clang had the fastest average runtimes and the best reproducibility. Our future work aims to enhance the open-source FLiT tool into a strong community resource and to follow up with found compiler oddities.

A16: Diagnosing Parallel I/O Bottlenecks in HPC Applications

Authors: *Peter Z. Harrington (University of California, Santa Cruz)*

HPC applications are generating increasingly large volumes of data (up to hundreds of TBs), which need to be stored in parallel to be scalable. Parallel I/O is a significant bottleneck in HPC applications, and is especially challenging in Adaptive Mesh Refinement (AMR) applications because the structure of output files changes dynamically during runtime. Data-intensive AMR applications run on the Cori supercomputer show variable and often poor I/O performance, but diagnosing the root cause remains challenging. Here we analyze logs from multiple levels of Cori's parallel I/O subsystems, and find bottlenecks during file metadata operations and during the writing of file contents that reduced I/O bandwidth by up to 40x. Such bottlenecks seemed to be system-dependent and not the application's fault. Increasing the granularity of file-system performance data will help provide conclusive causal relationships between file-system servers and metadata bottlenecks.

SF

A17: Toward Capturing Nondeterminism Motifs in HPC Applications

Authors: *Dylan Chapp (University of Delaware)*

High performance MPI applications employ nondeterministic asynchronous communication to achieve greater performance. However, this nondeterminism can significantly hamper debugging. Various software tools have been developed to control nondeterminism in HPC applications, but a high-level application-agnostic taxonomy for this nondeterminism is absent and limits these tools' effectiveness in practice. We propose to address this need by extracting common nondeterministic communication motifs from representative applications.

We present a first step toward capturing nondeterminism motifs by way of a workflow for detecting and summarizing sender nondeterminism in HPC applications.

A18: Understanding the Impact of Fat-Tree Network Locality on Application Performance

Authors: *Philip Taffet (Rice University)*

Network congestion can be a significant cause of performance loss and variability for many message passing programs. However, few studies have used a controlled environment with virtually no other extraneous sources of network traffic to observe the impact of application placement and multi-job interactions on overall performance. We study different placements and pairings for three DOE applications. We observe that for a job size typical for an LLNL commodity cluster, the impact of congestion and poor placement is typically less than 2%, which is less dramatic than on torus networks. In addition, in most cases, the cyclic MPI task mapping strategy increases performance and reduces placement sensitivity despite also increasing total network traffic. We also found that the performance difference between controlled placements and runs scheduled through the batch system was less than 3%.

SF

A19: Performance Analysis of a Parallelized Restricted Boltzmann Machine Artificial Neural Network Using OpenACC Framework and TAU Profiling System

Authors: *Abhishek Kumar (Brookhaven National Laboratory)*

Restricted Boltzmann Machines are stochastic neural networks that create probability distribution based off connection weight between nodes of the hidden and visible layer. The distribution makes the program optimal at classifying large amounts of data, which could be useful in work settings, such as a research lab. The parallelization of these neural networks would allow for the classification of data at a much faster rate than before. Using a high-performance computer it was determined that parallelizing the neural networks could decrease the runtime of the algorithm by over 35% when offloading the work to a GPU through OpenACC.

Using Tuning and Analysis Utilities Profiling Systems, it was found that scheduling the program would only be effective if the data size was large enough and an increase in the number of thread blocks used for scheduling would allow for greater performance gains than the number of threads in each thread block.

A20: Correctness Verification and Boundary Conditions for Chapel Iterator-Based Loop Optimization

Authors: Daniel A. Feshbach (*Haverford College*)

We explore two issues of correctness concerning iteration space transformation techniques: data dependencies and boundary conditions. First, we present a data structure which automatically verifies correctness of data dependencies for stencil computations with transformed iteration spaces. This further confirms the viability of Chapel iterators for defining iteration space transformations, by demonstrating that simple tool support can verify data dependencies and assist debugging. Second, we explore the performance and simplicity of three strategies for implementing boundary conditions in transformed iteration spaces: if statements, loop peeling, and an array of coefficients. We find that the coefficient array technique performs the best, often at 70 to 80 percent speed of the benchmark of ignoring the boundary condition. If statements are not far behind, while loop peeling performs much worse. The coefficient array and if statements are indifferent to the transformation technique applied, while loop peeling must be implemented within the transformation.

A21: Runtime Support for Concurrent Execution of Overdecomposed Heterogeneous Tasks

Authors: Jaemin Choi (*University of Illinois*)

With the rise of heterogeneous systems in high performance computing, how we utilize accelerators has become a critical factor in achieving the optimal performance. We explore several issues with using accelerators in Charm++, a parallel programming model that employs overdecomposition. We propose a runtime support scheme that enables concurrent execution of heterogeneous tasks and evaluate its performance. Using a synthetic benchmark that utilizes busy-waiting to simulate workload, we observe that the effectiveness of the runtime support varies with the application characteristics, with a maximum speedup of 4.79x. With a two-dimensional five-point stencil benchmark designed to represent a realistic workload, we obtain up to 2.75x speedup.

A22: Verifying Functional Equivalence Between C and Fortran Programs

Authors: Wenhao Wu (*University of Delaware*)

Software verification is a mature research area with many techniques. These verification approaches can be applied to programs written in different programming languages; nevertheless, most verification tools are only designed for programs written in C or Java. As a result, verification tools are inadequate for other languages, such as Fortran. A high level of software safety is mandatory in most of its application scenarios, which makes verification tools for Fortran programs necessary and significant.

In this poster, the author illustrates the motivation and objectives of the project with examples. Also, this poster shows an extension (as a Fortran program verifier) of an existing verification platform -- CIVL. Additionally, the results of a set of extensive experiments conducted by the author is shown in this poster to indicate that the performance is satisfactory.

A23: Evaluation of Data-Intensive Applications on Intel Knights Landing Cluster

Authors: Tao Gao (*University of Delaware*)

Analyzing and understanding large datasets on high performance computing platforms is becoming more and more important in various scientific domains. MapReduce is the dominant programming model for processing these datasets. Platforms for data processing are empowered by many-core nodes with cutting-edge processing units. Intel Knights Landing (KNL) is the new arrival in the field. However, this new architecture has not been fully evaluated for data-intensive applications. In this poster, we present the assess of KNL on the performance of three key data-intensive applications based on a high-performance MapReduce programming framework on the latest KNL-cluster, Stampede2. We focus on the impact of different KNL memory models, we compare Stampede2 with other clusters such as Tianhe-2 and Mira, and we measure the scalability of large datasets. We observe how KNL-based clusters are a promising architecture for data-intensive applications. We also identify key aspects to enable more efficient usage of KNL-based clusters.

A24: Comparison of Machine Learning Algorithms and Their Ensembles for Botnet Detection

Authors: Songhui Ryu (*Purdue University*)

A Botnet is a network of compromised devices that is controlled by malicious 'botmaster' in order to perform various tasks, such as executing DoS attack, sending SPAM and obtaining personal data etc. As botmasters generate network traffic while communicating with their bots, analyzing network traffic to detect Botnet traffic can be a promising feature of Intrusion Detection System(IDS). Although IDS has been applying various machine learning (ML) techniques, comparison of ML algorithms including their ensembles on Botnet detection has not been figured out yet. In this study, not only the three most popular classification ML algorithms – Naïve Bayes, Decision tree, and Neural network are evaluated, but also the ensemble methods known to strengthen ML algorithms are tested to see if they indeed provide enhanced predictions on Botnet detection. This evaluation is conducted with CTU-13 public dataset, measuring running time of each ML and its f measure and MCC score.

A25: Investigating Performance of Serialization Methods for Networked Data Transfer in HPC Applications

Authors: Max Yang (Georgia Institute of Technology)

Cluster-to-user data transfers present challenges with cross-platform endianness (byte-order) compatibility and handling a variety of numeric types, and may occur over suboptimal network links. Two serialization libraries, Protocol Buffers and Conduit, were selected for their ability to handle endianness and their cross-language support, and their performance in both size and speed was measured. It was found that the throughput of Protocol Buffers was significantly more than that of Conduit while exhibiting less protocol overhead. Adding a compression stage after serialization dramatically reduced the size of messages on certain types of data, but had some impact on throughput.

A26: Co-Designing MPI Runtimes and Deep Learning Frameworks for Scalable Distributed Training on GPU Clusters

Authors: Ammar Ahmad Awan (Ohio State University)

Deep Learning frameworks like Caffe, TensorFlow, and CNTK have brought forward new requirements and challenges for communication runtimes like MVAPICH2-GDR. These include support for low-latency and high-bandwidth communication of very-large GPU-resident buffers. This support is essential to enable scalable distributed training of Deep Neural Networks on GPU clusters. However, current MPI runtimes have limited support for large-message GPU-based collectives. To address this, we propose the S-Caffe framework; a co-design of distributed training in Caffe and large-message collectives in MVAPICH2-GDR. We highlight two designs for MPI_Bcast, one that exploits NVIDIA NCCL and the other that exploits ring-based algorithms. Further, we present designs for MPI_Reduce that provide up to 2.5X improvement. We also present layer-wise gradient aggregation designs in S-Caffe that exploit overlap of computation and communication as well as the proposed reduce design. S-Caffe provides a scale-out to 160 GPUs for GoogLeNet training and delivers performance comparable to CNTK for AlexNet training.

A27: High-Performance and Scalable Broadcast Schemes for Deep Learning on GPU Clusters

Authors: Ching-Hsiang Chu (Ohio State University)

Broadcast operations are a widely used operation in many streaming and deep learning applications to disseminate large amounts of data on emerging heterogeneous High-Performance Computing (HPC) systems. Further, traditional broadcast schemes are not well optimized for upcoming large-scale Graphics Processing Unit (GPU)-based systems. However, utilizing cutting-edge features of modern HPC technologies such like InfiniBand (IB) and NVIDIA GPUs to enable scalable heterogeneous broadcast operations remains an open challenge.

Toward delivering the best performance for streaming and deep learning workloads, we propose high-performance and scalable broadcast schemes that exploit IB hardware multicast (IB-MCAST) and NVIDIA GPUDirect technology. We present experimental results and find that they indicate improved scalability and up to 68% reduction of latency compared to the state-of-the-art solutions in the benchmark-level evaluation. Furthermore, the proposed design yields up to 24% performance improvement for the popular deep learning framework, Microsoft cognitive toolkit (CNTK), with no application changes.

A28: Exploring Use Cases for Non-Volatile Memories in Support of HPC Resilience

Authors: Onkar Patil (North Carolina State University)

Improving resilience and creating resilient architectures is one of the major goals of exascale computing. With the advent of Non-volatile memory technologies, memory architectures with persistent memory regions will be a significant part of future architectures. There is potential to use them in more than one way to benefit different applications. We look to take advantage of this technology to enable more fine-grained and novel methodology that will improve resilience and efficiency of exascale applications. We have developed three modes of memory usage for persistent memory to enable efficient checkpointing in HPC applications. We have developed a simple API that is evaluated with the DGEMM benchmark on a 16-node cluster with independent SSDs on every node. Our aim is to build on this work and enable static and dynamic runtime systems that will inherently make the HPC applications more fault-tolerant and resistant to errors.

P01: Cache-Blocking Tiling of Large Stencil Codes at Runtime

Authors: Istvan Z. Reguly (Pazmany Peter Catholic University), Gihan R. Mudalige (University of Warwick), Mike B. Giles (University of Oxford)

Stencil codes on structured meshes are well-known to be bound by memory bandwidth. Previous research has shown that compiler techniques that reorder loop schedules to improve temporal locality across loop nests, such as tiling, work particularly well. However in large codes the scope of such analysis is limited by the large number of code paths, compilation units, and run-time parameters. We present how, through run-time analysis of data dependencies across stencil loops enables the OPS domain specific language to tile across a large number of different loops. This lets us tackle much larger applications than previously studied: we demonstrate 1.7-3.5x performance improvement on CloverLeaf 2D, CloverLeaf 3D, TeaLeaf and OpenSBLI, tiling across up to 650 subsequent loopnests accessing up to 30 different state variables per gridpoint with up to 46 different stencils. We also demonstrate excellent strong and weak scalability of our approach on up to 4608 Broadwell cores.

P02: Strassen's Algorithm for Tensor Contraction

Authors: Jianyu Huang (University of Texas), Devin A. Matthews (University of Texas), Robert A. van de Geijn (University of Texas)

Tensor contraction(TC) is an important computational kernel widely used in numerous applications. It is a multi-dimensional generalization of matrix multiplication(GEMM). While Strassen's algorithm for GEMM is well studied in theory and practice, extending it to accelerate TC has not been previously pursued. Thus, we believe this to be the first work to demonstrate how one can in practice speed up tensor contraction with Strassen's algorithm. By adopting a Block-Scatter-Matrix format, a novel matrix-centric tensor layout, we can conceptually view TC as GEMM for general stride storage, with an implicit tensor-to-matrix transformation. This insight enables us to tailor a recent state-of-the-art implementation of Strassen's algorithm to TC, avoiding explicit transpositions(permutations) and extra workspace, and reducing the overhead of memory movement that is incurred. Performance benefits are demonstrated with a performance model as well as in practice on modern single core, multicore, and distributed memory parallel architectures, achieving up to 1.3× speedup.

P03: BEM4I: A Massively Parallel Boundary Element Solver

Authors: Michal Merta (*Technical University of Ostrava*), Jan Zapletal (*Technical University of Ostrava*), Michal Kravcenko (*Technical University of Ostrava*), Lukas Maly (*Technical University of Ostrava*)

In this work we present a library of parallel solvers based on the boundary element method (BEM). We provide a brief description of BEM and its parallelization, focus on SIMD vectorization and shared- and distributed-memory parallelization by OpenMP and MPI, respectively. Two approaches for distributed parallelization of BEM are discussed - the first one based on a novel parallel adaptive cross approximation (ACA) method, the second one on the boundary element tearing and interconnecting (BETI) domain decomposition method. To demonstrate the efficiency of the library we provide results of numerical experiments on the Xeon and Xeon Phi based clusters.

BP

P04: Unstructured-Grid CFD Algorithms on Many-Core Architectures

Authors: Aaron Walden (*NASA Langley Research Center*), Eric J. Nielsen (*NASA Langley Research Center*), Mohammad Zubair (*Old Dominion University*), John C. Linford (*ParaTools*), John G. Wohlbiel (*Engility Corporation*), Justin P. Luitjens (*Nvidia Corporation*), Jason Orender (*Old Dominion University*), Izaak Beekman (*ParaTools*), Samuel Khuvis (*ParaTools*), Sameer S. Shende (*ParaTools*)

In the field of computational fluid dynamics (CFD), the Navier-Stokes equations are often solved using an unstructured-grid approach to accommodate geometric complexity. Furthermore, turbulent flows encountered in aerospace applications generally require highly anisotropic meshes, driving the need for implicit solution methodologies to efficiently solve the discrete equations. These approaches require frequent construction and solution of large, tightly-coupled systems of block-sparse linear equations.

We explore the transition of two representative CFD kernels from a coarse-grained MPI-based model originally developed for multi-core systems to a shared-memory model suitable for many-core platforms. Results for the Intel Xeon Phi Knights Landing, NVIDIA Pascal P100, and NVIDIA Volta V100 architectures are compared with the aforementioned MPI-based implementation for the multi-core Intel Xeon Broadwell (BWL) processor. We observe substantial speedups over BWL as well as higher performance per dollar MSRP and performance per watt for the many-core architectures.

P05: ooc_cuDNN : A Deep Learning Library Supporting CNNs over GPU Memory Capacity

Authors: Yuki Ito (*Tokyo Institute of Technology*), Ryo Matsumiya (*Tokyo Institute of Technology*), Toshio Endo (*Tokyo Institute of Technology*)

GPUs are widely used to accelerate deep learning with convolutional neural network (CNN). However, since GPU memory capacity is limited, it is difficult to implement efficient programs that compute large CNN on GPU. This poster describes the design and implementation of out-of-core cuDNN (ooc_cuDNN) library, which supports to compute CNN exceeding GPU memory capacity using capacity of CPU memory. ooc_cuDNN is an extension of cuDNN, which is high performance and popular deep learning library. ooc_cuDNN divides CNN computation based on its performance model for better performance. In addition, ooc_cuDNN provides fused functions combined some computation to reduce extra communication. With ooc_cuDNN, we successfully computed CNN requiring more than 60 GB memory on a single GPU with 16 GB memory. Compared with an in-core case using cuDNN, performance degradation was 13 %.

P06: Large Scale FFT-Based Stress-Strain Simulations with Irregular Domain Decomposition

Authors: Anuva Kulkarni (*Carnegie Mellon University*), Franz Franchetti (*Carnegie Mellon University*), Jelena Kovacevic (*Carnegie Mellon University*)

Large-scale stress-strain simulations involving parallel Fast Fourier Transforms (FFTs) suffer from high memory requirements and high communication overhead. We propose an irregular domain decomposition method to reduce the memory requirement of an FFT-based stress-strain simulation algorithm for composite materials, the Moulinec-Suquet Composite (MSC) - Basic Scheme. This algorithm uses Green's functions to solve a partial differential equation. FFTs are used for convolution of large 3-D tensor fields with the Green's function.

In this preliminary work, we propose a modified algorithm, the MSC-Alternate Scheme, to show that processing the composite with smaller, local FFTs on irregular domains (grains in the material's microstructure) can reduce memory usage without adversely impacting accuracy of the result. Additionally, data models can reduce communication by compressing the data in the domains before the communication step. Our poster presents our proof-of-concept results and charts out the path towards a GPU

implementation.

P07: PORTAGE - A Flexible Conservative Remapping Framework for Modern HPC Architectures

Authors: Rao V. Garimella (Los Alamos National Laboratory), Peter D. Crossman (Los Alamos National Laboratory), Gary A. Dilts (Los Alamos National Laboratory), Rajeev S. Erramilli (Los Alamos National Laboratory), Charles R. Ferenbaugh (Los Alamos National Laboratory), Angela M. Herring (Los Alamos National Laboratory), Eugene Kinkinon (Los Alamos National Laboratory), Chris M. Malone (Los Alamos National Laboratory), Navamita Ray (Los Alamos National Laboratory), Mike L. Rogers (Los Alamos National Laboratory)

Portage is a massively parallel remapping framework to transfer fields between general polyhedral meshes while conserving integral quantities of interest. The framework also has the capability to remap data between two point clouds. Portage is templated on the component classes required in conservative remapping - search, intersection and interpolation as well as on the mesh and field managers. Applications supply Portage with custom components while the framework takes care of distributed parallelism using MPI and threaded parallelism using NVIDIA Thrust to scale to many thousands of cores. Moreover, the imposition of a functional design on the components used by Portage makes it very amenable to achieve task parallelism with runtime systems such as Legion. Portage has been tested in 2D/3D for remapping between general polygonal and polyhedral meshes and between point clouds. We present scaling results for distributed (MPI) and on-node parallelism (OpenMP) on LANL's HPC machines.

P08: Performance Optimization of Matrix-free Finite-Element Algorithms within deal.II

Authors: Martin Kronbichler (Technical University Munich), Karl Ljungkvist (Uppsala University), Momme Allalen (Leibniz Supercomputing Centre), Martin Ohlerich (Leibniz Supercomputing Centre), Igor Pasichnyk (IBM), Wolfgang A. Wall (Technical University Munich)

We present a performance comparison of highly tuned matrix-free finite element kernels from the deal.II finite element library on three contemporary computer architectures, an NVIDIA P100 GPU, an Intel Knights Landing Xeon Phi, and two multi-core Intel CPUs. The algorithms are based on fast integration on hexahedra using sum factorization techniques. On Cartesian meshes with a relatively high arithmetic intensity, the four architectures provide a surprisingly similar computational throughput. On curved meshes, the kernel is heavily memory bandwidth limited which reveals distinct differences between the architectures: the P100 is twice as fast as KNL, and almost four times as fast as the Haswell and Broadwell CPUs, effectively leveraging the higher memory bandwidth and the favorable shared memory programming model on the GPU.

P09: Adaptive Multistep Predictor for Accelerating Dynamic Implicit Finite-Element Simulations

Authors: Kohei Fujita (University of Tokyo, RIKEN), Tsuyoshi Ichimura (University of Tokyo, RIKEN), Masashi Horikoshi (Intel Corporation), Muneo Hori (University of Tokyo, RIKEN), Lalith Maddegadara (University of Tokyo, RIKEN)

We develop an adaptive multistep predictor for accelerating memory bandwidth-bound dynamic implicit finite-element simulations. We predict the solutions for future time steps adaptively using highly-efficient matrix-vector product kernels with multiple right-hand sides to reduce the number of iterations required in the solver. By applying the method to a conjugate gradient solver with 3 x 3 block Jacobi preconditioning, we were able to achieve a 42% speedup on a Skylake-SP Xeon Gold cluster for a typical earthquake ground motion problem. As the method enables the number of iterations, and thus the communication frequency, to be reduced, the developed solver was able to attain high size-up scalability: 80.6% up to 32,768 compute nodes on the K computer. The developed predictor can also be applied to other iterative solvers and is thus expected to be useful for wide range of dynamic implicit finite-element simulations.

P10: HiCOO: A Hierarchical Sparse Tensor Format for Tensor Decompositions

Authors: Jiajia Li (Georgia Institute of Technology), Jimeng Sun (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

This paper proposes a new Hierarchical COOrdinate (HiCOO) format for sparse tensors, which compresses its indices to units of sparse tensor blocks. HiCOO format does not favor one tensor mode over the others, thus can be used as a replacement of the traditional COOrdinate (COO) format. In this paper, we use HiCOO format for the Matriced Tensor Times Khatri-Rao Product (MTTKRP) operation, the most expensive computational core in the popular CANDECOMP/PARAFAC decomposition, then accelerate it on multicore CPU architecture using two parallel strategies for irregular shaped tensors. Parallel MTTKRP using HiCOO format achieves up to 3.5 $\times$ (2.0 $\times$ on average) speedup over COO format and up to 4.3 $\times$ (2.2 $\times$ on average) speedup over CSF format.

P11: Energy-Efficient Transprecision Techniques for Iterative Refinement

Authors: JunKyu Lee (Queen's University Belfast), Hans Vandierendonck (Queen's University Belfast), Dimitrios S. Nikolopoulos (Queen's University Belfast)

This paper presents transprecision techniques for iterative refinement, which utilize various precision arithmetic dynamically according to numeric properties of the algorithm and computational latencies depending on precisions. The transprecision techniques were plugged into a mixed precision iterative refinement on an Intel Xeon E5-2650 2GHz core with MKL 2017 and XBLAS 1.0. The transprecision techniques brought further 2.0-3.4X speedups and 3.0-4.1X energy reductions to a mixed precision iterative refinement when double precision solution accuracy was required for forward error and a matrix size was ranged from 4K

to 32K.

P12: Multi-Size Optional Offline Caching Algorithms

Authors: Andrew Y. Choliy (Rutgers University), Max D. Whitmore (Brandeis University), Gruia Calinescu (Illinois Institute of Technology)

The optional offline caching (paging) problem, where all future file requests are known, is a variant of the heavily studied online caching problem. This offline problem has applications in web caching and distributed storage systems. Given a set of unique files with varying sizes, a series of requests for these files, fast cache memory of limited size, and slow main memory, an efficient replacement policy is necessary to decide when it is best to evict some file(s) from the cache in favor of another. It is known that this problem is NP-complete, and few approximation algorithms have been proposed. We propose three new heuristics, as well as a 4-approximation algorithm. We then evaluate each algorithm by the metrics of runtime complexity and proximity to the optimal solutions of many synthetic data sets.

P13: Large-Scale GW Calculations on Pre-Exascale HPC Systems

Authors: Mauro Del Ben (Lawrence Berkeley National Laboratory), Felipe H. da Jornada (University of California, Berkeley), Andrew Canning (Lawrence Berkeley National Laboratory), Nathan Wichmann (Cray Inc), Karthik Raman (Intel Corporation), Ruchira Sasanka (Intel Corporation), Chao Yang (Lawrence Berkeley National Laboratory), Steven G. Louie (Lawrence Berkeley National Laboratory; University of California, Berkeley)

The accurate determination of excitation spectra of materials, such as the electronic band gap, is critical for the design of novel devices, including photovoltaics, transistors, batteries, and LEDs. Many-body perturbation-theory methods, and the ab-initio GW approach in particular, have emerged over the last decades as the gold standard for computing these quantities. However, the ab-initio GW formalism is often limited to systems of at most 100 atoms due to its computational complexity. We present here large scale GW calculations of crystalline defect problems, relevant for the performance of semiconductors, with up to 1000 atoms, on the Cori system at NERSC. We show that the GW method is particularly well suited for exascale/pre-exascale systems. Our implementation, which uses a combination of new algorithms and optimizations targeted at many-core CPU architectures, scales well to the entire Cori system, and obtains a significant fraction of peak performance.

P14: Robust SA-AMG Solver by Extraction of Near-Kernel Vectors

Authors: Naoya Nomura (University of Tokyo), Kengo Nakajima (University of Tokyo), Akihiro Fujii (Kogakuin University)

The smoothed aggregation algebraic multigrid (SA-AMG) method is among the fastest solvers for large-scale linear equations, $Ax=b$. The SA-AMG method achieves good convergence and scalability by damping various wavelength components efficiently. To achieve this damping, this method creates multi-level matrices which are hierarchically smaller in dimension than the original matrix. Moreover, the convergence can be further improved by setting near-kernel vectors p , which satisfy $Ap \approx 0$ and $p \neq 0$. Generally, the same number of near-kernel vectors are used at each level. In the present work, we propose a method that extracts and adds near-kernel vectors at each level. We evaluate the performance of the solver that extracts the near-kernel vectors and adds them at each level. We use the three-dimensional elastic problem and employ up to 512 processes on the FX10 supercomputer system. By using this method, the performance is improved compared with previous work.

P15: Toward Decoupling the Selection of Compression Algorithms from Quality Constraints

Authors: Julian Kunkel (German Climate Computing Center), Anastasiia Novikova (University of Hamburg), Eugen Betke (German Climate Computing Center)

With the Scientific Compression Library (SCIL), we are developing a meta-compressor that allows users to set various quantities that define the acceptable error and the expected performance behavior. The library then chooses the appropriate chain of algorithms to yield the users requirements. This approach is a crucial step towards a scientifically safe use of much-needed lossy data compression, because it disentangles the tasks of determining scientific ground characteristics of tolerable noise, from the task of determining an optimal compression strategy given target noise levels and constraints. Without changing applications, it allows these codes to utilize future algorithms once they are integrated into the library.

P16: Scaling Analysis of a Hierarchical Parallelization of Large Inverse Multiple-Scattering Solutions

Authors: Mert Hidayetoglu (University of Illinois), Carl Pearson (University of Illinois), Izzat El-Hajj (University of Illinois), Weng Cho Chew (University of Illinois), Levent Gurel (University of Illinois), Wen-Mei Hwu (University of Illinois)

We propose a hierarchical parallelization strategy to improve the scalability of inverse multiple-scattering solutions. The inverse solver parallelizes the independent forward solutions corresponding to different illuminations. For further scaling out on large numbers of computing nodes, each forward solver parallelizes the dense and large matrix-vector multiplications accelerated by the multilevel fast multipole algorithm. An inverse problem involving a large Shepp-Logan phantom is solved on up to 1,024 CPU nodes of the Blue Waters supercomputer in order to demonstrate the strong-scaling efficiency of the proposed parallelization scheme. The results show that parallelizing illuminations has almost perfect scaling efficiency of 95% because of the independent nature of forward-scattering solutions, however, parallelization of MLFMA has 73% efficiency due to MPI communications in MLFMA multiplications. Nevertheless, the proposed strategy improves granularity and allows spreading DBIM solutions on large numbers of nodes.

P17: Fully Non-Blocking Communication-Computation Overlap Using Assistant Cores toward Exascale Computing

Authors: Motoki Nakata (National Institute for Fusion Science), Masanori Nunami (National Institute for Fusion Science), Shinsuke Satake (National Institute for Fusion Science), Yoshihiro Kasai (Fujitsu Ltd), Shinya Maeyama (Nagoya University), Tomo-Hiko Watanabe (Nagoya University), Yasuhiro Idomura (Japan Atomic Energy Agency)

A fully non-blocking optimized Communication-Computation overlap technique using assistant cores (AC), which are independent from the calculation cores, is proposed for the application to the five-dimensional plasma turbulence simulation code with spectral (FFT) and finite-difference schemes, toward exascale supercomputing. The effects of optimization are examined in Fujitsu FX100 (2.62PFlop/s) with 32 ordinary cores and 2 Assistant cores/node, where AC enables us to employ the fully non-blocking MPI communications overlapped by the thread-parallelized calculations with OpenMP Static scheduling with much less overheads. It is clarified that the combination of the non-blocking communications by AC and the static scheduling leads to not only reduction in OpenMP overhead, but also improved load/store and cash performance, where about 22.5% improved numerical performance is confirmed in comparison to the conventional overlap by the master thread communications with dynamic scheduling.

P18: A Parallel Python Implementation of BLAST+ (PPIB) for Characterization of Complex Microbial Consortia

Authors: Amina Jackson (Naval Research Laboratory), William Connor Horne (Naval Research Laboratory), Daniel Beall (Naval Research Laboratory), Kenneth Jiang (Naval Research Laboratory), William Judson Hervey (Naval Research Laboratory)

Technological advancements in analytical instrumentation have enabled large-scale data acquisitions among the ‘-omics’ sciences of genomics, transcriptomics, and proteomics. An essential application among ‘-omics’ disciplines is the Basic Local Alignment Search Tool (BLAST) for functional inference of biomolecules. Though implementations of BLAST+ have been modified to address data volume growth, such improvements have neither been consistently maintained for high performance computing (HPC), nor have they been applied to complex microbiomes. Further, such implementations do not scale well to microbiomes of Naval interest on HPC systems in our hands.

Here, we compare 2 HPC implementations: BLAST+ and a Parallel Python Implementation of BLAST+ (PPIB) for protein functional inference. PPIB enabled functional inference of 2 complex microbiomes, which may be attributed to a combination of MPI and Python multiprocessing to query up to 3,600 proteins simultaneously. In contrast, BLAST+ did not complete functional assignments relative to PPIB at a comparable walltime.

P19: MPI-GIS: An MPI System for Big Spatial Data

Authors: Satish Puri (Marquette University)

In recent times, geospatial datasets are growing in terms of size, complexity and heterogeneity. High performance systems are needed to analyze such data to produce actionable insights in an efficient manner. For polygonal a.k.a vector datasets, operations such as I/O, data partitioning, and communication becomes challenging in a cluster environment.

In this work, we present MPI-GIS equipped with MPI-Vector-IO, a parallel I/O library that we have designed using MPI-IO specifically for irregular polygonal (vector) data formats such as Well Known Text, XML, etc. Our system can perform spatial in-memory indexing and join efficiently for an order of magnitude larger datasets compared to our previous work. It makes MPI aware of spatial data and spatial primitives and provides support for spatial data types embedded within collective computation and communication using MPI message-passing library. It takes less than 2 minutes to scan through 2.7 billion geometries in 96GB file using 160 processes.

P20: Facilitating the Scalability of ParSplice for Exascale Testbeds

Authors: Vinay B. Ramakrishnaiah (University of Wyoming), Jonas L. Landsgesell (University of Stuttgart), Ying Zhou (Loughborough University), Iris Linck (University of Colorado, Denver), Mouad Ramil (National School of Bridges and Roads - ParisTech), Joshua Bevan (University of Illinois), Danny Perez (Los Alamos National Laboratory), Louis J. Vernon (Los Alamos National Laboratory), Thomas D. Swinburne (Los Alamos National Laboratory), Robert S. Pavel (Los Alamos National Laboratory), Christoph Junghans (Los Alamos National Laboratory)

Parallel trajectory splicing (ParSplice) is an attempt to solve the enduring challenge of simulating the evolution of materials over long time scales for complex atomistic systems. A novel version of ParSplice is introduced with features that could be useful in its scaling to exascale architectures. A two-pronged approach is used. First, latent parallelism is exploited by extending support to heterogeneous architectures, including GPUs and KNLs. Second, the efficiency of the Kinetic Monte Carlo predictor is improved, allowing enhanced parallel speculative execution. The key idea in these predictor modifications is to include statistics from higher temperature simulations. The issue of inherent uncertainty in the prediction model was addressed in order to improve the performance, as the current predictor only takes into account the previous observations to formulate the problem. The predictor was also improved by using a hybrid approach of message-passing + multi-threading. (LA-UR-17-26181)

P21: The First Real-Scale DEM Simulation of a Sandbox Experiment Using 2.4 Billion Particles

Authors: Mikito Furuichi (Japan Agency for Marine-Earth Science and Technology), Daisuke Nishiura (Japan Agency for Marine-Earth Science and Technology), Mitsuteru Asai (Kyushu University), Takane Hori (Japan Agency for Marine-Earth Science and Technology)

A novel implementation of the Discrete Element Method (DEM) for a large parallel computer system is presented to simulate a sandbox experiment with realistic particle sizes. To save memory in the pairwise tangential forces and halve the arithmetic costs, interactions are calculated using the action-reaction law. An iterative load-balancer the flexible 2D orthogonal domain decomposition is applied to overcome the load-imbalance problem caused by the Lagrangian nature of DEM. An overlapping communication technique combined with cell-ordering with space-filling curves is also applied to hide the overhead cost because of the MPI communication tasks. We verify our complex parallel implementation with the action-reaction law via a reproducibility test. The parallel scaling test shows good strong, and weak scalabilities up to 2.4 billion particles on the Earth Simulator and the K computer. The world's first real-scaled numerical sandbox simulation successfully captures the characteristics of real observations.

P22: Numerical Simulation of Snow Accretion by Airflow Simulator and Particle Simulator

Authors: Kohei Murotani (Railway Technical Research Institute), Koji Nakade (Railway Technical Research Institute), Yasushi Kamata (Railway Technical Research Institute)

In this research, to take countermeasures for the snow accretion damage, we developed a simulator of realizing the snow accretion process in the following steps. Firstly, air flow analysis is performed by "Airflow simulator" developed by RTRI (Railway Technical Research Institute). Secondly, trajectory of flying snow is calculated by Basset-Boussinesq-Oseen equation using distributed of velocity of air flow. Thirdly, snow accretion analysis is performed by "Particle simulator" developed by RTRI. The shape modified by snow accretion is reflected onto the boundary conditions of the air flow analysis. In this year, snow accretion analysis for simple cubic shapes is performed in order to aim at system development and validation.

P23: AI with Super-Computed Data for Monte Carlo Earthquake Hazard Classification

Authors: Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Muneo Hori (University of Tokyo, RIKEN), Madgededara Lalith (University of Tokyo, RIKEN), Naonori Ueda (RIKEN)

Many problems associated with earthquakes are yet to be solved using heroic computing, which is defined as computing at the largest scale possible using the best supercomputers and algorithms. Thus, a continuous effort has been pursued in HPC to solve these problems. However, even when heroic computing is applied, its practical use is difficult without considering the uncertainties in models. In this study, we constructed an AI methodology that uses super-computed data generated using heroic computing. We applied this AI to an earthquake hazard classification including uncertainty analyses in order to demonstrate its utility. This study can be regarded as an innovative step towards realizing high quality computing for Earthquakes by exploiting the potential of HPC through AI.

BP

P24: A Deployment of HPC Algorithm into Pre/Post-Processing for Industrial CFD on K-Computer

Authors: Keiji Onishi (RIKEN), Niclas Jansson (KTH Royal Institute of Technology), Rahul Bale (RIKEN), Wei-Hsiang Wang (RIKEN), Chung-Gang Li (Kobe University, RIKEN), Makoto Tsubokura (Kobe University, RIKEN)

Pre- and post-processing is still a major problem in industrial computational fluid dynamics (CFD). With the rapid development of computers, physical solvers are getting faster, while pre- remains slow because it's mainly a serial process. A methodology using MPI+OpenMP hybrid parallelization has been proposed to eliminate the manual work required during pre-processing for correcting the surface imperfections of CAD data. Compared to the rapidly increasing amount of data in recent years, the speed-up of visualization is insufficient. We address this limitation of post- by adapting the in-situ visualization to parallelize the post-processing using libsim (Visit) library. The performance of pre-/post- processing is investigated in this work, and we show that the pre-processing time has been reduced from several days in the conventional framework to order of minutes. The post-processing time has been reduced seconds order per frame, and approximately 30% increase of computational time was observed in vehicle aerodynamics cases.

P25: Large-Scale Adaptive Mesh Simulations Through Non-Volatile Byte-Addressable Memory

Authors: Bao Nguyen (Washington State University, Vancouver), Hua Tan (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

Octree-based mesh adaptation has enabled simulations of complex physical phenomena. Existing meshing algorithms were proposed with the assumption that computer memory is volatile. Consequently, for failure recovery, the in-core algorithms need to save memory states as snapshots with slow file I/Os. The out-of-core algorithms store octants on disks for persistence. However, neither of them was designed to leverage unique characteristics of non-volatile byte-addressable memory (NVBM). We propose a novel data structure Persistent Merged octree (PM-octree) for both meshing and in-memory storage of persistent octrees using NVBM. It is a multi-version data structure and can recover from failures using its earlier persistent version stored in NVBM. In addition, we design a feature-directed sampling approach to help dynamically transform the PM-octree layout for reducing NVBM-induced memory write latency.

P26: Optimizing Gravity and Nuclear Physics in FLASH for Exascale

Authors: Hannah Klion (University of California, Berkeley; Oak Ridge National Laboratory), Bronson Messer (Oak Ridge National Laboratory, University of Tennessee), J. Austin Harris (Oak Ridge National Laboratory), Thomas Papatheodore (Oak Ridge National Laboratory)

In a Type Ia supernova, runaway fusion ignites in a white dwarf, causing it to explode. The heavy element yields of these events remain uncertain, and high-performance multiphysics simulations with tools like FLASH are critical for our understanding. Current simulations track approximately a dozen nuclear isotopes, as opposed to the thousands required to completely capture the event's nuclear physics.

Simulating nuclear physics and self-gravity accurately and efficiently is critical for modeling a Type Ia supernova, since supernovae are competitions between energy-releasing nuclear reactions and gravity. Currently, the FLASH nuclear reaction network and self-gravity solver requires substantial inter-node communication. We use non-blocking MPI collectives to overlap communication in the self-gravity calculation with the computation-heavy nuclear burning calculation. We find that speedups from this technique are possible, but are MPI implementation-dependent. We highlight some of the challenges associated with this type of optimization.

P27: Parallelization of the Particle-In-Cell Monte Carlo Collision (PIC-MCC) Algorithm for Plasma Simulation on Intel MIC Xeon Phi Architecture

Authors: Keval Shah (*Dhirubhai Ambani Institute of Information and Communication Technology*), Anusha Phadnis (*Dhirubhai Ambani Institute of Information and Communication Technology*), Miral Shah (*Dhirubhai Ambani Institute of Information and Communication Technology*), Bhaskar Chaudhury (*Dhirubhai Ambani Institute of Information and Communication Technology*)

The implementation of 2D-3v (2D in space and 3D in velocity space) PIC-MCC (Particle-In-Cell Monte Carlo Collision) method involves the computational solution of Vlasov-Poisson equations. This provides the spatial and temporal evolution of the charged-particle velocity distribution functions in plasma under the effect of self-consistent electromagnetic fields and collisions. Stringent numerical constraints associated with the PIC code makes it computationally prohibitive on CPU.

In our work, parallelization and optimization techniques have been extended to this simulation, along with a novel approach that involves developing a 'self-aware' code that triggers sorting in order to maintain cache-coherence while reducing the total sorting time during iterations.

We present the effect of important numerical parameters on speed-up. Finally, we compare the scalability and performance of the parallelization and optimization strategies on Intel® Xeon™ E5-2630, Xeon Phi™ 5110p and Xeon Phi™ 7250 relative to a serial implementation on Intel® i5.

P28: High-Fidelity Blade-Resolved Wind Plant Modeling

Authors: Andrew C. Kirby (*University of Wyoming*), Zhi Yang (*University of Wyoming*), Michael J. Brazell (*University of Wyoming*), Behzad R. Ahrabi (*University of Wyoming*), Jay Sitaraman (*Parallel Geometric Algorithms LLC*), Dimitri J. Mavriplis (*University of Wyoming*)

Blade-resolved numerical simulations of wind energy applications using full blade and tower models are presented. The computational methodology combines solution technologies in a multi-mesh, multi-solver paradigm through a dynamic overset framework. The coupling of a finite-volume solver and a high-order, hp-adaptive finite-element solver is utilized. Additional technologies including in-situ visualization and atmospheric micro-scale modeling are incorporated into the analysis environment. Validation of the computational framework is performed on the NREL 5MW wind turbine, the unsteady aerodynamics experimental NREL Phase VI turbine, and the Siemens SWT-2.3-93 wind turbine. The power and thrust results of all single turbine simulations agree well with low-fidelity model simulation results and field experiments when available. Scalability of the computational framework is demonstrated using 6, 12, 24, 48, and 96 wind turbine wind plant set-ups including the 48 turbine wind plant known as Lillgrund. Demonstration of the coupling of atmospheric micro-scale and CFD solvers is presented.

P29: A Deep Learning Tool for Fast Simulation

Authors: Sofia Vallecorsa (*CERN*), Gul rukh Khattak (*CERN*), Shruti Sharan (*CERN*)

We present the first application of Volumetric Generative Adversarial Network (VGAN) to High Energy Physics simulation. We generate three dimensional images of particles depositing energy in calorimeters. This is the first time such an approach is taken in HEP where most of data is three dimensional in nature but it is customary to convert it into two dimensional slices. The volumetric approach leads to a larger number of parameters, but two dimensional slicing loses the volumetric dependencies inherent in the dataset. The present work proves the success of handling those dependencies through VGANs. Energy showers are faithfully reproduced in all dimensions and show a reasonable agreement with standard techniques. We also demonstrate the ability to condition training on several parameters such as particle type and energy. This work aims at proving Deep Learning techniques represent a valid fast alternative to standard MonteCarlo approaches and is part of the GEANTV project.

P30: MPI/OpenMP Parallelization of the Hartree-Fock Method for the Second Generation Intel Xeon Phi

Authors: Kristopher Keipert (*Iowa State University*), Vladimir Mironov (*Lomonosov Moscow State University*), Yuri Alexeev (*Argonne National Laboratory*), Michael D'mello (*Intel Corporation*), Alexander Moskovsky (*RSC Technologies*), Mark S. Gordon (*Iowa State University*)

Replication of critical data structures in the MPI-only GAMESS Hartree-Fock algorithm limits the full utilization of the manycore Intel Xeon Phi processor. In this work, modern OpenMP threading techniques are used to implement hybrid MPI/OpenMP algorithms. Two separate implementations that differ by the sharing and replication details of key data structures among threads are

considered. The hybrid MPI/OpenMP implementations reduce the memory footprint by approximately 200 times compared to the legacy code. The MPI/OpenMP code was shown to run up to six times faster than the original for a range of molecular system sizes. The implementation details and strategies will be presented for both hybrid algorithms. Benchmark scaling results utilizing up to 3000 Intel Xeon Phi processors will also be discussed.

BP

P31: Understanding the Performance of Small Convolution Operations for CNN on Intel Architecture

Authors: Alexander Heinecke (Intel Corporation), Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalmakar (Intel Corporation), Narayanan Sundaram (Intel Corporation), Anand Venkat (Intel Corporation), Greg Henry (Intel Corporation), Hans Pabst (Intel Corporation)

Convolution layers are prevalent in many classes of deep neural networks, including Convolutional Neural Networks (CNNs) which provide state-of-the-art results for tasks like image recognition, natural language processing, and speech recognition. The computationally expensive nature of a convolution operation has led to the proliferation of implementations including matrix-matrix multiplication formulation, FFT-formulation, Winograd transformation, and direct convolution primarily targeting GPUs. In this paper, we optimize a direct convolution and Winograd implementation for x86 architectures, in particular for Xeon Phi systems, via a dynamic compilation approach. We then show how these JIT optimizations can be integrated in a high-level domain-specific language setting. We shed light on what is possible and what is not possible based on different data-formats and blocking techniques. Our JIT-based Ninja implementation shows close to theoretical peak results on modern x86 architectures, depending on setting and the CPU architecture at hand.

P32: Exploring the Performance of Electron Correlation Method Implementations on Kove XPDs

Authors: Colleen Berton (Argonne National Laboratory), Brian Toonen (Argonne National Laboratory), William Allcock (Argonne National Laboratory), Spencer R. Pruitt (Worcester Polytechnic Institute), Mark S. Gordon (Iowa State University)

In electron correlation methods in quantum chemistry, there are often high memory requirements which can reach terabytes for medium-sized molecular systems. For second-order perturbation theory (MP2), the two-electron integral arrays are the main memory bottleneck. Previously the two-electron integrals were recomputed, stored in distributed memory, or stored on disk. A way of storing the arrays which would remove the dependence on compute node memory and large latency associated with using disk is by using an external memory appliance, like Kove's XPD.

In this work, we modified a distributed memory implementation of MP2 to use XPDs instead of distributed memory. We evaluated the performance of our implementation against the distributed memory version for several molecular systems by considering scaling behavior with respect to compute processes and connections to XPDs. In the poster, we present an outline of the MP2 implementation using XPDs and the scaling results.

P33: Massively Parallel Evolutionary Computation for Empowering Electoral Reform: Quantifying Gerrymandering via Multi-objective Optimization and Statistical Analysis

Authors: Wendy K. Cho (National Center for Supercomputing Applications, University of Illinois), Yan Liu (National Center for Supercomputing Applications, University of Illinois)

Important insights into redistricting can be gained by formulating and analyzing the problem within a large-scale optimization framework. Redistricting is an application of the set-partitioning problem that is NP-hard. We design a hybrid metaheuristic as the base search algorithm. With our grant on the Blue Waters supercomputer, we extend our algorithm to the high-performance-computing realm by using MPI to implement an asynchronous processor communication framework. We experimentally demonstrate the effectiveness of our algorithm to utilize multiple processors and to scale to 131,072 processors. The massive computing power allows us to extract new substantive insights that closely mesh with the framework that the Supreme Court has elucidated for electoral reform.

P34: GPU Acceleration for the Impurity Solver in GW+DMFT Packages

Authors: Kwangmin Yu (Brookhaven National Laboratory), Patrick Semon (Brookhaven National Laboratory), Nicholas D'Imperio (Brookhaven National Laboratory)

The combination of dynamical mean field theory (DMFT) and GW (or density functional theory) has become a powerful tool to study and predict properties of real materials with strongly correlated electrons, such as high temperature superconductors. At the core of this combined theory lies the solution of a quantum impurity model, and continuous-time quantum Monte Carlo (CT-QMC) has proven an indispensable algorithm in this respect. However, depending on the material, this algorithm is computationally very expensive, and enhancements are crucial for bringing new materials within reach of GW+DMFT. Based on a CPU implementation, GPU acceleration is added, and two times speedup is achieved. New techniques are invented and implemented to deal with various GPU acceleration environment.

P36: A Novel Feature-Preserving Spatial Mapping for Deep Learning Classification of Ras Structures

Authors: Thomas Corcoran (Lawrence Berkeley National Laboratory), Rafael Zamora-Resendiz (Lawrence Berkeley National Laboratory), Xinlian Liu (Lawrence Berkeley National Laboratory), Silvia Crivelli (Lawrence Berkeley National Laboratory)

A protein's 3D structure determines its functionality, and is therefore a topic of great importance. This work leverages the power of

Convolutional Neural Networks (CNNs) to classify proteins and extract features directly from their 3D structures. So far, researchers have been unable to fully exploit 3D structural information with 2D CNNs, partly because it is difficult to encode 3D data into the 2D format that can be ingested by such networks. We designed and implemented a novel method that maps 3D models to 2D data grids as a preprocessing step for 2D CNN use. Our experiments focused on the Ras protein family, which has been linked to various forms of cancer. Our trained CNNs are able to distinguish between two branches within the Ras family, HRas and KRas, which are similar in sequence and structure. Analysis of saliency maps suggests classification is accomplished by detection of structurally and biologically-meaningful sites.

P37: PaSTRI: A Novel Data Compression Algorithm for Two-Electron Integrals in Quantum Chemistry

Authors: Ali Murat Gok (Argonne National Laboratory, Northwestern University), Dingwen Tao (University of California, Riverside), Sheng Di (Argonne National Laboratory), Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Franck Cappello (Argonne National Laboratory)

Integral computations for two-electron repulsion energies are very frequently used applications in quantum chemistry. Computational complexity, energy consumption and the size of the output data generated by these computations scales with $O(N^4)$, where N is the number of atoms simulated. Typically, the same integrals are calculated multiple times. Storing these values and reusing them requires impractical amounts of storage space; whereas recalculating them requires a lot of computations. We propose PaSTRI (Pattern Scaling for Two-electron Repulsion Integrals), a fast novel compression algorithm which makes it possible to calculate these integrals only once, store them, and reuse them at much smaller computational cost than recalculation. PaSTRI is "lossy" compared to floating point numbers, but still maintains the precision level required by the integral computations. PaSTRI is an extension to SZ compressor package as a part of ECP-EZ. PaSTRI achieves 17.5:1 compression ratio whereas vanilla SZ achieves 8.0:1 and ZFP achieves 7.1:1.

P38: Benchmarking Parallelized File Aggregation Tools for Large Scale Data Management

Authors: Tiffany Li (National Center for Supercomputing Applications, University of Illinois), Craig Steffen (National Center for Supercomputing Applications, University of Illinois), Ryan Chui (National Center for Supercomputing Applications, University of Illinois), Roland Haas (National Center for Supercomputing Applications, University of Illinois), Liudmila S. Mainzer (National Center for Supercomputing Applications, University of Illinois)

Large-scale genomic data analyses have given rise to bottlenecks in data management due to the production of many small files. Existing file-archiving utilities, such as tar, are unable to efficiently package large datasets with upward of multiple terabytes and hundreds of thousands of files. To create parallelized and multi-threaded alternatives, ParFu (parallel archiving file utility), MPItar, and ptgz (parallel tar gzip) were developed by the Blue Waters team and the NCSA Genomics team as efficient data management tools, with the ability to perform parallel archiving (and eventually extracting). Scalability was tested for each tool as a function of the number of ranks executed and stripe count on a Lustre filesystem. We used two datasets typically seen in genomic analyses to measure the effects of different file-size distributions. These tests suggest the best user parameters and subsequent costs for usage as efficient replacements of data-packaging tools.

P39: Extremely Large, Wide-Area Power-Line Models

Authors: Ross N. Adelman (US Army Research Laboratory)

The electric and magnetic fields around power lines carry an immense amount of information about the power grid, and can be used to improve stability, balance loads, and reduce outages. To study this, extremely large models of transmission lines over a 49.5-sq-km tract of land near Washington, DC have been built. The terrain is modeled accurately using 1-m-resolution LIDAR data. The models are solved using the boundary element method, and the solvers are parallelized across Army Research Laboratory's Centennial supercomputer using a modified version of the domain decomposition method. The code on each node is accelerated using the fast multipole method and, when available, GPUs. Additionally, larger test models were used to characterize the scaling properties of the code. The largest test model had 10,020,913,152 elements, and was solved across 1024 nodes in 3.0 hours.

P40: Running Large-Scale Ultrasound Simulations on Piz Daint with 512 Pascal GPUs

Authors: Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Ultrasound simulation is a critical component of model-based treatment planning for ultrasound therapy. However, the domains are typically thousands of wavelengths in size, leading to large-scale numerical models with 10s of billions of unknowns. This paper presents a novel local Fourier basis domain decomposition for full-wave ultrasound propagation simulations with a custom bell function which ensures that the numerical error stays below 0.1% while enabling almost ideal strong scaling. Realistic benchmarks with 512 Nvidia P100 GPUs in the best EU supercomputer Piz Daint achieved efficiency between 90 and 100% with a speed-up over 100 and computational cost reduction by a factor of 12 compared to 1024 Haswell cores.

P41: OpenCL-Based High-Performance 3D Stencil Computation on FPGAs

Authors: Hamid Reza Zohouri (Tokyo Institute of Technology), Artur Podobas (Tokyo Institute of Technology), Naoya Maruyama (RIKEN), Satoshi Matsuoka (Tokyo Institute of Technology)

With the recent advancements in OpenCL-based High-Level Synthesis, FPGAs are now more attractive choices for accelerating

High Performance Computing workloads. Despite their power efficiency advantage, FPGAs usually fall short in terms of sheer performance against GPUs due to having multiple times lower memory bandwidth and compute performance. In this work, we show that due to the architectural advantage of FPGAs for stencil computation, apart from power efficiency, these devices can also offer comparable performance to high-end GPUs. We achieve this goal using a parameterized OpenCL-based implementation that employs both spatial and temporal blocking, and multiple advanced FPGA-specific optimizations to maximize performance. We show that it is possible to achieve up to 60 GBps and 230 GBps of effective throughput for 3D stencil computation on Intel Stratix V and Arria 10 FPGAs, respectively, which is comparable to a highly-optimized implementation on high-end GPUs.

P42: TRIP: An Ultra-Low Latency, TeraOps/s Reconfigurable Inference Processor for Multi-Layer Perceptrons

Authors: Ahmed Sanaullah (Boston University), Chen Yang (Boston University), Yuri Alexeev (Argonne National Laboratory), Kazutomo Yoshii (Argonne National Laboratory), Martin C. Herbordt (Boston University)

Multi-Layer Perceptron (MLP) is one of the most commonly deployed Deep Neural Networks, representing 61% of the workload in Google data-centers. MLP Inference, a memory bound problem, typically has hard response time deadlines and prefers latency over throughput. In our work, we designed a TeraOps/s Reconfigurable Inference Processor for MLPs (TRIP) on FPGAs that alleviates the memory bottleneck by storing all application specific weights on-chip. It can be deployed in multiple configurations, including host-independent operation. We have shown that TRIP achieves 60x better performance than the current state-of-the-art Google Tensor Processing Unit (TPU) for MLP Inference. It was demonstrated on the cancer patient datasets used in the Candle Exascale Computing Project (ECP).

P43: Deep Packet/Flow Analysis Using GPUs

Authors: Qian Gong (Fermi National Laboratory), Wenji Wu (Fermi National Laboratory), Phil DeMar (Fermi National Laboratory)

Deep packet inspection (DPI) faces severe challenges in high-speed networks as it requires high I/O throughputs and intensive computations. The parallel architecture of GPUs fits exceptionally well for per-packet traffic processing. However, TCP data stream need to be reconstructed in a per-flow level to deliver a consistent content analysis. Since the flow-centric operations are naturally anti-parallel and often require large memory space for buffering out-of-sequence packets, they can be problematic for GPUs. Here, we present a highly efficient DPI framework, which includes a purely GPU-implemented TCP flow tracking and stream reassembly. Instead of buffering till the TCP packets become in sequence, we process the packets in batch with pattern matching states between consecutive batches connected by a Aho-Corasick with a prefix-/suffix- tree method. Evaluation shows that our code can reassemble tens of millions of packets per second and conduct a signature-based DPI at 55 Gbit/s using an NVIDIA K40 GPU.

P44: Increasing Throughput of Multiprogram HPC Workloads: Evaluating a SMT Co-Scheduling Approach

Authors: Elias Lundmark (University West Sweden), Chris Persson (University West Sweden), Andreas de Blanche (University West Sweden, Tetra Pak), Thomas Lundqvist (University West Sweden)

Simultaneous Multithreading (SMT) is a technique that allows for more efficient processor utilization by scheduling multiple threads on a single physical core, thus increasing the instruction level parallelism as it mixes instructions from several threads. Previous research has shown an average throughput increase of around 20% with an SMT level of two, e.g. two threads per core. However, a bad combination of threads can actually result in decreased performance. To be conservative, many HPC-systems have SMT disabled, thus, limiting the number of scheduling slots in the system to one per core. However, for SMT to not hurt performance, we need to determine which threads should share a core. In this poster, we use 30 random SPEC CPU job mixed on a twelve-core Broadwell based node, to study the impact of enabling SMT using two different co-scheduling strategies. The results show that SMT can increase performance especially when using no-same-program co-scheduling.

P45: Campaign Storage: Erasure Coding with GPUs

Authors: Walker Haddock (University of Alabama, Birmingham), Matthew L. Curry (Sandia National Laboratories), Purushotham Bangalore (University of Alabama, Birmingham), Anthony Skjellum (University of Tennessee, Chattanooga)

Cloud computing has developed high capacity, reliable and economical storage systems based on object technology. Los Alamos National Labs has designed the storage systems for Trinity to include a cloud type object storage system as a layer between the Parallel File System (PFS). This pre-archive system has been dubbed "Campaign Storage" with the purpose of storing data products to be quickly accessible during the life of a research project. Data stored on the Campaign Storage can be pulled into the PFS or moved to archive after the data has been curated. Campaign Storage reduces the capacity requirements for PFS storage and reduces the data transfer bandwidth requirements for the archive storage.

We make these contributions to the pre-archive storage layer: * GPU assisted erasure coding * Demonstrating erasure on File Transfer Agents * Reducing erasure recovery costs with "lazy recovery" * Enabling larger erasure coded disk pools

P46: Understanding How OpenCL Parameters Impact on Off-Chip Memory Performance of FPGA Platforms

Authors: Yingyi Luo (Northwestern University), Zheming Jin (Argonne National Laboratory), Kazutomo Yoshii (Argonne National Laboratory), Seda Ogreni-Memik (Northwestern University)

Reconfigurability has strong potential to achieve higher performance and energy efficiency in the post-Moore era. Field-programmable gate arrays (FPGAs), the most practical reconfigurable architecture today, are becoming more relevant to scientific

computing thanks to hardened floating-point circuits and emerging FPGA high-level synthesis (HLS) technology. Most notably, FPGA vendors started supporting OpenCL for FPGA platforms, and some OpenCL-based codes have been ported to FPGAs. However, OpenCL offers no guarantee for performance portability; optimal OpenCL parameters such as global size and local size are different between platforms, which could lead to unfair comparisons. In this study, our objective is twofold: 1) to understand how OpenCL parameters impact off-chip memory access performance of the current generation of OpenCL-FPGA platforms and 2) to find effective OpenCL parameters empirically from microbenchmark results.

P47: Understanding Congestion on Omni-Path Fabrics

Authors: Lauren Gillespie (Southwestern University), Christopher Leap (University of New Mexico), Dan Cassidy (Los Alamos National Laboratory)

High-performance computing systems require high-speed interconnects, such as InfiniBand (IB), to efficiently transmit data. Intel's Omni-Path Architecture (OPA) is a new interconnect similar to IB that is implemented on some of Los Alamos National Laboratory's recent clusters. Both interconnects suffer from degraded performance under heavy network traffic loads, resulting in packet discards. However, unlike IB, OPA specifically calls out these drops in the form of the performance counter, congestion discards. Owing to the relative immaturity of the OPA fabric technology, the correlation between performance degradation and congestion discards has not been fully evaluated to date. This research aims to increase the level of understanding of the effects congestion has on cluster performance by presenting a sufficiently high data injection load to the OPA fabric such that performance degradation is induced and the cause of this performance degradation can be evaluated. LA-UR-17-26341

P48: Prototyping of Offloaded Persistent Broadcast on Tofu2 Interconnect

Authors: Yoshiyuki Morie (RIKEN), Masayuki Hatanaka (RIKEN), Masamichi Takagi (RIKEN), Atsushi Hori (RIKEN), Yutaka Ishikawa (RIKEN)

With the increasing scale of parallel computers, it has become more important to reduce communication time. Overlapping computation and communication is one effective method for hiding communication delay. Although standard non-blocking collective communication is an overlap method, it requires generating a communication command sequence for each collective communication. In contrast, persistent non-blocking collective communication can generate the sequence at initialization and reuse it at the start of collective communication. Moreover, if the sequence can be offloaded to a network device, more efficient execution is possible without using CPU cycles.

In this poster, a persistent non-blocking broadcast is implemented using the offloading functionality of the Tofu2 interconnect on the Fujitsu FX100 supercomputer, the successor to the K computer. We report the performance improvement by offloading persistent non-blocking collective communication in a real machine.

P49: Toward Exascale HPC Systems: Exploiting Advances in High Bandwidth Memory (HBM2) through Scalable All-to-All Optical Interconnect Architectures

Authors: Pouya Fotouhi (University of California, Davis), Roberto Proietti (University of California, Davis), Paolo Grani (University of California, Davis), Mohammad Amin Nazirzadeh (University of California, Davis), S. J. Ben Yoo (University of California, Davis)

As we reach the limits of miniaturization in fabrication processes, the interpretation of Moore's law has changed from doubling the frequency every eighteen months to doubling the core count every three to four years (from 2 cores in 2004 to 16 cores in 2015). To reach exascale-level computation, the communication and data transfers between processors and memory is expected to increase drastically; the on-chip interconnect plays a key role in the overall system latency and energy-efficiency. Therefore, novel solutions providing one order of magnitude higher bandwidth and lower energy consumption than what is possible with current electrical interconnects are needed. This poster discusses an optical interconnect compute node that makes use of embedded photonic interconnects together with emerging high bandwidth memory technologies (such as HBM and HBM2). Two different multi-processors architectures with different requirements in terms of number of lasers, high-speed SERDES, and memory bandwidth per processors are presented.

P50: Energy-Efficient and Scalable Bio-Inspired Nanophotonic Computing

Authors: Mohammadamin Nazirzadeh (University of California, Davis), Pouya Fotouhi (University of California, Davis), Mohammad Sadegh Shamsabadeh (University of California, Davis), Roberto Proietti (University of California, Davis), S. J. Ben Yoo (University of California, Davis)

This paper discusses bio-inspired neuromorphic computing utilizing nanophotonic, nanoelectronic, and NEMS technologies integrated into reconfigurable 2D-3D integrated circuits as hierarchical neural networks. The goal is to achieve $\geq 1000\times$ improvements in energy-per-operation compare to the state-of-the-art implementations of neural networks on Von-Neumann based computers. We combine nanophotonic and nanoelectronic technologies to build energy-efficient (~ 10 fJ/b) artificial spiking neurons with required functionality (spiking, integration, thresholding, reset). Photonic interconnects exploiting 2×2 NEMS-MZIs enables distance independent propagation of signal with weighted addition among the neurons as well as possibility of on-line learning capability. Using low-leakage nanophotonic and nanoelectronic devices, and NEMS, the static power consumption of the system can be decreased down to nearly zero. Realizing 2D-3D photonic integrated circuit technologies, the proposed system can overcome the scalability limitations of current neuromorphic computing architectures.

P51: TuPiX-Flow: Workflow-Based Large-Scale Scientific Data Analysis System

Authors: Sunggeun Han (Korea Institute of Science and Technology Information), Jung-Ho Um (Korea Institute of Science and Technology Information), Hyunwoo Kim (Korea Institute of Science and Technology Information), Kyongseok Park (Korea Institute of Science and Technology Information)

With more research being actively conducted on big data, there has been a growing interest towards data-intensive data science in various fields. Against this backdrop, extensive efforts have been made to apply High Performance Computing (HPC) technology such as distributed computing or parallel computing. However, many researchers are unable to fully utilize such technology due to their lack of knowledge, programming skills, and analytical skills for large-scale data. TuPiX-Flow, which provides a workflow-based user interface, enables researchers to easily analyze large-scale data using workflow diagrams for end-to-end analytical processes even without programming knowledge or implementation technology. In addition, large-scale data can be efficiently analyzed in software compatible with HPC, including distributed computing and parallel computing. As a case study of large-scale data analysis using TuPiX-Flow, a model for the detection of red tides surrounding the Korean peninsula was analyzed based on ocean color satellite data.

P52: A Simulation-Based Analysis on the Configuration of Burst Buffer

Authors: Tianqi Xu (Tokyo Institute of Technology), Kento Sato (Lawrence Livermore National Laboratory), Satoshi Matsuoka (Tokyo Institute of Technology)

Burst buffers have been widely deployed in many supercomputer systems to absorb bursty I/O and accelerate I/O performance. Previous work has shown that with burst buffer systems, I/O operations from computer nodes can be greatly accelerated. Different configurations of burst buffers can have huge impact on performance of applications. However, the proper configuration of burst buffers for given systems and workloads still remains an open problem. In this paper, we address this challenge by simulating the behavior of a burst buffer under different buffer sizes with trace logs from a set of HPC applications in a distributed environment. We tuned our simulator with a production level burst buffer system. From the results of the simulation, we found that for most of HPC applications, using a buffer size that is less than half of the total access space of the application can still achieve high performance.
BP

P53: TensorViz: Visualizing the Training of Convolutional Neural Network Using ParaView

Authors: Xinyu Chen (University of New Mexico), Qiang Guan (Los Alamos National Laboratory), Xin Liang (University of California, Riverside), Li-Ta Lo (Los Alamos National Laboratory), Simon Su (US Army Research Laboratory), Trilce Estrada (University of New Mexico), James Ahrens (Los Alamos National Laboratory)

Deep Convolutional Networks have been very successful in visual recognition tasks recently. Previous works visualize learned features at different layers to help people understand how CNNs learn visual recognition tasks. However, they do not help to accelerate the training process. We use ParaView to provides both qualitative and quantitative visualization that help understand the learning procedure, tune the learning parameters, and direct merging and pruning of neural networks.

P54: Investigating Hardware Offloading for Reed-Solomon Encoding

Authors: John W. Dermer (Los Alamos National Laboratory), Gustavo De Leon (Los Alamos National Laboratory; University of California, Berkeley), Tyler S. Rau (Los Alamos National Laboratory)

Reed-Solomon (RS) encoding is a storage scheme which offers better scalability, but requires heavier computation, compared to other models. This presents a problem as it requires users to purchase brawnier CPUs than would be otherwise necessitated. However, Mellanox's ConnectX-4 Infiniband cards have the capability to perform RS encoding on the HCA hardware; removing the need for powerful CPUs to calculate it. We investigated the performance, measured in throughput, between these cards and Intel's ISA-Library, with regard to various block sizes and concurrency. We conclude that the MLX cards encoded faster and more consistently than ISA-L. Furthermore, the ConnectX-5 cards support the Galois Field (GF) 2^8 , this grants compatibility with data encoded by any system using GF(2^8) or less, including ISA-L. These cards enable users to substantially increase encoding and decoding throughput by using more cards; additionally enabling the use of less powerful CPUs to achieve similar performance. LA-UR-17-26333

P55: Incorporating Proactive Data Rescue into ZFS Disk Recovery for Enhanced Storage Reliability

Authors: Zhi Qiao (University of North Texas), Song Fu (University of North Texas), Hsing-bung Chen (Los Alamos National Laboratory), Michael Lang (Los Alamos National Laboratory)

As tremendous amount of data are generated every day, storage systems store exabytes of data on hundreds of thousands of disk drives. At such a scale, disk failures become the norm. Data recovery takes longer time due to increased disk capacity. ZFS is a widely used filesystem, providing data recovery from corruption. Many factors may affect ZFS's recovery performance in a production environment. Additionally, disk failure prediction techniques enables ZFS to proactively rescue data prior to disk failures. In this poster, we extensively evaluate the recovery performance with a variety of ZFS configurations. We also compare the performance of different data rescue strategies, including post-failure disk recovery, proactive disk cloning, and proactive data recovery. Our proposed analytic model uses the collected zpool utilization data and system configuration to derive the optimal data rescue strategy that best suits the storage array in the current state.

P56: ZoneTier: A Zone-Based Storage Tiering and Caching Co-Design to Integrate SSDs with Host-Aware SMR Drives

Authors: Xuchao Xie (National University of Defense Technology), Liquan Xiao (National University of Defense Technology), David H.C. Du (University of Minnesota)

Integrating solid state drives (SSDs) and host-aware shingled magnetic recording (HA-SMR) drives can potentially build a cost-effective high-performance storage system. However, existing SSD tiering and caching designs in such a hybrid system are not fully matched with the intrinsic properties of HA-SMR drives due to their handling of non-sequential writes (NSWs) from both workloads and data migration. We propose ZoneTier, a zone-based storage tiering and caching co-design, to effectively control all the NSWs by leveraging the host-aware property of HA-SMR drives. ZoneTier exploits the real-time data layouts of HA-SMR zones to optimize zone placements, reshape NSWs generated from zone demotions to HA-SMR drive preferred sequential writes, and transforms the inevitable NSWs to HA-SMR zones to cleaning-friendly write traffics. Our experiments show that ZoneTier can utilize SSDs with high efficiency, minimize relocation overhead, shorten performance recovery time of HA-SMR drives, and finally accomplish better system performance than existing hybrid storage designs.

P57: Adaptive Tier Selection for NetCDF and HDF5

Authors: Jakob Luettgau (German Climate Computing Center), Eugen Betke (German Climate Computing Center), Olga Perevalova (University of Hamburg), Julian Kunkel (German Climate Computing Center), Michael Kuhn (University of Hamburg)

Scientific applications on supercomputers tend to be I/O-intensive. To achieve portability and performance, data description libraries such as HDF5 and NetCDF are commonly used. Unfortunately, the libraries often default to suboptimal access patterns for reading/writing data to multi-tier distributed storage. This work explores the feasibility of adaptively selecting tiers depending on an application's I/O behavior.

P58: Wharf: Sharing Docker Images across Hosts from a Distributed Filesystem

Authors: Chao Zheng (University of Notre Dame), Lukas Rupperecht (IBM), Vasily Tarasov (IBM), Mohamed Mohamed (IBM), Dimitrios Skourtis (IBM), Amit S. Warke (IBM), Dean Hildebrand (IBM), Douglas Thain (University of Notre Dame)

Due to their portability and less overhead compared to traditional virtual machines, containers are becoming an attractive solution for running HPC workloads. Docker is a popular toolset which enables convenient provisioning and management of containers and their corresponding images. However, Docker does not natively support running on shared storage, a crucial requirement in large-scale HPC clusters which are often diskless or access data via a shared burst buffer layer. This lack of distributed storage support can lead to overhead when running containerized HPC applications. In this work, we explore how Docker images can be served efficiently from a shared distributed storage layer. We implement a distributed layer on top of Docker that allows multiple Docker daemons to access container images from a shared filesystem such as IBM Spectrum Scale or NFS. Our design is independent of the underlying storage layer and minimizes the synchronization overhead between different daemons.

P59: Secure Enclaves: An Isolation-Centric Approach for Creating Secure High-Performance Computing Environments

Authors: Ferrol Aderholdt (Oak Ridge National Laboratory), Susan Hicks (Oak Ridge National Laboratory), Thomas Naughton (Oak Ridge National Laboratory), Lawrence Sorrillo (Oak Ridge National Laboratory), Blake Caldwell (University of Colorado, Boulder), James Pogge (Tennessee Technological University), Stephen L. Scott (Tennessee Technological University)

High performance computing environments are used for a wide variety of workloads. These systems may process data at various security levels but in so doing are often enclaved at the highest security posture, which may limit usability or performance. The traditional approach used to provide isolation is effective at the creation of secure enclaves, but poses significant challenges with respect to the use of shared infrastructure in HPC environments. We evaluate the use of system-level (i.e., hypervisor-based) and operating system level (i.e., containers) virtualization as well as software defined networking (SDN) as possible mechanisms for secure, isolation-centric enclaves (secure enclaves). We describe our approach to secure HPC enclaves and provide benchmark results for three focus areas (compute, network and data storage) where isolation mechanisms are most significant.

P60: Managing dbGaP Data with Stratus, a Research Cloud for Protected Data

Authors: Evan F. Bollig (University of Minnesota), Graham Allan (University of Minnesota), Benjamin J. Lynch (University of Minnesota), Yectli Huerta (University of Minnesota), Mathew Mix (University of Minnesota), Brent Swartz (University of Minnesota), Edward A. Munsell (University of Minnesota), Joshua Leibfried (University of Minnesota), Naomi Hospodarsky (University of Minnesota)

Modern research computing needs at academic institutions are evolving. While traditional HPC continues to satisfy most workflows, a new generation of researcher has emerged looking for sophisticated, self-service control of compute infrastructure in a cloud-like environment. Often, these demands are not for their own interest, but nonetheless present due to constraints imposed by data governance and protection policies that cannot be satisfied by traditional HPC.

To cater to these modern users, the Minnesota Supercomputing Institute deployed a cloud service for research computing called Stratus. Stratus is designed expressly to satisfy the requirements set forth by the NIH Genomic Data Sharing (GDS) Policy for dbGaP data. It is powered by the Newton version of OpenStack, and backed by Ceph storage. The service offers three features not available on traditional HPC systems: a) on-demand availability of compute resources; b) long-running jobs (i.e., > 30 days); and c) container-based computing with Docker applications.

P61: Cloud Resource Selection Based on PLS Method for Deploying Optimal Infrastructures for Genomic Analytics Application

Authors: Katsunori Miura (*Kitami Institute of Technology*), Courtney Powell (*Hokkaido University*), Masaharu Munetomo (*Hokkaido University*)

This poster proposes a method for determining infrastructures composed of cloud resources that concurrently meet satisfiability and optimality system requirements, such as computational performance, maximum price payable, and deployment location of a genomic analytics application. The input to the proposed method is a mathematical formula that captures the system requirements given by the user, which is defined in accordance with first-order predicate logic, whereas the output is a set of unit clauses representing infrastructures for deploying the genomic analytics application. In the proposed method, an equivalent transformation algorithm is used to generate valid solutions with respect to system requirements, and a genetic algorithm is used to evaluate the optimality of the solutions.

P62: How To Do Machine Learning on Big Clusters

Authors: Thomas Ashby (*IMEC*), Tom Vander Aa (*IMEC*), Stanislav Bohm (*Technical University of Ostrava*), Vojtech Cima (*Technical University of Ostrava*), Jan Martinovic (*Technical University of Ostrava*), Vladimir Chupakhin (*Janssen Global Services LLC*)

Scientific pipelines, such those in chemogenomics machine learning applications, often compose of multiple interdependent data processing tasks. We are developing HyperLoom - a platform for defining and executing workflow pipelines in large-scale distributed environments. HyperLoom users can easily define dependencies between computational tasks and create a pipeline which can then be executed on HPC systems. The high-performance core of HyperLoom dynamically orchestrates the tasks over available resources respecting task requirements. The entire system was designed to have a minimal overhead and to efficiently deal with varying computational times of the tasks. HyperLoom allows to execute pipelines that contain basic built-in tasks, user-defined Python tasks, tasks wrapping third-party applications or a combination of those.

BP

P63: FleCSPH: a Parallel and Distributed Smoothed Particle Hydrodynamics Framework Based on FleCSI

Authors: Julien Loiseau (*University of Reims Champagne-Ardenne*), Hyun Lim (*Brigham Young University*), Ben Bergen (*Los Alamos National Laboratory*), Nicholas Moss (*Los Alamos National Laboratory*)

In this poster session, we introduce our new parallel and distributed Smoothed Particle Hydrodynamics implementation, FleCSPH, which is based on the open-source framework FleCSI. This framework provides us the basic data structures and runtime required in our work. We intend to provide a parallel and distributed version of Binary, Quad, and Oct trees data structure, respectively for 1, 2 and 3 dimensions dedicated for SPH problems.

Also, we present various test problems in several dimensions that indicate the flexibility and the scalability of our toolkit. For application and tests we simulate classical physics test cases like Sod Shock Tube, Sedov Blast Wave or Fluid Flows. The aim of this work is to simulate binary compact object mergers such as white dwarfs and neutron stars that address many interesting astrophysical phenomena in the universe.

P64: romeoLAB : HPC Training Platform on HPC facility

Authors: Jean-Matthieu Etancelin (*University of Reims Champagne-Ardenne*), Arnaud Renard (*University of Reims Champagne-Ardenne*)

In this pre-exascale era, we are observing a dramatic increase of the necessity of computer science courses dedicated to parallel programming with advanced technologies on hybrid architectures. The full hybrid cluster Romeo has long been used for that purpose in order to train master students and cluster users. The main issue for trainees is the cost of accessing and exploiting a production facility in a pedagogic context. The use of some specific techniques and software (SSH, workload manager, remote file system, ...) is mandatory without being part of courses prerequisites nor pedagogic objectives. The romeoLAB platform we developed at ROMEO HPC Center is an online interactive pedagogic platform for HPC technologies courses. Its main purpose is to simplify the process of resource usage in order to focus on the taught subjects. This paper presents the context, the romeoLAB architecture and its motivations, usages and pedagogical contents.

P65: CAPES: Unsupervised System Performance Tuning Using Neural Network-Based Deep Reinforcement Learning

Authors: Yan Li (*University of California, Santa Cruz*), Kenneth Chang (*University of California, Santa Cruz*), Oceane Bel (*University of California, Santa Cruz*), Ethan Miller (*University of California, Santa Cruz*), Darrell Long (*University of California, Santa Cruz*)

Parameter tuning is an important task of storage performance optimization. Current practice usually involves numerous tweak-benchmark cycles that are slow and costly. To address this issue, we developed CAPES, a model-less deep reinforcement learning-based unsupervised parameter tuning system driven by a deep neural network (DNN). It is designed to find optimal values for computer systems that have tunable parameters when human tuning can be costly and often cannot achieve optimal performance. CAPES takes periodic measurements of a target computer system's state, and trains a DNN which uses Q-learning to suggest changes to the system's current parameter values. CAPES is minimally intrusive, and can be deployed into a production system to collect training data and suggest tuning actions during the system's daily operation. Evaluation of a prototype on a Lustre file system demonstrates an increase in I/O throughput up to 45% at saturation point.

P66: Analyzing Multi-Layer I/O Behavior of HPC Applications

Authors: Ronny Tschüter (Technical University Dresden), Christian Herold (Technical University Dresden), Bert Wesarg (Technical University Dresden), Matthias Weber (Technical University Dresden)

In this work, we present an approach to analyze multi-layer I/O behavior of HPC applications. For a comprehensive analysis it is not sufficient to track I/O operations on a single layer because applications may combine multiple I/O paradigms (e.g., MPI I/O, NetCDF, HDF5). Furthermore, I/O libraries may use another I/O paradigm internally. With our approach, I/O activities can be captured at any layer. This work introduces methodologies to intercept calls to I/O libraries and presents an event design to record applications' I/O activities. We implement our approach in the Score-P measurement system and prove its usability with a study of the Met Office/NERC Cloud Model (MONC) code.

P67: Measuring I/O Behavior on Upcoming Systems with NVRAM

Authors: Christian Herold (Technical University Dresden), Ronny Tschüter (Technical University Dresden)

Upcoming HPC systems will use NVRAM to address existing I/O bottlenecks. The I/O performance thereby is one of the keys for the exascale challenges. NVRAM introduces a new level in the memory hierarchy and can be utilized by different technologies, e.g. memory mapped files or block transfer operations. Using NVRAM without considering the complete hierarchy may lead to an inefficient usage and bad I/O performance. Therefore, in this work, we evaluate techniques for measuring the I/O behavior of applications that utilize NVRAM in various use cases. We extended the application tracing of our tool Score-P and introduced several metrics as well as events for different use cases of NVRAM.

P68: Continuous Clock Synchronization for Accurate Performance Measurements

Authors: Johannes Ziegenbalg (Technical University Dresden), Matthias Weber (Technical University Dresden)

Clock synchronization is a key prerequisite for accurate performance analysis. This is particularly true for most HPC systems, due to the lack of a system-wide timer. The effects of different clock properties lead to false measurements and wrong conclusions. Often these errors remain undetected by the user. Different environmental effects cause continuous changes to clock properties over time. This behavior is ignored by the usual post-mortem approach for clock synchronization. In order to improve time measurement accuracy we implemented a method for continuous clock synchronization. In this work, we share our experiences and draw conclusions for tool development.

P69: Portable Methods for Measuring Cache Hierarchy Performance

Authors: Tom Deakin (University of Bristol), James Price (University of Bristol), Simon McIntosh-Smith (University of Bristol)

There has been a recent influx of different processor architecture designs into the market, with many of them targeting HPC applications. When estimating application performance, developers are used to considering the most common figures of merit, such as peak FLOP/s, memory bandwidth, core counts, and so on. In this study, we present a detailed comparison of on-chip memory bandwidths, including single core and aggregate across a node, for a set of next-generation CPUs.

We do this in such a way as to be portable across different architectures and instruction sets. Our study indicates that, while two processors might look superficially similar when only considering the common figures of merit, those two processors might have radically different on-chip memory bandwidth, a fact which may be crucial when understanding observed application performance. Our results and methods will be made available on GitHub to aid the community in evaluating cache bandwidths.

P70: FFT, FMM, and Multigrid on the Road to Exascale: Performance Challenges and Opportunities

Authors: Huda Ibeid (University of Illinois), Luke Olson (University of Illinois), William Gropp (University of Illinois)

FFT, FMM, and multigrid methods are widely used fast and highly scalable solvers for elliptic PDEs. However, emerging systems are introducing challenges in comparison to current petascale computers. The International Exascale Software Project Roadmap identifies several constraints on the design of exascale software. Challenges include massive concurrency, energy efficiency, resilience management, exploiting the high performance of heterogeneous systems, and utilizing the deeper and more complex memory hierarchy expected at exascale. In this study, we perform model-based comparison of the FFT, FMM, and multigrid methods in the context of these constraints and use the performance models to offer predictions about the methods performance on possible exascale system configurations, based on current technology trends.

P71: Is ARM Software Ecosystem Ready for HPC?

Authors: Fabio Banchelli (Barcelona Supercomputing Center), Daniel Ruiz (Barcelona Supercomputing Center), Ying Hao Xu Lin (Barcelona Supercomputing Center), Filippo Mantovani (Barcelona Supercomputing Center)

In recent years, the HPC community has increasingly grown its interest towards the ARM architecture with research projects targeting primarily the deployment of ARM-based clusters. Attention is usually given to hardware platforms, however the availability of a mature software ecosystem and the possibility of running large and complex HPC applications plays a key role in the consolidation process of a new technology.

For this reason in this poster we present a preliminary evaluation of the ARM system software ecosystem, limited here to the ARM HPC Compiler and the ARM Performance Libraries, together with a porting and testing of three fairly complex HPC code suites: QuantumESPRESSO, WRF, and FEniCS.

These codes have been proposed as HPC challenges during the last two editions of the Student Cluster Competition at ISC where all the authors have been involved operating an ARM-based cluster and awarded with the Fan Favorite award.

P72: New Developments for PAPI 5.6+

Authors: *Anthony Danalis (University of Tennessee), Heike Jagode (University of Tennessee), Vince Weaver (University of Maine), Yan Liu (University of Maine), Jack Dongarra (University of Tennessee)*

The HPC community has relied on PAPI to track low-level hardware operations for over 15 years. In that time, the needs of software developers have changed immensely, and the PAPI team aims to meet these demands through a better understanding of deep and heterogeneous memory hierarchies and finer-grain power-management support.

This poster demonstrates how PAPI enables power-tuning to reduce overall energy consumption without, in many cases, a loss in performance. Furthermore, we discuss efforts to develop microbenchmarks intended to assist application developers who are interested in performance analysis by automatically categorizing and disambiguating performance counters. Finally, the poster illustrates efforts to update PAPI's internal sanity checks, designed to inspect that PAPI's predefined events are in fact measuring the values they claim to measure, and modernize the implementation of critical API functions, e.g., `PAPI_read()`, and the sampling interface so that more information can be captured and reported with lower overhead.

BP

P73: HPC Production Job Quality Assessment

Authors: *Omar Aaziz (New Mexico State University), Jonathan Cook (New Mexico State University)*

Users of HPC systems would benefit from more feedback about the quality of their application runs, such as knowing whether or not the performance of a particular run was good, or whether the resources requested were enough, or too much. Such feedback requires more information to be kept regarding production application runs, and requires some analytics to assess any new runs. In this research, we assess the practicality of using job data, system data, and hardware performance counters in a near-zero overhead manner to assess job performance, in particular whether or not the job runtime was in line with expectations from historical application performance. We show over four proxy applications and two real application that our assessment is within 10% of actual performance.

P74: A Methodology for Bridging the Native and Simulated Executions of Parallel Applications

Authors: *Ali Mohammed (University of Basel), Ahmed Eleliemy (University of Basel), Florina M. Ciorba (University of Basel)*

Simulation is considered as the third pillar of science, following experimentation and theory. Bridging the native and simulated executions of parallel applications is needed for attaining trustworthiness in simulation results. Yet, bridging the native and simulated executions of parallel applications is challenging. This work proposes a methodology for bridging the native and simulated executions of message passing parallel applications on high performance computing (HPC) systems in two steps: Expression of the software characteristics, and representation and verification of the hardware characteristics in the simulation. This work exploits the capabilities of the SimGrid [3] simulation toolkit's interfaces to reduce the effort of bridging the native and simulated executions of a parallel application on an HPC system. For an application from computer vision, the simulation of its parallel execution using straightforward parallelization on an HPC cluster approaches the native performance with a minimum relative percentage difference of 5.6%.

P75: Model-Agnostic Influence Analysis for Performance Data

Authors: *Rahul Sridhar (University of California, Irvine; Lawrence Livermore National Laboratory), Rushil Anirudh (Lawrence Livermore National Laboratory), Jayaraman J. Thiagarajan (Lawrence Livermore National Laboratory), Nikhil Jain (Lawrence Livermore National Laboratory), Todd Gamblin (Lawrence Livermore National Laboratory)*

Execution time of an application is affected by several performance parameters, e.g. number of threads, decomposition, etc. Hence, an important problem in high performance computing is to study the influence of these parameters on the performance of an application. Additionally, quantifying the influence of individual parameter configurations (data samples) on performance also aids in identifying sub-domains of interest in high-dimensional parameter spaces. Conventionally, such analysis is performed using a surrogate model, which introduces its own bias that is often non-trivial to undo, leading to inaccurate results. In this work, we propose an entirely data-driven, model-agnostic influence analysis approach based on recent advances in analyzing functions on graphs. We show that the problem of identifying influential parameters (features) and configurations (samples) can be effectively addressed within this framework.

P76: A Compiler Agnostic and Architecture Aware Predictive Modeling Framework for Kernels

Authors: *William Killian (University of Delaware), Ian Karlin (Lawrence Livermore National Laboratory), David Beckingsale (Lawrence Livermore National Laboratory), John Cavazos (University of Delaware)*

Multi-architecture machines make program characterization for modeling a regression outcome difficult. Determining where to offload compute-dense portions requires accurate prediction models for multiple architectures. To productively achieve portable performance across these diverse architectures, users are adopting portable programming models such as OpenMP and RAJA.

When adopted, portable models make traditional high-level source code analysis inadequate for program characterization. In this poster, we introduce a common microarchitecture instruction format (ComIL) for program characterization. ComIL is capable of representing programs in an architecture-aware compiler-agnostic manner. We evaluate feature extraction with ComIL by constructing multiple regression-objective models for performance (execution time) and correctness (maximum absolute error). These models perform better than the current state of the art -- achieving a mean error rate of only 4.7% when predicting execution time. We plan to extend this work to handle multiple architectures concurrently and evaluate with more representative physics kernels.

P77: AutoTuneTMP: Auto Tuning in C++ With Runtime Template Metaprogramming

Authors: David Pfander (University of Stuttgart), Dirk Pflüger (University of Stuttgart)

Maximizing the performance on modern hardware platforms has become more and more difficult, due to different levels of parallelism and complicated memory hierarchies. Auto tuning can help developers to address these challenges by writing code that automatically adjusts to the underlying hardware platform. AutoTuneTMP is a new C++-based auto tuning framework that uses just-in-time compilation to enable runtime-instantiable C++ templates. We use C++ template metaprogramming to provide data structures and algorithms that can be used to develop tunable compute kernels. These compute kernels can be tuned with different optimization strategies. We demonstrate for a first prototype the applicability and usefulness of our approach by tuning 6 parameters of a DGEMM implementation, achieving 68% peak performance on an Intel Skylake processor.

P78: Performance Evaluation of Graph500 Considering CPU-DRAM Power Shifting

Authors: Yuta Kakibuka (Kyushu University), Yuichiro Yasui (Kyushu University), Takatsugu Ono (Kyushu University), Katsuki Fujisawa (Kyushu University), Koji Inoue (Kyushu University)

There are power constraints on computer systems which comes from technical, costly or social demands. Power wall is one of the most serious issues for post petascale high-performance computing. A promising solution to tackle this problem is to effectively manage power resources based on the characteristics of workloads. In power constrained computing, the key is to translate the limited power budget into sustained performance effectively. To achieve this goal, assigning the appropriate amount of power budget to each hardware component, or power shifting, is a critical challenge.

In this work, we focus on large-scale graph processing. Graph analysis algorithms are increasing its importance with growing demands of big data analysis. However, the impact of power constraint on the performance of graph processing application is not declared. Our work is the performance evaluation of Graph500 under power constraints to CPU and DRAM.

P79: Porting the Opacity Client Library to a CPU-GPU Cluster Using OpenMP 4.5

Authors: Jason S. Kimko (College of William and Mary), Michael M. Pozulp (Lawrence Livermore National Laboratory), Riyaz Haque (Lawrence Livermore National Laboratory), Leopold Grinberg (IBM)

The poster accompanying this summary exhibits our experience porting the Opacity client library to IBM's "Minsky" nodes using OpenMP 4.5. We constructed a GPU-friendly container class that mimics existing library functionality. We benchmarked our implementation on Lawrence Livermore National Laboratory's (LLNL) RZManta, a Minsky cluster. In our benchmarks on a single POWER8 CPU and Tesla P100 GPU, we observed up to a 4x speedup including CPU-GPU data transfers and up to a 30x speedup excluding data transfers. Optimizing to reduce register pressure and increase occupancy may improve speedups. Our results demonstrate a successful and beneficial library port to the CPU-GPU architecture.

P80: Adaptive Loop Scheduling with Charm++ to Improve Performance of Scientific Applications

Authors: Vivek Kale (University of Southern California), Harshitha Menon (Lawrence Livermore National Laboratory), Karthik Senthil (University of Illinois)

Supercomputers today employ a large number of cores on each node. The Charm++ parallel programming system provides an intelligent runtime which has been highly effective at providing dynamic load balancing across nodes of a supercomputer. Modern multi-core nodes present new challenges and opportunities for Charm++. The large degree of over-decomposition required may lead to high overhead. We modified the Charm++ Runtime System (RTS) to assign Charm++ objects to nodes, thus reducing over-decomposition, and spreading work across cores via parallel loops. We modify a library of the Charm++ software suite that supports loop parallelism by adding to it a loop scheduling strategy that maximizes load balance across cores while minimizing data movement. We tune parameters of the RTS and the loop scheduling strategy to improve performance of benchmark codes run on a variety of architectures. Our technique improves performance of a Particle-in-Cell code run on the Blue Waters supercomputer by 17.2%.

BP

P81: Offloading Python Kernels to Micro-Core Architectures

Authors: Nick Brown (University of Edinburgh)

Micro-core architectures combine many low memory, low power computing cores together in a single package. These can be used as a co-processor or standalone but due to limited on-chip memory and esoteric nature of the hardware, writing efficient parallel codes for them is challenging. We previously developed ePython, a low memory (24Kb) implementation of Python supporting the rapid development of parallel Python codes and education for these architectures. In this poster we present our work on an offload abstraction to support the use of micro-cores as an accelerator. Programmers decorate specific functions in their Python code, running under any Python interpreter on the host, with our underlying technology then responsible for the low-level data movement, scheduling and execution of kernels on the micro-cores. Aimed at education and fast prototyping, a machine learning code for detecting lung cancer, where computational kernels are offloaded to micro-cores, is used to illustrate the approach.

P82: Performance Evaluation of the NVIDIA Tesla P100: Our Directive-Based Partitioning and Pipelining vs. NVIDIA's Unified Memory

Authors: *Xuewen Cui (Virginia Tech), Thomas R. W. Scogland (Lawrence Livermore National Laboratory), Bronis R. de Supinski (Lawrence Livermore National Laboratory), Wu-chun Feng (Virginia Tech)*

We need simpler mechanisms to leverage the performance of accelerators, such as GPUs, in supercomputers. Programming models like OpenMP offer simple-to-use but powerful directive-based offload mechanisms. By default, these models naively copy data to or from the device without overlapping computation. Achieving performance can require extensive hand-tuning to apply optimizations such as pipelining. Users must manually partition data whenever it exceeds device memory. Our directive-based partitioning and pipelining extension for accelerators overlaps data transfers and kernel computation without explicit user data-splitting. We compare a prototype implementation of our extension to NVIDIA's Unified Memory on the Pascal P100 GPU and find that our extension outperforms Unified Memory on average by 68% for data sets that fit into GPU memory and 550% for those that do not.

P83: Contracts for Message-Passing Programs

Authors: *Ziqing Luo (University of Delaware), Stephen F. Siegel (University of Delaware)*

Verification of message-passing parallel programs is challenging because of the state-explosion problem. A procedure-level contract system for parallel programs can help overcome this challenge by verifying contracts individually, in a modular way. A contract system is used to specify intended behaviors of programs. Contracts can be checked at run-time or verified formally. There is a mature theory of contracts for sequential programs, but little work has been done on parallel programs, and even less for message-passing parallel programs. We developed a theory of contracts for message-passing programs and realize this theory in a contract language for programs that use the Message Passing Interface (MPI). We are also developing a verification tool that uses symbolic execution and model checking techniques to prove that MPI programs satisfy their contracts.

P84: PRESAGE: Selective Low Overhead Error Amplification for Easy Detection

Authors: *Vishal Chandra Sharma (University of Utah), Arnab Das (University of Utah), Ian Briggs (University of Utah), Ganesh Gopalakrishnan (University of Utah), Sriram Krishnamoorthy (Pacific Northwest National Laboratory)*

Soft-errors remain a vexing challenge. Today's error detectors either come with high false positives or high omissions. Our work focuses on not losing an error sown, but amplifying it so that cheap detection is enabled.

Consider structured address generation in loops where data from a base address plus offset is used. An erroneous offset no longer crashes today's applications thanks to large address spaces; instead, it silently corrupts data (unintended fetch). We relativize address generation using LLVM, thus each new address is not base plus offset but previous relative address plus offset. If one address is corrupted, all future address are corrupted in a chain. This permits efficient loop exit-point detection.

Relativization has low overhead, actually lowered by some ISAs down to zero. These advantages survive crucial compiler memory access optimizations. We demonstrate 100% SDC detection for a class of benchmarks with respect to structured address protection.

P85: GPU Mekong: Simplified Multi-GPU Programming Using Automated Partitioning

Authors: *Alexander Matz (University of Heidelberg)*

GPU accelerators are pervasively used in the HPC community, because they provide excellent computational performance at a reasonable power efficiency. While programming single-GPU applications is comparatively productive, programming multiple GPUs using data-parallel languages is tedious and error prone as the user has to manually orchestrate data movements and kernel launches.

The Mekong research project is driven by the motivation to improve productivity of multi-GPU systems by compiler based partitioning of single-device data-parallel programs. Key to scalable performance improvement is the resolution of data dependencies between kernels and the orchestration of these kernels. Mekong relies on polyhedral compilation to identify memory access patterns in order to compile a single-GPU application into a multi-GPU application.

In this work, the Mekong project is introduced and its components explained. While the tool is still under development, preliminary results are available and are shortly discussed demonstrating the potential of this approach.

P86: HyGraph: High Performance Graph Processing on Hybrid CPU+GPUs platforms

Authors: *Stijn Heldens (University of Twente), Ana Lucia Varbanescu (University of Amsterdam), Alexandru Iosup (Vrije University Amsterdam)*

Graph analytics is becoming increasingly important in many domains, such as in biology, social sciences, and data mining. Many large-scale graph-processing systems have been proposed, either targeting distributed clusters or GPU-based accelerated platforms. However, little research exists on designing systems for hybrid CPU-GPU platforms, i.e., exploiting both the CPU and the GPU efficiently.

In this work, we present HyGraph, a novel graph-processing system for hybrid platforms which delivers performance by using both the CPU and GPUs concurrently. Its core feature is dynamic scheduling of tasks onto both the CPU and the GPUs, thus providing load balancing, contrary to the state-of-the-art approach based on static partitioning. Additionally, communication overhead is minimized by overlapping computation and communication.

Our results demonstrate that HyGraph outperforms CPU-only and GPU-only solutions, delivering close-to-optimal performance. Moreover, it supports large-scale graphs which do not fit into GPU memory and is competitive against state-of-the-art systems.

P87: EoCoE Performance Benchmarking Methodology for Renewable Energy Applications

Authors: *Paul Gibbon (Forschungszentrum Juelich), Matthieu Haefele (French Alternative Energies and Atomic Energy Commission), Sebastian Luehrs (Forschungszentrum Juelich, Juelich Supercomputing Center)*

An optimisation strategy developed by the Energy Oriented Centre of Excellence (EoCoE) is presented for computational models used in a variety of renewable energy domains. It is found that typical applications in this comparatively new sector exhibit a huge range of HPC maturity, from simple parallelization needs to near-exascale readiness. An essential part of this process has therefore been the introduction of a flexible, quantitative performance assessment of applications using the benchmarking tool JUBE to automatically extract up to 28 different metrics taken with several state-of-the-art performance tools. An initial hands-on workshop to establish this baseline status is consolidated by follow-up actions by joint code-teams comprising members of both developer groups and HPC centres involved with the EoCoE consortium. Examples of early successes achieved with this policy are given, together with an outlook on challenges faced for energy applications with next-generation, pre-exascale architectures.

P88: PetaVision Neural Simulation Toolbox on Intel KNLs

Authors: *Boram Yoon (Los Alamos National Laboratory), Pete Schultz (Los Alamos National Laboratory, New Mexico Consortium), Garrett Kenyon (Los Alamos National Laboratory, New Mexico Consortium)*

We implemented a large-scale neuromorphic algorithm called the Sparse Prediction Machine (SPM), on the Los Alamos Trinity supercomputer. Specifically, we used PetaVision, an open source high-performance neural simulation toolbox, to implement a 4-layer SPM applied to ImageNet video. Various optimization techniques were applied to efficiently utilize up to 8192 KNL nodes. The goal of the SPM is to predict future states of a system from a sequence of previous states, or in the case of video, to predict a subsequent frame from previous frames. In our training, the SPM was able to predict the 8th frame from the preceding 7 frames, including successful separation of foreground and background motion.

P89: Desh: Deep Learning for HPC System Health Resilience

Authors: *Anwesha Das (North Carolina State University), Abhinav Vishnu (Pacific Northwest National Laboratory), Charles Siegel (Pacific Northwest National Laboratory), Frank Mueller (North Carolina State University)*

HPC systems are well known to endure service downtime due to increasing failures. With enhancements in HPC architectures and design, enabling resilience is extremely challenging due to component scaling and absence of well defined failure indicators. HPC system logs are notorious to be complex and unstructured. Efficient fault prediction to enable proactive recovery mechanisms is the need of the hour to make such systems more robust and reliable. This work addresses such faults in computing systems using a recurrent neural network based technique called LSTM (long short-term memory).

We present our framework Desh : Deep Learning for HPC System Health, which entails a procedure to diagnose and predict failures with acceptable lead times. Desh indicates prospects of indicating failure indicators with enhanced training and classification for generic applicability to other systems. This deep learning based framework gives interesting insights for further work on HPC system reliability.

P90: Global Survey of Energy and Power-Aware Job Scheduling and Resource Management in Supercomputing Centers

Authors: *Siddhartha Jana (Intel Corporation), Gregory A. Koenig (Energy Efficient HPC Working Group), Matthias Maiterth (Intel Corporation), Kevin T. Pedretti (Sandia National Laboratories), Andrea Borghesi (University of Bologna), Andrea Bartolini (ETH Zurich), Bilel Hadri (King Abdullah University of Science and Technology), Natalie J. Bates (Energy Efficient HPC Working Group)*

The two major driving forces that are leading centers to investigate dynamic power and energy management strategies are (1) limitations to the availability of resources from the electricity service provider, and (2) the desire to spend limited budgets on computational cycles rather than infrastructure requirements such as electricity. In addition, supercomputer systems have increasingly rapid, unpredictable, and large power fluctuations. In addition, electricity service providers may request supercomputing

centers to change their timing and/or magnitude of demand to help address electricity supply constraints. To adapt to this new landscape, centers may employ energy and power-aware job scheduling and resource management (EPA-JSRM) strategies to dynamically, and in real-time, control their electricity demand. This poster summarizes the lessons learned from one of the first global surveys of supercomputing centers that are actively investigating such strategies.

P91: Assessing the Availability of Source Code in Computational Physics

Authors: *Matthew Krafczyk (National Center for Supercomputing Applications, University of Illinois), Victoria Stodden (University of Illinois), Yantong Zheng (University of Illinois), David Wong (University of Illinois)*

Replicability of scientific work based on computation is a subject which has been receiving increased scrutiny recently. One approach to replicating a computational finding is to run the original source code. Availability of source code however is not routine; Only 3/33 computationally based article authors released source code from JASA in 2006, and a 2015 study showed that only 44% of computer science article authors released their source code. The field of Computational Physics has yet to be examined in this way, nor has the effect of author knowledge of such a study been measured.

We present our findings regarding the availability of source code in recent articles of the Journal of Computational Physics as well as how author knowledge of the study affects their willingness to make source code available. This work extends current reproducibility efforts being explored by the ACM, SIGHPC, and the SC conference community.

P92: Characterization and Comparison of Application Resilience for Serial and Parallel Executions

Authors: *Kai Wu (University of California, Merced), Qiang Guan (Los Alamos National Laboratory, Ultrascale Systems Research Center), Nathan DeBardeleben (Los Alamos National Laboratory, Ultrascale Systems Research Center), Dong Li (University of California, Merced)*

Soft error of exascale application is a challenge problem in modern HPC. In order to quantify an application's resilience and vulnerability, the application-level fault injection method is widely adopted by HPC users. However, it is not easy since users need to inject a large number of faults to ensure statistical significance, especially for parallel version program. Normally, parallel execution is more complex and requires more hardware resources than its serial execution. Therefore, it is essential that we can predict error rate of parallel application based on its corresponding serial version. In this poster, we characterize fault pattern in serial and parallel executions. We find first there are same fault sources in serial and parallel execution. Second, parallel execution also has some unique fault sources compared with serial executions. Those unique fault sources are important for us to understand the difference of fault pattern between serial and parallel executions.

P93: Spacehog: Evaluating the Costs of Dedicated Resources to In Situ Analysis

Authors: *Rebecca Kreitinger (University of New Mexico), Scott Levy (Sandia National Laboratories), Kurt B. Ferreira (Sandia National Laboratories), Patrick Widener (Sandia National Laboratories)*

Using in situ analytics requires that computational resources be shared between the simulation and the analysis. With space-sharing, there is a possibility for contention over these shared resources such as memory, memory bandwidth, network bandwidth, or filesystem bandwidth. In our analysis, we explore the sensitivity of different applications with a set of microbenchmarks that are representative of analytics that may be used with scientific simulation. These tasks are modeled using a library called libspacehog. The experimentation consisted of examining three different dimensions of how simulation workloads might be space-shared with analysis codes. The results indicate that contention does need to be considered when applying in situ analytic techniques and can be of greater concern than simply the number of analysis processes or overall process density. This research provides an explanation on how the application's performance is affected by space-sharing to further understand in situ analytic techniques.

P94: Fully Hierarchical Scheduling: Paving the Way to Exascale Workloads

Authors: *Stephen Herbein (University of Delaware), Tapasya Patki (Lawrence Livermore National Laboratory), Dong H. Ahn (Lawrence Livermore National Laboratory), Don Lipari (Lawrence Livermore National Laboratory), Tamara Dahlgren (Lawrence Livermore National Laboratory), David Domyancic (Lawrence Livermore National Laboratory), Michela Taufer (University of Delaware)*

Exascale workloads, such as uncertainty quantification (UQ), represent an order of magnitude increase in both scheduling scale and complexity. Batch schedulers with their decades-old, centralized scheduling model will fail to address the needs of these new workloads. To address these upcoming challenges, we claim that HPC schedulers must transition from the centralized to the fully hierarchical scheduling model. In this work, we assess the impact of the fully hierarchical model on both a synthetic stress test and a real-world UQ workload. We observe over a 100x increase in scheduler scalability on the synthetic stress test and a 37% decrease in the runtime of real-world UQ workloads under the fully hierarchical model. Our empirical results demonstrate that the fully hierarchical scheduling model can overcome the limitations of existing schedulers to meet the needs of UQ and other exascale workloads.

BP

P95: GEOPM: A Scalable Open Runtime Framework for Power Management

Authors: *Siddhartha Jana (Intel Corporation), Asma H. Al-rawi (Intel Corporation), Steve S. Sylvester (Intel Corporation), Christopher M. Cantalupo (Intel Corporation), Brad Geltz (Intel Corporation), Brandon Baker (Intel Corporation), Jonathan M. Eastep (Intel Corporation)*

The power scaling challenges associated with exascale systems is a well-known issue. In this work, we introduce the Global Extensible Open Power Manager (GEOPM): a tree-hierarchical, open source runtime framework we are contributing to the HPC community to foster increased collaboration and accelerated progress toward software-hardware co-designed energy management solutions that address exascale power challenges and improve performance and energy efficiency in current systems. Through its plugin extensible architecture, GEOPM enables rapid prototyping of new energy management strategies. Different plugins can be tailored to the specific performance or energy efficiency priorities of each HPC center. To demonstrate the potential of the framework, this work develops an example plugin for GEOPM. This power rebalancing plugin targets power-capped systems and improves efficiency by minimizing job time-to-solution within a power budget. Our results demonstrate up to 30% improvements in the time-to-solution of CORAL system procurement benchmarks on a Xeon Phi cluster.

P96: Correcting Detectable Uncorrectable Errors in Memory

Authors: Grzegorz Pawelczak (*University of Bristol*), Simon McIntosh-Smith (*University of Bristol*)

With the expected decrease in Mean Time Between Failures, Fault Tolerance has been identified as one of the major challenges for exascale computing. One source of faults are soft errors caused by cosmic rays, which can cause bit corruptions to the data held in memory. Current solutions for protection against these errors include Error Correcting Codes, which can detect and/or correct these errors. When an error that can be detected but not corrected occurs, a Detectable Uncorrectable Error (DUE) results, and unless checkpoint-restart is used, the system will usually fail. In our work we present a probabilistic method of correcting DUEs which occur in the part of the memory where the program instructions are stored. We devise a correction technique for DUEs for the ARM A64 instruction set which combines extended Hamming code with Cyclic Redundancy Check code to provide near 100% Successful Correction Rate of DUEs.

P97: Profile Guided Kernel Optimization for Individual Container Execution on Bare-Metal Container

Authors: Kuniyasu Suzuki (*National Institute of Advanced Industrial Science and Technology*), Hidetaka Koie (*National Institute of Advanced Industrial Science and Technology*), Ryousei Takano (*National Institute of Advanced Industrial Science and Technology*)

Container technologies become popular on supercomputers as well as in data centers. They use a container image as a package of an application, which makes easy to customize the computing environment. Unfortunately, they are not allowed to change the kernel. It means that an application cannot get the benefit of kernel optimization. Especially, Profile Guided Kernel Optimization (PGKO) is not applied.

Bare-Metal Container (BMC) tries to solve this problem. BMC utilizes remote machine management technologies (IPMI, Intel AMT, and WakeupOnLAN) to run a container image on a remote machine with a suitable Linux kernel. It enables to use PGKO easily, because the trial execution to get a profile and the optimized execution are executed automatically. Furthermore, BMC easily changes the target machine, and the user can compare the effects. We measured the performance of PGKO on big data workloads (Apache and Redis) on Xeon and i7 and found the difference.

P98: Energy Efficiency in HPC with Machine Learning and Control Theory

Authors: Connor Imes (*University of Chicago*), Steven Hofmeyr (*Lawrence Berkeley National Laboratory*), Henry Hoffmann (*University of Chicago*)

Performance and power management in HPC has historically favored a race-to-idle approach in order to complete applications as quickly as possible, but this is not energy-efficient on modern systems. As we move toward exascale and hardware over-provisioning, power management is becoming more critical than ever for HPC system administrators, opening the door for more balanced approaches to performance and power management. We propose two projects to address balancing application performance and system power consumption in HPC during application runtime, using closed loop feedback designs based on the Self-Aware Computing Model to observe, decide, and act.

P99: The Intersection of Big Data and HPC: Using Asynchronous Many Task Runtime Systems for HPC and Big Data

Authors: Joshua Daniel Suetterlein (*Pacific Northwest National Laboratory*), Joshua Landwehr (*Trovares Inc*), Andres Marquez (*Pacific Northwest National Laboratory*), Joseph Manzano (*Pacific Northwest National Laboratory*), Kevin Barker (*Pacific Northwest National Laboratory*), Guang Gao (*University of Delaware*)

Although the primary objectives of the HPC and Big data fields seem disparate, HPC is beginning to suffer from a growing size of its workloads and the limitation of its techniques to handle large amount of data. This places interesting research challenges for both HPC and Big Data on how to marriage both fields together. This poster presents a case study which uses Asynchronous Many Task Runtimes (AMTs) as an exploratory vehicle to highlight possible solutions to these challenges. AMTs presents the unique opportunity for better load balancing, reconfigurable schedulers and data layouts that can take advantage of introspection frameworks, and the ability to exploit a massive amount of concurrency. We use the Performance Open Community Runtime (P-OCR) as a vehicle to port MapReduce operators to the HPC realm. We conduct experiments with both strong and weak scaling experimental format using WordCount and TeraSort as our kernels.

Wednesday, November 15th

Room: Four Seasons Ballroom
8:30 am - 5:00 pm

Research Posters

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

8:30 am - 5:00 pm

ACM Student Research Competition

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

Room: 701
3:00 pm - 4:45 pm

ACM Student Research Competition: Presentations by Semi-Finalists

A13: Deep Learning with HPC Simulations for Extracting Hidden Signals: Detecting Gravitational Waves

Authors: Daniel George (National Center for Supercomputing Applications, University of Illinois)

We introduce Deep Filtering, new machine learning method for end-to-end time-series signal processing, which combines two deep one-dimensional convolutional neural networks for classification and regression to detect and characterize signals much weaker than the background noise. We trained this method with a novel curriculum learning scheme on data derived from HPC simulations and applied it for gravitational wave analysis specifically for mergers of black holes and demonstrated that it significantly outperforms conventional machine learning techniques, is far more efficient than matched-filtering, offering several orders-of-magnitude speed-up, allowing real-time processing of raw big data with minimal resources, and extends the range of detectable signals. This initiates a new paradigm for scientific research which employs massively-parallel numerical simulations to train artificial intelligence algorithms that exploit emerging hardware architectures such as deep-learning-optimized GPUs. Our approach offers a unique framework to enable coincident detection campaigns of gravitational wave sources and their electromagnetic counterparts.

SF

A01: GEMM-Like Tensor-Tensor Contraction (GETT)

Authors: Paul Springer (RWTH Aachen University), Paolo Bientinesi (RWTH Aachen University)

Tensor contractions (TC) are a performance critical component in numerous scientific computations. Despite the close connection between matrix-matrix products (GEMM) and TCs, the performance of the latter is in general vastly inferior to that of an optimized GEMM. To close such a gap, we propose a novel approach: GEMM-like Tensor-Tensor multiplication (GETT). GETT mimics the design of a high-performance GEMM implementation; as such, it systematically reduces an arbitrary tensor contractions to a highly-optimized "macro-kernel". This macro-kernel operates on suitably "packed" sub-tensors that reside in specified levels of the cache hierarchy. GETT's decisive feature is its ability to pack sub-tensors via tensor transpositions, yielding efficient packing routines. In contrast to previous approaches to TCs, GETT attains the same I/O cost as an equally-sized GEMM, making GETT especially well-suited for bandwidth-bound TCs. GETT's excellent performance is highlighted across a wide range of random tensor contractions.

SF

A04: Optimization of the AIREBO Many-Body Potential for KNL

Authors: Markus Höhnerbach (RWTH Aachen University)

Molecular dynamics simulations are an indispensable research tool for computational chemistry and material science. Empirical many-body potentials promise high-fidelity simulations that capture bonding and reaction behavior accurately, providing a level of detail in between more classical molecular dynamics and quantum methods.

The AIREBO potential is one such example that provides forces and energies for molecular dynamics (MD) simulations of carbon and carbohydrate structures. Allowing many-body potentials to profit from the recent architectural advances still poses a challenge due to deeply nested, short loops. We develop an optimized, vectorized AIREBO implementation for Intel's Xeon and Xeon Phi (co)processors and integrate it into the open-source LAMMPS molecular dynamics code. By both introducing improvements to the code and vectorization, we achieve a sustained real-world speedup of two on Broadwell, and a speedup of four on KNL. The optimized code will be distributed with each LAMMPS download as part of the USER-INTEL package.

SF

A08: Virtualized Big Data: Reproducing Simulation Output on Demand

Authors: Salvatore Di Girolamo (ETH Zurich)

Scientific simulations are being pushed to the extreme in terms of size and complexity of the addressed problems, producing astonishing amount of data. If the data is stored on disk, analysis applications can randomly access simulation output. Yet, storing the massive amounts simulation data is challenging. This is primarily due to the high storage costs and the fact that compute capabilities grow faster than storage capacities and bandwidths. In-situ analysis removes the storage costs but applications lose random access.

We propose to not store the full simulation output data but to produce it on demand. Our system intercepts I/O requests of both analysis tools and simulators, enabling data virtualization. This new paradigm allows us to explore the computation-storage tradeoff, by trading computation power for storage space. Overall, SDaVi offers a viable path towards exa-scale scientific simulations, by exploiting the growing computing power and relaxing the storage capacity requirements.

SF

A02: Accelerating the Higher Order Singular Value Decomposition Algorithm for Big Data with GPUs

Authors: Yuhsiang M. Tsai (National Taiwan University)

With the explosion of big data, finding ways of compressing large datasets with multi-way relationship - i.e., tensors - quickly and efficiently has become critical in HPC.

High-order singular value decomposition (HOSVD) method provides us with the means to attain both extremely high compression ratio and low error rate through low-rank approximation.

However, parallelizing HOSVD efficiently on GPUs remains a challenging problem, largely due to the lack of a fast SVD implementation that can stream data to the limited GPU memory through the PCIe bottleneck.

Our work studies, optimizes, and then contrasts four different methods for calculating singular vectors for performance, weak/strong scalability and accuracy in the context of HOSVD. We also discuss ways of load balancing the problem across multiple GPUs on a single node, and discuss the pros and cons of these different algorithms for GPU acceleration.

SF

Break

A16: Diagnosing Parallel I/O Bottlenecks in HPC Applications

Authors: Peter Z. Harrington (University of California, Santa Cruz)

HPC applications are generating increasingly large volumes of data (up to hundreds of TBs), which need to be stored in parallel to be scalable. Parallel I/O is a significant bottleneck in HPC applications, and is especially challenging in Adaptive Mesh Refinement (AMR) applications because the structure of output files changes dynamically during runtime. Data-intensive AMR applications run on the Cori supercomputer show variable and often poor I/O performance, but diagnosing the root cause remains challenging. Here we analyze logs from multiple levels of Cori's parallel I/O subsystems, and find bottlenecks during file metadata operations and during the writing of file contents that reduced I/O bandwidth by up to 40x. Such bottlenecks seemed to be system-dependent and not the application's fault. Increasing the granularity of file-system performance data will help provide conclusive causal relationships between file-system servers and metadata bottlenecks.

SF

A11: Finding a Needle in a Field of Haystacks: Lightweight Metadata Search for Large-Scale Distributed Research Repositories

Authors: Anna Blue Keleher (University of Maryland)

Fast, scalable, and distributed search services are commonly available for single nodes, but lead to high infrastructure costs when scaled across tens of thousands of filesystems and repositories, as is the case with Globus. Endpoint-specific indexes may instead be stored on their respective nodes, but while this distributes storage costs between users, it also creates significant query overhead. Our solution provides a compromise by introducing two levels of indexes: a single centralized "second-level index" (SLI) that aggregates and summarizes terms from each endpoint; and many endpoint-level indexes that are referenced by the SLI and used only when needed. We show, via experiments on Globus-accessible filesystems, that the SLI reduces the amount of space needed on central servers by over 96% while also reducing the set of endpoints that need to execute user queries.

SF

A12: Applying Image Feature Extraction to Cluttered Scientific Repositories

Authors: Emily Herron (Mercer University)

Over time many scientific repositories and file systems become disorganized, containing poorly described and error-ridden data. As a result, it is often difficult for researchers to discover crucial data. In this poster, we present a collection of image processing modules that collectively extract metadata from a variety of image formats. We implement these modules in Skluma—a system designed to automatically extract metadata from structured and semi-structured scientific formats. Our modules apply several image metadata extraction techniques that include processing file system metadata, header information, color content statistics, extracted text, feature-based clusters, and predicting tags using a supervised learning model. Our goal is to collect a large number of

metadata that may then be used to organize, understand, and analyze data stored in a repository.
SF

A18: Understanding the Impact of Fat-Tree Network Locality on Application Performance

Authors: *Philip Taffet (Rice University)*

Network congestion can be a significant cause of performance loss and variability for many message passing programs. However, few studies have used a controlled environment with virtually no other extraneous sources of network traffic to observe the impact of application placement and multi-job interactions on overall performance. We study different placements and pairings for three DOE applications. We observe that for a job size typical for an LLNL commodity cluster, the impact of congestion and poor placement is typically less than 2%, which is less dramatic than on torus networks. In addition, in most cases, the cyclic MPI task mapping strategy increases performance and reduces placement sensitivity despite also increasing total network traffic. We also found that the performance difference between controlled placements and runs scheduled through the batch system was less than 3%.

SF

A03: A High-Speed Algorithm for Genome-Wide Association Studies on Multi-GPU Systems

Authors: *Yen Chen Chen (National Taiwan University)*

We develop an algorithm as long as a CUDA code for GWAS (Genome-Wide Associate Studies). This algorithm can work efficiently on GPU and has high scalability. The core of the algorithm is an accurate and fast p-value integration reformation, which accelerates the most time-consuming part of the algorithm. With the algorithm, researchers can now deal with tens of billions of SNP to trait pair in only a few minutes. Even better, since this algorithm is highly scalable, you can increase the problem size as long as you have enough computing power.

SF

Thursday, November 16th

Room: Four Seasons Ballroom

8:30 am - 5:00 pm

Research Posters

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

8:30 am - 5:00 pm

ACM Student Research Competition

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

Awards Presentation [\(back to top\)](#)

Wednesday, November 15th

Room: Mile High Ballroom
8:30 am - 10:00 am

ACM and IEEE-CS Award Presentations

Fernbach Award Presentation: Particles, HPC, and the Ukulele Syndrome *Steve Plimpton (Sandia National Laboratories)*
The recipient of the 2017 IEEE Computer Society Sidney Fernbach Award is Steve Plimpton from Sandia Lab.

In the first part of my talk, I'll give some vignettes of different flavors of large-scale particle simulations (molecular dynamics, kinetic Monte Carlo, direct simulation Monte Carlo) that run efficiently on current HPC platforms. They span length scales from microns to centimeters, and timescales from milliseconds to seconds. In the second part, I'll share some observations about the 30-year journey from gigaflops to teraflops to petaflops to exaflops, and explain what I think ukuleles have to do with high performance computing.

Kennedy Award Presentation: The Real Revolution ... from the Latency to the Throughput Age *Jesus Labarta (Barcelona Supercomputing Center)*
The recipient of the 2017 ACM and IEEE Computer Society Ken Kennedy Award is Jesus Labarta from the Technical University of Catalonia and the Barcelona Supercomputing Center.

The talk will address how architectural evolutions and multicores have impacted the way we program our machines and my vision on how we should proceed with the objective of ensuring productivity and performance. In this context, programming models and performance analysis tools are two key technologies to succeed in our efforts towards exascale and before.

Following this vision, I will briefly describe some of our developments at BSC in the area of task based models and performance analysis and prediction tools. I will overview the OmpSs and PyCOMPSs programming models as well as BSC tools (Paraver, Dimemas, performance analytics modules). Rather than focusing on the details, I will present a personal vision on the underlying considerations that have steered this research and on the methodologies and strategies towards achieving practical impact in our community.

Based on those considerations, I will explain how the title of the talk summarizes my personal perception of the times we are living in and the fundamental cultural transformation that I think will characterize the evolution of computing in the next decade.

Thursday, November 16th

Room: Mile High Ballroom
12:45 pm - 1:30 pm

Awards Ceremony

SC17 Awards Session

The SC17 conference awards, as well as selected ACM and IEEE awards, will be presented. The awards include: Best Paper, Best Student Paper, Best Poster, Best Scientific Visualization, ACM Student Research Competition, ACM Gordon Bell Prize, ACM/IEEE-CS George Michael Memorial HPC Fellowship, ACM SIGHPC / Intel Computational & Data Science Fellowships, ACM SIGHPC Emerging Woman Leader in Technical Computing Award, IEEE TCHPC Award for Excellence for Early Career Researchers in High Performance Computing, and Student Cluster Competition.

Birds of a Feather [\(back to top\)](#)

Tuesday, November 14th

Room: 201-203

12:15 pm - 1:15 pm

SIGHPC Annual Member Meeting

SIGHPC Annual Member Meeting

The annual business meeting of SIGHPC is your opportunity to hear about and discuss the status of SIGHPC and its chapters. We will also be discussing upcoming plans for the year. All of the elected officers and many of the other volunteers will be present to answer your questions about SIGHPC. Representatives from our chapters: Education, Big Data, Resource Constrained Environments (RCE), and System Administration will also be available. Come tell us how SIGHPC can meet your needs.

Room: 205-207

12:15 pm - 1:15 pm

Impacting Cancer with HPC: Opportunities and Challenges

Patricia Kovatch (Icahn School of Medicine at Mount Sinai), Thomas Steinke (Zuse Institute Berlin), Sean Hanlon (National Cancer Institute), Eric Stahlberg (Frederick National Laboratory)

High-performance computing has long been employed in cancer research and clinical applications. Ongoing challenges of deriving insight from increasing amounts of cancer related information have accelerated the significance of HPC and collaborations in cancer research, in areas including simulations, data analytics and deep learning. Underscored by several recent high-profile efforts, communities are rapidly developing where HPC is applied to cancer challenges. This SC17 BoF provides an important venue, bringing these communities together around a common theme of HPC and cancer. Following the successful format of past years, the BoF will be informative and highly-interactive with presentations and significant attendee discussion.

Room: 301-302-303

12:15 pm - 1:15 pm

Exascale Challenges and Opportunities

Douglas Kothe (Oak Ridge National Laboratory), James Ang (Sandia National Laboratories)

This BoF explores the challenges and opportunities of Exascale Computing. The motivations for exascale are centered on drivers such as economic competitiveness, national security, health care, energy, manufacturing, materials, and science. This session gathers a forum of HPC experts to identify what they consider to be key challenges in achieving exascale computing and how they would best propose to meet them, with special attention to collaboration opportunities. A few of the key topics include: exascale eco-system elements, lessons from establishing public/private partnerships, impact of exascale R&D investments on the broader HPC community. Significant audience participation will be encouraged.

Room: 402-403-404

12:15 pm - 1:15 pm

Batched, Reproducible, and Reduced Precision BLAS

Jack Dongarra (University of Tennessee), Cris Cecka (Nvidia Corporation), Timothy Costa (Intel Corporation), Sivasankaran Rajamanickam (Sandia National Laboratories), Azzam Haidar (University of Tennessee), Mawussi Zounon (University of Manchester), Piotr Luszczek (University of Tennessee)

This BoF will bring together the community focused on extending the Basic Linear Algebra Software (BLAS). The existing BLAS have proven to be effective in assisting portable, efficient software for sequential and the current class of high-performance computers. We'd like to investigate the possibility of extending the currently accepted standards to provide greater parallelism for small size operations, reproducibility, and reduced precision support. This is an open forum to discuss and formalize details. The agenda and talks from past workshops can be found here: <http://bit.ly/Batch-BLAS-2017> <http://bit.ly/Batch-BLAS-2016>

A standard interface will be considered for the Batched, Reproducible, and Reduced Precision BLAS.

Room: 405-406-407
12:15 pm - 1:15 pm

Resilient Programming Environments

Keita Teranishi (Sandia National Laboratories), George Bosilca (University of Tennessee)

Dealing with process faults were early introduced and heartily embraced by the parallel programming paradigms used in industry where resilience was always a core component of the toolbox proposed to the user. As resiliency is becoming more critical in HPC, user communities are left searching for alternative extensions to their traditional programming paradigms.

This BoF is intended as a forum between experts in different programming paradigms and users looking for solutions to circumvent the pitfalls of coping with failures. Join and become a voice that shapes the resilience support provided by parallel programming paradigms and their supporting software infrastructure.

Room: 501-502
12:15 pm - 1:15 pm

SKA: The Ultimate Big Data Project

Nicolás Erdödy (Open Parallel Ltd), Happy Sithole (Center for High Performance Computing in South Africa)

The Square Kilometre Array (SKA) will be the world's largest radio-telescope, to be built in Australia and South Africa in an 11 countries' effort. SKA1 design will finish in 2018 with construction scheduled 2019 – 2025. SKA will produce Big Data on a scale that cannot be processed with today's technologies. This BoF brings together vendors and OEMs with participants from SKA's computing platform design to discuss topics such as: * Data Management Challenges, * HPDA, * Frameworks for SKA-scale projects, * Compute platform design, OS, Middleware, * Accelerators and HPC software, * Power considerations toward exascale computing and others.

Room: 503-504
12:15 pm - 1:15 pm

Interactivity in Supercomputing

Fernanda Foertter (Oak Ridge National Laboratory), Sadaf Alam (Swiss National Supercomputing Centre), Peter Messmer (Nvidia Corporation)

The HPC landscape has undergone a lot of changes in the past years. In addition to more heterogeneous system architectures, a wealth of novel workloads specifically from the data sciences have emerged. With these novel workloads, a broader user base with different workflows and expectations about interactivity have arrived. In addition, traditional HPC applications have embraced more interactive workflows. The goal of this BoF is to bring together domain scientists, tool developers and HPC center administrators to identify the scientific impact and technical challenges of highly interactive access to HPC resources.

Room: 507
12:15 pm - 1:15 pm

Americas HPC Collaboration

Phillipe Navaux (Federal University of Rio Grande do Sul, Advanced Computing Service for Latin America and the Caribbean), Mark Dietrich (Compute Canada), Nancy Wilkins-Diehr (San Diego Supercomputer Center), Carlos Jaime Barrios Hernandez (Advanced Computing Service for Latin America and the Caribbean, Industrial University of Santander)

This BoF Americas HPC Collaboration reach to present different experiences of the collaboration between XSEDE, Compute Canada and different countries in Latin America after the first meeting developed during SC15 at Austin, Texas. The goals of this BoF is to show the state of continental of the HPC collaboration, the development of regional HPC networks (XSEDE, Compute Canada and SCALAC) and define a new roadmap for the next two years following the experience of PRACE in Europe to create an Americas HPC Partnership. (A official letter of intention will be proposed to formalize the collaboration). More information at: <http://www.sc3.uis.edu.co/bofamericas-hpc-collaboration/>

Room: 601
12:15 pm - 1:15 pm

LLVM in HPC: Uses and Desires

Hal Finkel (Argonne National Laboratory, LLVM Foundation), Jim Cownie (Intel Corporation)

LLVM is the prime environment for developing new compilers and language-processing tools. In this BoF a group of LLVM experts who are driving the use of LLVM in HPC-relevant projects will give short presentations of their work, answer your questions, and discuss the future of LLVM in HPC.

As well as Hal Finkel, our experts include Doug Miles (Fortran), Tobias Grosser (Polly), Keno Fischer (Julia), and Carlo Bertolli (OpenMP).

Our goal is to connect you with LLVM experts so that you understand some of the uses of LLVM, and they understand what tools and LLVM enhancements you want.

Room: 603

12:15 pm - 1:15 pm

Software Engineers: Careers in Research

Daniel Katz (University of Illinois), Jay Jay Billings (Oak Ridge National Laboratory), Alys Brett (UK Atomic Energy Authority), Jeffrey Carver (University of Alabama), Catherine Jones (Science and Technology Facilities Council), Lauren Michael (University of Wisconsin), David Perez-Suarez (University College London), Andrew Turner (University of Edinburgh), Neil Chue Hong (University of Edinburgh)

Many people in research organizations around the world combine expertise in programming with an intricate understanding of research. Although this combination of skills is extremely valuable, these people often lack a formal career path, particularly in the academic system. There is no easy way to recognise their contributions, reward them, or represent their views.

In the UK (and increasingly in Europe), the term Research Software Engineer is used to describe this important role, while the US does not have a single term. This BoF aims to raise awareness of the role and bring the community together to address challenges.

Room: 605

12:15 pm - 1:15 pm

Total Cost of Ownership and HPC System Procurement

Eric Boyer (GENCI), James Rogers (Oak Ridge National Laboratory), James Laros (Sandia National Laboratories), Natalie Bates (Energy Efficient HPC Working Group), Jason Hick (Los Alamos National Laboratory), Steve Martin (Cray Inc), Gert Svensson (KTH Royal Institute of Technology)

The predominant goal for procurement of HPC systems is to identify the optimal solution to both technical and financial targets that maximizes the contribution of that system to the organization's mission. Beyond the acquisition cost of the system, it is also important to consider the total costs of ownership, including the improvements necessary to host the system, the infrastructure that supports its operation, and the significant operational costs associated with that new HPC system. HPC leaders from across the globe will discuss and debate key procurement requirements and lessons learned that can contribute to greater efficiency and reduced operational costs.

Room: 701

12:15 pm - 1:15 pm

Ceph Applications in HPC Environments

Douglas Fuller (Red Hat Inc), Benjamin Lynch (University of Minnesota)

With a series of lightning talks, we will present and discuss the use of Ceph in HPC environments. Ceph is an open-source distributed object store with an associated file system widely used in cloud and distributed computing. In addition, both the object store and file system components are seeing increasing deployments as primary data storage for traditional HPC. Ceph is backed by a robust, worldwide open source community effort with broad participation from major HPC and storage vendors.

Room: 703

12:15 pm - 1:15 pm

PBS Pro Open Source Project Community BoF

Greg Matthews (NASA Ames Research Center), Andrew Wellington (Australian National University), Bill Nitzberg (Altair Engineering)

PBS Pro schedules jobs and manages workloads for HPC clusters and clouds. The software was born at NASA in the 1990s, and, by the 2010s, became one of the top tools of its kind. In mid-2016, PBS Pro became a fully open source solution, with a growing community of developers and users around the globe. Join the community (www.pbspro.org) -- users, developers, contributors, and

open source enthusiasts -- to learn what's new and to drive the future of PBS Pro.

Room: 702

12:15 pm - 1:15 pm

Omni-Path User Group (OPUG) Meeting

Philip Murphy (Intel Corporation), J. Ray Scott (Pittsburgh Supercomputing Center)

The goal of this BoF is to provide a forum for users and other interested parties to share their experiences and insights using the Omni-Path interconnect fabric from Intel. The format will consist of an introduction, an Omni-Path product update on the hardware, new software features, important IFS updates, and more, followed by panel presentations from sites who have Omni-Path installations, and group discussion, targeting 50% of the time to the short presentations and 50% to audience engagement.

Room: 704-706

12:15 pm - 1:15 pm

Fortran Is 60 Years Old - Has It Changed for the Better?

John Levesque (Cray Inc)

Fortran is 60 years old and maligned by many in the HPC community for being non-productive. While its usage in HPC is not as dominant as it was prior to 2000, it is still the major language employed for high performance. The Fortran standards committee is changing Fortran to be a more productive language. Have these changes to Fortran benefitted the community? We will hear from ex-Fortran coders who have moved to other languages and from current Fortran coders who still believe it has advantages over the alternatives.

Short five minutes talks from 3-4 panelists will stimulate discussion.

Room: 710-712

12:15 pm - 1:15 pm

Regression Testing and Monitoring Tools

Reuben Budiardja (Oak Ridge National Laboratory), Guilherme Peretti-Pezzi (Swiss National Supercomputing Centre), Bilel Hadri (King Abdullah University of Science and Technology)

Supercomputers are becoming larger and more complex tightly integrated systems consisting of many different hardware components, tens of thousands of processors and memory chips, kilometers of networking cables, large numbers of disks, and hundreds of applications and libraries. To increase scientific productivity and ensure that applications efficiently and effectively exploit a system's full potential, all the components must deliver reliable, stable, and performant service. This BoF discusses the best practice from supercomputing centers using different strategies on system performance assessments and seeks those interested in sharing experiences to detect issues related to the performance and functionality of HPC systems.

Room: 201-203

5:15 pm - 7:00 pm

Containers in HPC

Jeffrey Sica (University of Michigan), Gregory Kurtzer (SingularityWare LLC), Reid Friedhorsky (Los Alamos National Laboratory), Andrew Younge (Sandia National Laboratories), Abdulrahman Azab (University of Oslo, Partnership for Advanced Computing in Europe (PRACE)), Tim Randles (Los Alamos National Laboratory), Richard Canon (Lawrence Berkeley National Laboratory)

Containers are rapidly gaining traction in HPC because they promise significantly greater software flexibility, reliability, and portability for users. We will provide an opportunity to engage with leaders who are developing container solutions (e.g., Charliecloud, Singularity, and Shifter) and promoting the container model. The agenda includes an overview of container options for HPC environments and use cases they are enabling, as well as an interactive period to ask questions and share experiences. A key goal of the session is to understand and develop an agenda for solving questions and concerns that are holding back broader adoption.

Room: 205-207

5:15 pm - 7:00 pm

Tracking and Analyzing Job-level Activity Using Open XDMoD, XALT and OGRT

Georg Rath (Lawrence Berkeley National Laboratory), Thomas Furlani (University at Buffalo), Robert DeLeon (University at Buffalo), Robert McLay (University of Texas; Texas Advanced Computing Center, University of Texas)

This BoF is for those interested in the increasingly important need to track and analyze activity on large-scale systems: usage, performance, and impact, down to the level of each individual job. Open XDMoD primarily displays aggregated data: it provides your own web portal to view, summarize, and analyze this data. We will discuss recent developments and improvements in XDMoD. XALT and OGRT are about collections: these tools are battle-tested tools focused on job-level usage data. They track executables and libraries with the lowest possible overhead. Join us for demos, discussions, and a wide-ranging exchange of information!

Room: 210-212
5:15 pm - 7:00 pm

OpenACC API User Experience, Vendor Reaction, Relevance, and Roadmap

Sunita Chandrasekaran (University of Delaware), Fernanda Foertter (Oak Ridge National Laboratory), Guido Juckeland (Helmholtz-Zentrum Dresden-Rossendorf), Duncan Poole (Nvidia Corporation)

OpenACC, a directive-based high-level parallel programming model, has gained rapid momentum among scientific application users - the key drivers of specification. The user-friendly programming model has facilitated migration of several applications such as CAM, ANSYS Fluent, Gaussian, VASP, COSMO and is also seen as the entry-level programming model on the top 5 supercomputers (Top500 list). As in previous years, this BoF invites scientists, programmers and researchers to discuss their experiences in adopting OpenACC for scientific legacy applications, learn about the roadmaps from implementers and the latest developments in the specification.

Room: 301-302-303
5:15 pm - 7:00 pm

HPC Carpentry - Practical, Hands-On HPC Training

Christina Koch (University of Wisconsin), Tracy Teal (Michigan State University), Robert Freeman Jr (Harvard University), Chris Bording (Pawsey Supercomputing Centre), Martin Callaghan (University of Leeds), Andrew Turner (University of Edinburgh)

Carpentry training has proven hugely successful; providing researchers with the tools required to exploit software and data to improve their research. The increasing role HPC is playing in research means that more researchers need HPC skills in order to progress. We are designing an HPC Carpentry course to bring the benefit of the Carpentry approach to current and potential HPC users. We will describe the state of HPC Carpentry and facilitate discussions on its Goals, Design, Structure and Implementation. We will use this BoF to employ the expertise and experience of the HPC community to deliver the best course possible.

Room: 402-403-404
5:15 pm - 7:00 pm

HPC Graph Toolkits and the GraphBLAS Forum

Mahantesh Halappanavar (Pacific Northwest National Laboratory), Alex Pothen (Purdue University), Michael Wolf (Sandia National Laboratories), Fabrizio Petrini (Intel Corporation), Aydin Buluc (Lawrence Berkeley National Laboratory), Timothy Mattson (Intel Corporation), Antonino Tumeo (Pacific Northwest National Laboratory)

Government agencies and companies are looking for a new generation of tools able to efficiently solve large scale combinatorial problems for applications ranging from big data science to national security. This BoF aims at gathering the community of people interested at frameworks and workflows for large scale graph analytics, surveying the current approaches, identifying new challenges and opportunities, and laying a path towards future interoperable infrastructures. The BoF will discuss GraphBLAS as a key component of such toolkits and bring together the GraphBLAS community. We will report: (1) the essential design patterns using GraphBLAS, and (2) requirements for GraphBLAS 2.0.

Room: 405-406-407
5:15 pm - 7:00 pm

Reconfigurable Computing in Exascale

David Donofrio (Lawrence Berkeley National Laboratory), John Leidel (Tactical Computing Laboratories), Hans-Christian Hoppe (Intel Corporation), Marie-Christine Sawley (Intel Corporation), Kazutomo Yoshii (Argonne National Laboratory), Kentaro Sano (RIKEN, Tohoku University), Franck Cappello (Argonne National Laboratory), Sven Karlsson (Technical University of Denmark)

While traditional HPC systems continue making great advances, it is argued that paradigm shifts, including radical new architectures, are needed to continue scaling performance beyond exascale at sustainable cost and energy levels. Such architectures will likely employ reconfigurability or FPGAs for certain types of parallelism.

This BoF targets persons interested in these subjects aiming to further understand their role, usage, run-time systems, programming models, the role of open-source technology and what challenges to widespread adoption remain. One goal is also to continue to build a HPC FPGA community.

The BoF will combine interactive presentations with audience driven discussions.

Room: 501-502
5:15 pm - 7:00 pm

European Exascale Projects and Their Global Contributions

Marcin Ostasz (European Technology Platform for High Performance Computing, Barcelona Supercomputing Center), Jean-Philippe Nominé (European Technology Platform for High Performance Computing, French Alternative Energies and Atomic Energy Commission), François Bodin (University of Rennes), Jean-Pierre Panziera (European Technology Platform for High Performance Computing, Atos)

Europe will present a holistic view of its HPC development and showcase its selected projects that contribute to the global exascale R&D effort.

The European HPC Technology and Application programme covers the entire HPC system stack and application expertise; including system prototypes (see www.etp4hpc.eu/euexascale for a full description).

We will summarize our ambitions and the contribution of our projects, present selected technology (Mont-Blanc, SAGE) and application projects (BioExcel) and discuss our role in the global HPC effort. The European EXDCI (www.exdci.eu) booth, publications (European HPC Handbook) and the ETP4HPC website will facilitate networking before and after the BoF.

Room: 503-504
5:15 pm - 7:00 pm

Fabric APIs - libfabric User Perspective and C++ Standardization

Sean Hefty (Intel Corporation), Chris Taylor (US Department of Defense), Jeffrey Squyres (Cisco Systems)

Application performance and scalability is affected by the design and usability of network APIs. Open Fabrics libfabric promises high-performance access to fabric services, through an implementation agnostic interface.

Based on the architectural work that went into the creation of a fabric-agnostic API, Open Fabrics is a natural fit for discussing expanding the adoption of fabric APIs.

This BoF will begin with a series of lightning talks to share experiences developing to libfabric. It will conclude by proposing that Open Fabrics develop fabric extensions to the C++ networking technical specification, for submission to the ISO C++ standards committee.

Room: 507
5:15 pm - 7:00 pm

Memory-Centric Architectures for the Cloud and HPC

Peter Kogge (University of Notre Dame), Sudhakar Yalamanchili (Georgia Institute of Technology)

Computing systems are at an inflection point where fundamental changes in technology and applications are poorly matched to memory systems of modern computing-system designs. At all levels of the system, memory is becoming the engine of performance growth leading to new memory hierarchies, near data computing architectures, and new data centric programming models and system management frameworks.

The goal of this BoF is to bring together application developers, system software developers, memory system designers, and system architects to create community that can identify major trends and technical challenges, and catalyze collaborative opportunities between industry, academia, and government to address them.

Room: 601
5:15 pm - 7:00 pm

Software Engineering and Reuse in Computational Science and Engineering

Maxim Belkin (University of Illinois), Alys Brett (Culham Centre for Fusion Energy), Jeffrey Carver (University of Alabama), Neil

Chue Hong (University of Edinburgh), Mike Folk (HDF Group), Gerard Gorman (Imperial College, London), William Gropp (University of Illinois), Michael Heroux (Sandia National Laboratories, St. John's University), Daniel Katz (University of Illinois), David Keyes (King Abdullah University of Science and Technology), Scott Lathrop (National Center for Supercomputing Applications, University of Illinois; Shodor Education Foundation), Randall LeVeque (University of Washington), James Lin (Shanghai Jiao Tong University), Lois Curfman McInnes (Argonne National Laboratory), Kengo Nakajima (University of Tokyo), Andy Terrell (NumFOCUS), David Bernholdt (Oak Ridge National Laboratory)

Software developers, researchers, trainers, and outreach staff will meet to raise awareness of the challenges and opportunities for developing and promoting software engineering practices, including the development of reusable software to enhance the computational science and engineering ecosystem for HPC systems to accelerate the “time to science”.

The BoF attendees will help develop an international “community of practice” to share experiences in developing quality software, gathering community feedback, addressing interoperability, and promoting software to applications developers and users.

We will identify strategies for ongoing efforts to inform, engage, and benefit the community and thereby advance the HPC software ecosystem.

Room: 603

5:15 pm - 7:00 pm

Usability, Scalability and Productivity on Many-Core Processors: Intel Xeon Phi

Douglas Doerfler (Lawrence Berkeley National Laboratory), Estela Suarez (Forschungszentrum Juelich), Hai Ah Nam (Los Alamos National Laboratory)

Recent deployments of large-scale Intel Xeon Phi processor (Knight Landing)-based systems have provided users with early experiences on many-core processors. This BoF, conducted by the Intel Xeon Phi Users Group (IXPUG), will provide a forum for application and tool developers, HPC center staff, and industry experts to discuss their successes and challenges. This BoF will showcase code optimization successes, particularly, how to achieve usability, scalability and productivity on KNL-based systems through invited talks, a panel session and a community Q&A. IXPUG is an independent users group for anyone interested in application performance on the Intel Xeon Phi. See <http://ixpug.org>

Room: 607

5:15 pm - 7:00 pm

Cross-Layer Allocation and Management of Hardware Resources in Shared Memory Nodes

Pete Beckman (Argonne National Laboratory), Brice Goglin (French Institute for Research in Computer Science and Automation (INRIA)), Emmanuel Jeannot (French Institute for Research in Computer Science and Automation (INRIA))

The goal of this BoF is to gather the community (from runtime system to compilers) working in the area of hardware resource allocation for threads. We will discuss this problem, share visions, propose solutions, and coordinate a worldwide effort.

We will consider all the resources of a shared memory node and discuss how to coordinate resource sharing by different parts of the software stack to avoid competition for these resources.

Participants will be able to provide their own vision through discussions. The goal is to come up with a document specifying possible solutions and discuss possible implementations.

Room: 701

5:15 pm - 7:00 pm

The ARM User Experience: Testbeds and Deployment at HPC Centers

Mitsuhisa Sato (RIKEN), Filippo Mantovani (Barcelona Supercomputing Center), Simon McIntosh-Smith (University of Bristol), Jack Wells (Oak Ridge National Laboratory)

The ARM architecture is gaining a lot of traction in the HPC community as evidenced by several ARM-based projects including the Japanese Post-K, European Mont-Blanc, and the UK's GW4/EPSRC efforts. This BoF will bring together users and computational centers in the HPC community to present and discuss their experiences of using or deploying state-of-the-art ARM HPC systems. The BoF will consist of a series of brief presentations and a panel discussion session. It is the intent of the organizers that more than 50 percent of the time will be invested in community discussion, including vigorous participation from general attendees.

Room: 703

5:15 pm - 7:00 pm

Special Interest Group on HPC in Resource Constrained Environments (SIGHPC-RCE)

Elizabeth Leake (STEM-Trek), Hensley Omorodion (University of Benin)

The SC17 BoF continue discussions that were begun during SC16, share results from the “Challenges” annual survey (now in its seventh year and provides great longitudinal information of interest to systems administrators who wish to know what others are using for scheduling, security, etc.), develop competency (workforce development) for HPC, explore how federated and interfederated cyberinfrastructures can be shared with regions that lack resources, improve cybersecurity in shared cyberecosystems, and facilitate access cloud resources(CI) through federated identity. New chapter officers will be announced, and audience-driven discussions will be fostered to help shape the SIG’s scope for the coming year.

Room: 702

5:15 pm - 7:00 pm

Accelerating Big Data Processing and Machine/Deep Learning Middleware on Modern HPC Clusters

Gil Bloch (Mellanox Technologies), Dhableswar Panda (Ohio State University), Xiaoyi Lu (Ohio State University), Gilad Shainer (Mellanox Technologies)

The convergence of HPC, Big Data, and Machine/Deep Learning is the next game-changing business opportunity. Machine/Deep Learning is a pillar of today’s technological world and enables making better decisions based on the great amounts of data being collected. This BoF will involve all the speakers and the audience to identify the most critical challenges facing the community and coming up with a roadmap for the next 5-10 years in accelerating Big Data processing and Machine/Deep Learning middleware (e.g., Hadoop/Spark/TensorFlow/Caffe) on modern HPC clusters. Recent examples from organizations such as Baidu, Tencent, NVIDIA, Stanford, OSU, and more will be discussed.

Room: 704-706

5:15 pm - 7:00 pm

Improving Numerical Computation with Practical Tools and Novel Computer Arithmetic

*Michael Lam (James Madison University, Lawrence Livermore National Laboratory), John Gustafson (National University of Singapore; A*STAR Computational Resource Center, Singapore)*

Efficient use of computer arithmetic is crucial to HPC. Unfortunately, few software analysis tools exist to analyze floating-point accuracy or to recommend robust mixed-precision modifications for full-sized applications. At the same time, next-generation arithmetic formats such as Gustafson’s latest Type III unums (posits and valids) offer the possibility of better accuracy, performance, energy efficiency, and reproducibility, but their implementations and tool ecosystems are still young. The goal of this BoF is to connect researchers and industry practitioners interested in improving the state of the art in this area, and to begin to identify concrete ways to work together.

Room: 710-712

5:15 pm - 7:00 pm

PowerAPI, GEOPM and Redfish: Open Interfaces for Power/Energy Measurement and Control

Jonathan Eastep (Intel Corporation), Jeff Autor (Hewlett Packard), Nicolas Dube (Hewlett Packard), Vitali Morozov (Argonne National Laboratory), Todd Rosedahl (IBM), Steve Martin (Cray Inc), James Laros, III (Sandia National Laboratories), Stephen Oliver (Sandia National Laboratories), Ramkumar Nagappan (Intel Corporation), Vinanti Phadke (Hewlett Packard Enterprise), Ryan Grant (Sandia National Laboratories)

With every generation of new hardware, more power measurement and control capabilities are exposed. As software applications for these capabilities are developed and deployed, there is a need for open-interfaces and frameworks that allow for ease of communication between components within a platform as well as portability across platforms. In this BoF, we discuss three efforts for such open-interfaces and frameworks: these are PowerAPI, GEOPM and Redfish. The BoF will introduce each effort and feature an interactive panel discussion with experts from currently implementing and adopting organizations. There will also be a focus on synergy and compatibility between them.

Room: Mile High Ballroom

5:15 pm - 7:00 pm

TOP500 Supercomputers

Jack Dongarra (University of Tennessee), Horst Simon (Lawrence Berkeley National Laboratory), Martin Meuer (ISC Events), Erich Strohmaier (Lawrence Berkeley National Laboratory)

Now in its 24th year, the TOP500 list of supercomputers serves as a "Who's Who" in the field of High Performance Computing (HPC). The TOP500 list was started in 1993 as a project to compile a list of the most powerful supercomputers in the world. It has evolved from a simple ranking system to a major source of information to analyze trends in HPC. The TOP500 has traditionally been published at the SC in November and at the ISC in June of each year. The 50th TOP500 list will be published in November 2017 just in time for SC17.

Wednesday, November 15th

Room: 201-203
12:15 pm - 1:15 pm

IEEE CS TCHPC Meeting

IEEE CS TCHPC Meeting

The annual IEEE Computer Society Technical Consortium on High Performance Computing (TCHPC) Birds of a Feather session at SC provides a forum to discuss the consortium and its mission, activities and initiatives with members of the community, as well as to get inputs from the community. It give members of the community a chance to interact with the TCHPC volunteers and representatives from its member technical committees and to explore opportunities to get involved.

Room: 205-207
12:15 pm - 1:15 pm

The Future of NSF Advanced Cyberinfrastructure

Alejandro Suarez (National Science Foundation), Edward Walker (National Science Foundation)

The National Science Foundation's vision and investment plans for cyberinfrastructure (CI) are designed to address the evolving needs of the science and engineering research community. Program Directors from NSF's Office of Advanced Cyberinfrastructure will update attendees on new NSF cyberinfrastructure strategies and activities, and discuss the latest funding opportunities in advanced computing, software infrastructure, data infrastructure, networking, cybersecurity, and learning and workforce development. Presentations will also discuss the NSF CI 2030 Request for Information and associated plans, new science-CI collaborative opportunities, and strategic priorities such as NSF Big Ideas. Ample time will be provided for Q&A with NSF staff.

Room: 210-212
12:15 pm - 1:15 pm

The HDF5 Dataverse

Quincey Koziol (Lawrence Berkeley National Laboratory), David Pearah (HDF Group)

The HDF Group will provide a forum for its diverse user community to share ideas and discuss current initiatives. David Pearah will present the latest HDF5 roadmap, including upcoming releases and soliciting input on the future roadmap. Quincey Koziol will moderate a panel with representatives from research, commercial, and government organizations who will present case studies on how they leverage HDF technologies to solve their big compute problems as well as discuss the challenges of using HDF5.

Room: 301-302-303
12:15 pm - 1:15 pm

Women in HPC: Non-Traditional Paths to HPC and How They Can and Do Enrich the Field

Toni Collis (University of Edinburgh), Neelofer Banglawala (University of Edinburgh), Rebecca Hartman-Baker (National Energy Research Scientific Computing Center)

There is often the unspoken expectation that those best equipped to work in HPC have studied computer science or a physical science to PhD level. However, those who have not taken such a 'standard' or 'traditional' route to HPC can and do contribute meaningfully to the field. In this BoF we examine how to broaden the community's perceptions to access a more diverse talent pool. We will hear from three women who came to HPC from non-standard directions, followed by discussion with audience participation on how to recruit from non-traditional backgrounds, broadening the HPC talent pool.

Room: 402-403-404
12:15 pm - 1:15 pm

Modeling and Simulation of Communication in HPC Systems

Nikhil Jain (Lawrence Livermore National Laboratory), Jeremiah Wilke (Sandia National Laboratories), Misbah Mubarak (Argonne National Laboratory)

Modeling and simulation play a key role in analyzing the impact of system design choices on applications' communication performance. The HPC community currently offers several predictive design tools with varying focus and strengths. To make the best use of these tools, it is critical that the tool developers and end users interact to identify the user requirements and capabilities of these tools. In this BoF, we will discuss recent advances in modeling and simulation techniques and identify opportunities for cross-pollination among various predictive design techniques. We will also discuss the requirements and challenges posed by the end users.

Room: 405-406-407
12:15 pm - 1:15 pm

Distributed and Heterogeneous Programming in C++ for HPC

Michael Wong (Codeplay Software Ltd, Khronos Group Inc.), Hal Finkel (Argonne National Laboratory)

In response to the HPC requirements from CORAL/SUMMIT to achieve exascale performance, we will discuss the programming models that support heterogeneous programming in C and C++ and future standardization work toward that exascale goal. We will discuss research in this domain and consolidate usage experience with the aim of passing that experience to ISO C and C++.

There are a number of C++ frameworks for parallel programming, including HPX, KoKKos, Raja, C++AMP, HCC, Boost.Compute, CUDA, and more. SYCL from Khronos provides heterogeneous computing built on OpenCL and C++, and Codeplay has released ComputeCpp Community Edition.

Room: 501-502
12:15 pm - 1:15 pm

State of the Practice: Energy and Power Aware Job Scheduling and Resource Management (EPA-JSRM)

Toshio Endo (Tokyo Institute of Technology), Carlo Cavazzoni (CINECA), Bilel Hadri (King Abdullah University of Science and Technology), Torsten Wilde (Leibniz Supercomputing Centre), Milos Puzovic (Hartree Centre), Francis Belot (Atomic Energy and Alternative Energies Commission), Toshihiro Hanawa (University of Tokyo), Josip Loncaric (Los Alamos National Laboratory), Keiji Yamamoto (RIKEN), Greg Koeing (Energy Efficient HPC Working Group), Kevin Pedretti (Sandia National Laboratories), Siddhartha Jana (Intel Corporation), Matthias Maiterth (Intel Corporation), Andrea Borghesi (University of Bologna), David Montoya (Los Alamos National Laboratory), Natalie Bates (Energy Efficient HPC Working Group)

Supercomputing centers are beginning a transition to "dynamic power and energy management." Cost control is a major factor. Supercomputer systems have increasingly rapid, unpredictable and large power fluctuations. In addition, electricity service providers may request supercomputing centers to change their timing and/or magnitude of demand to help address electricity supply constraints. To adapt to this new landscape, centers may employ JSRM strategies to dynamically and in real-time control their electricity demand. This BoF presents results of a global survey of supercomputing centers using these JSRM strategies and seeks those interested in sharing experiences with dynamic power and energy management.

Room: 503-504
12:15 pm - 1:15 pm

Characterizing Faults, Errors, and Failures in Extreme-Scale Systems

Stephen McNally (Oak Ridge National Laboratory), Christian Engelmann (Oak Ridge National Laboratory)

This session brings together a group of international experts from the Accelerated Data Analytics and Computing Institute to present their efforts in characterizing faults, errors, and failures in extreme-scale systems and to discuss practical experiences with software tools and infrastructures, including operational aspects. The ADAC Institute is a collaboration between Oak Ridge National Laboratory, the Swiss Federal Institute of Technology Zurich, Tokyo Institute of Technology, Lawrence Livermore National Laboratory, Juelich Research Centre, the University of Tokyo, Cray, Nvidia, and Intel. The session includes short presentations and a discussion that focuses on future research and development, collaboration opportunities, and vendor interactions.

Room: 507
12:15 pm - 1:15 pm

OpenHPC Community BoF

David Brayford (Leibniz Supercomputing Centre), Derek Simmel (Pittsburgh Supercomputing Center), Thomas Sterling (Indiana University), Scott Suchyta (Altair Engineering), Nirmala Sundararajan (Dell Inc), Karl Schulz (Intel Corporation)

There is a growing sense within the HPC community for the need to have an open community effort to more efficiently build, test, and deliver integrated HPC software components and tools. Formed initially in November 2015 and formalized as a Linux Foundation project in June 2016, OpenHPC is endeavoring to address this need. At this BoF, speakers from the OpenHPC Technical Steering Committee will provide a technical overview of the project and near-term roadmaps. We then invite open discussion giving attendees an opportunity to provide feedback on current conventions, packaging, request additional components and configurations, and discuss general future trends.

Room: 601
12:15 pm - 1:15 pm

Practical Reproducibility by Managing Experiments Like Software

Carlos Maltzahn (University of California, Santa Cruz), Jay Lofstead (Sandia National Laboratories), Michael Heroux (Sandia National Laboratories), Kate Keahey (Argonne National Laboratory), Ivo Jimenez (University of California, Santa Cruz)

This BoF is for HPC community members who want to make sharing and re-executing scientific experiments more practical. As a starting point, we will introduce the Popper Protocol (<http://falsifiable.us>), a set of guidelines for managing scientific explorations as software projects. We will moderate discussions, and conduct a survey among attendants on tools/services that can be used to "Popperize" scientific explorations, i.e. fully automate/document experiments by scripting all aspects of their execution/analysis and version-control their evolution using DevOps tools. Our goal: create a list of tool-chain templates that domain-specific communities can use to incentivize researchers/students to generate easily re-executable/shareable experiments.

Room: 603
12:15 pm - 1:15 pm

Second Annual Meeting of the SIGHPC - Big Data Chapter

Suzanne McIntosh (New York University), Stratos Efstathiadis (New York University)

The goal of the BoF is to gather for the second time members and non-members of the SIGHPC BigData Virtual Chapter who are interested in learning about the challenges of converging Big Data and HPC. The BoF will give people the opportunity to hear about existing challenges and openly discuss solutions, tools, and new approaches on how to best utilize available Big Data and HPC resources.

Room: 605
12:15 pm - 1:15 pm

15th Graph500 List

David Bader (Georgia Institute of Technology), Peter Kogge (University of Notre Dame), Andrew Lumsdaine (Pacific Northwest National Laboratory), Richard Murphy (Micron Technology Inc)

Data intensive supercomputer applications are increasingly important workloads, especially for "Big Data" problems, but are ill suited for most of today's computing platforms (at any scale!). As the Graph500 list has surpassed 200 entries, it has demonstrated the challenges of even simple analytics. Backed by a steering committee of 30 international HPC experts from academia, industry, and national laboratories, this effort serves to enhance data intensive workloads for the community. This BoF will unveil the 15th Graph500 list, initial rankings for the new SSSP kernel released at ISC2017, and enhance the new energy metrics the Green Graph500.

Room: 607
12:15 pm - 1:15 pm

Small Business and the Exascale Computing Project

Barbara Helland (US Department of Energy), Lucy Nowell (US Department of Energy, Networking and Information Technology Research and Development Program), Jack Wells (Oak Ridge National Laboratory), Richard Carlson (US Department of Energy), Ashley Barker (Oak Ridge National Laboratory), Richard Lethin (Reservoir Labs Inc, Yale University)

Working with the US Exascale Computing Project (ECP) is an important opportunity for Small Business and the United States. Small Business is a major source of US innovation, fueled by entrepreneurial spirit, venture capital, unique technology, and hard work. There are opportunities for Small Business to both benefit from ECP funding (providing R&D and NRE services and technology to ECP) and from ECP technology (utilizing the software, hardware, and systems from ECP). This session will bring together ECP leadership, SBIR PMs, and Small Business to make connections and understand opportunities, to enhance ECP success and impact with Small Business participation.

Room: 701
12:15 pm - 1:15 pm

The Internet of Things and HPC: Are They Teaming Up to Work Together?

Nicolás Erdödy (Open Parallel Ltd), Pete Beckman (Argonne National Laboratory, Northwestern University)

IoT devices with edge computing are being deployed everywhere. By almost every physical measure, IoT and HPC seem to be opposites. However, data-driven simulations, machine learning, and in-situ parallel processing are linking them. We will bring together projects for a panel discussion of these questions: a) How will multicore edge devices be programmed? b) Can low-power deep learning accelerators designed for IoT (e.g. Myriad 2) be adapted to HPC? c) Can machine-learning models be shared and connected between IoT and HPC?

Audience: Experts and practitioners in: simulation and modeling, architecture, machine learning, multicore computing, sensing systems, IoT, and mobile computing.

Room: 703
12:15 pm - 1:15 pm

The Message Passing Interface: On the Road to MPI 4.0 and Beyond

Martin Schulz (Technical University Munich)

The Message Passing Interface (MPI) is one of the most dominant programming models for HPC environments. Its specification is driven by the MPI Forum, an open forum consisting of MPI developers, vendors and users. This BoF Meeting will provide some insight into the current topics discussed in the forum as well as the process of how features are added to the standard. It is intended to keep the larger HPC community informed about current activities and long-term directions, as well as encourage larger community participation in this crucial standard for the supercomputing community.

Room: 710-712
12:15 pm - 1:15 pm

Contemporary Design of Supercomputer Experiments

Jed Brown (University of Colorado, Boulder), Stefan Wild (Argonne National Laboratory), Prasanna Balaprakash (Argonne National Laboratory), Paul Constantine (University of Colorado, Boulder), Jeffrey Hokanson (Colorado School of Mines), Felix Wolf (Technical University Darmstadt), Dmitry Duplyakin (University of Utah)

This BoF facilitates community discussion on the topics in contemporary design of supercomputer experiments. Design of Experiments (DOE) is a critical component of the knowledge discovery process in science and engineering and has impact in a wide variety of fields, including supercomputing. The scale of contemporary supercomputer simulations and the growth of computational capabilities toward exascale dictate the need for practical DOE that incorporate adaptive, multi-objective, and cost-efficient strategies. In this BoF, we propose to discuss promising state-of-the-art techniques and facilitate the community building around the research and development focused on innovative supercomputing DOE.

Room: 201-203
5:15 pm - 7:00 pm

The Virtual Institute of I/O and the IO-500

Jay Lofstead (Sandia National Laboratories), John Bent (Seagate Government Solutions), Julian Kunkel (German Climate Computing Center)

Due to the increasing complexity of HPC data management, activities in the storage research community have increased over the last few years. The general purpose of this BoF is to foster this community.

An important activity is the community driven development of an IO-500 benchmark. The speakers will briefly introduce the international Virtual Institute for I/O (VI4IO, <http://vi4io.org>) and focus on the status of the IO-500 development and the usage of the benchmark. A highlight is the presentation of the first IO-500 list. The direction of VI4IO and the standardization of the IO-500 is then discussed with the participants.

Room: 205-207
5:15 pm - 7:00 pm

OpenMP® is Twenty. Where Is It Going?

Michael Klemm (OpenMP Architecture Review Board, Intel Corporation), Jim Cownie (Intel Corporation)

OpenMP is twenty this year. In HPC it is the most commonly used way to exploit thread-level parallelism, and can now support accelerators.

The aim of this BoF is to give you information about OpenMP's current state and future evolution. A significant amount of time (75%) will be available for you to ask our OpenMP experts your OpenMP questions. Our experts are members of the OpenMP ARB from both companies which support OpenMP and sites which are important OpenMP users.

OpenMP is now much more than "parallel loops". If you use it, or expect to, this BoF is for you.

Room: 210-212
5:15 pm - 7:00 pm

From Outreach to Education to Researcher - Innovative Ways of Expanding the HPC Community

Nicholas Brown (University of Edinburgh), Julie Mullen (MIT Lincoln Laboratory), Tom Maiden (Pittsburgh Supercomputing Center), Lorna Rivera (Georgia Institute of Technology), Scott Callaghan (University of Southern California), Weronika Filingier (University of Edinburgh)

The HPC community has consistently identified public outreach and education of new and existing community members as vital to the growth of the field. New strategies and alternative teaching methods are needed, not only to improve access, promote diversity and attract new HPC practitioners, but also to encourage the next generation of scientists and gain the support of the general public. To help improve education and outreach efforts, this BoF will host an interactive discussion, demonstrations of existing training and outreach activities, and examination of alternative teaching approaches such as MOOCs and virtual workshops. Come and try them for yourself!

Room: 301-302-303
5:15 pm - 7:00 pm

Reconfigurable Supercomputing (RSC)

Alan George (University of Pittsburgh), Herman Lam (University of Florida), Martin Herboldt (Boston University)

Reconfigurable Supercomputing (RSC) is characterized by hardware that adapts to match the needs of each application, offering unique advantages in performance per unit energy for high-end computing. 2017 continues breakout for RSC. Last year's highlights included datacenter deployment by Microsoft, acquisition of Altera by Intel, and successful large-scale RSC in the NSF CHREC Center. This year is highlighted by deployment of publicly available RSC nodes and clusters by AWS and Baidu. This BoF introduces architectures of such systems, describes applications and tools being developed, and provides a forum for discussing emerging opportunities and issues for performance, productivity, and sustainability.

Room: 402-403-404
5:15 pm - 7:00 pm

The Green500: Trends in Energy-Efficient Supercomputing

Erich Strohmaier (Lawrence Berkeley National Laboratory), Natalie Bates (Energy Efficient HPC Working Group), Thomas Scogland (Lawrence Livermore National Laboratory), Wu Feng (Virginia Tech)

With power becoming a first-order design constraint on-par with performance, it is important to measure and analyze energy-efficiency trends in supercomputing. To raise the awareness of greenness as a first-order design constraint, the Green500 seeks to characterize the energy-efficiency of supercomputers for different metrics, workloads, and methodologies. This BoF discusses trends across the Green500 and highlights from the current Green500 list. In addition, the Green500, Top500, and Energy-Efficient HPC Working Group have been working together on improving power-measurement methodology and this BoF presents case studies from sites that have made power submissions that meet the highest quality of measurement methodology.

Room: 405-406-407
5:15 pm - 7:00 pm

Lustre Community BoF: Lustre Deployments for the Next 5 Years

Sarp Oral (Oak Ridge National Laboratory), Frank Baetke (Hewlett Packard, European Open File System)

Lustre is the leading open source file system for HPC and is widely used and in more mission-critical installations. Lustre currently supports many HPC infrastructures beyond its traditional stronghold of scientific research including financial services, oil and gas,

advanced manufacturing, and visual effects. At this year's Community BoF, the worldwide community of developers, administrators, and solution providers will gather to discuss the current status of the Lustre community, OpenSFS and EOFS organizations, and also discuss recent feature developments, new challenges and corresponding opportunities for the next 5 years of Lustre. Summaries of LUG2017 and LAD 2017 will be presented.

Room: 501-502
5:15 pm - 7:00 pm

OpenSHMEM in the Era of Exascale

Barbara Chapman (Stony Brook University), Tony Curtis (Stony Brook University), Oscar Hernandez (Oak Ridge National Laboratory), Vivek Sarkar (Georgia Institute of Technology), Jeff Kuehn (Los Alamos National Laboratory), Sameer Shende (ParaTools), Pavel Shamis (ARM Ltd), Manjunath Venkata (Oak Ridge National Laboratory), Steve Poole (Los Alamos National Laboratory)

OpenSHMEM is a PGAS API for single-sided asynchronous scalable communications in HPC applications. OpenSHMEM is a community driven standard for the SHMEM API across multiple architectures/implementations. This BoF brings together the OpenSHMEM community to present the latest accomplishments since release of the 1.3 specification, and discuss future directions for the OpenSHMEM community as we develop version 1.4. The BoF will consist of talks from end-users, implementers, middleware and tool developers to discuss their experiences and plans for using OpenSHMEM. We will then open the floor for discussion of the specification and our mid-to-long term goals.

Room: 503-504
5:15 pm - 7:00 pm

Open MPI State of the Union XI

George Bosilca (University of Tennessee), Jeffrey Squyres (Cisco Systems)

Open MPI continues to drive the state of the art in HPC. This year, we've added new features, fixed bugs, improved performance, and collaborated with many across the HPC community. We'll discuss what Open MPI has accomplished over the past year and present a roadmap for the next year.

One of Open MPI's strength lies in its diversity: we represent many different viewpoints across the HPC ecosystem. To that end, many developers from the community will be present to discuss and answer your questions both during and after the BoF.

Room: 507
5:15 pm - 7:00 pm

OpenStack For HPC: Best Practices for Optimizing Software-Defined Infrastructure

Stig Telfer (StackHPC Ltd), Martial Michel (Data Machines Corp), Mike Lowe (Indiana University), Timothy Randles (Los Alamos National Laboratory), Robert Budden (Pittsburgh Supercomputing Center), Benjamin Lynch (University of Minnesota), Bruno Silva (Francis Crick Institute), Tom King (Queen Mary University of London), Simon Thompson (University of Birmingham), Blair Bethwaite (Monash University)

OpenStack has become an ideal solution for managing HPC infrastructure and workloads. However, performance overheads, network integration, and system complexity pose daunting architectural challenges. High performance parallel file systems and the integration of HPC resources into on-premise hybrid clouds allow organizations to increase agility and reduce cost. Using OpenStack technology to control large pools of compute, storage and networking resources throughout the data center allows IT organizations to reduce storage cost while maintaining performance and utilization as it scales. This BoF is aimed at architects, administrators, software engineers, and scientists interested in designing and deploying OpenStack infrastructure for HPC.

Room: 601
5:15 pm - 7:00 pm

National Strategic Computing Initiative Update

Irene Qualters (National Science Foundation), Mark Sims (US Department of Defense), William Harrod (US Department of Energy)

The National Strategic Computing Initiative (NSCI) focuses the efforts of key US federal agencies into a unified strategy, with five key objectives, to ensure development and deployment of future generations of high-performance computing (HPC) systems. As the reliable progress in computer performance afforded by Moore's Law approaches its end, this initiative must investigate technologies and approaches for providing new sustainable pathways of both increased capability and capacity. The presenters in this session will describe their agency's plans and progress towards achieving NSCI's objectives. Discussions will contribute to furthering NSCI investment plans and identify opportunities for future global collaboration.

Room: 603
5:15 pm - 7:00 pm

HPC in Space: Supercomputing at 17,500 MPH

John Kichury (Hewlett Packard Enterprise), Eng Lim Goh (Hewlett Packard Enterprise), Mark Fernandez (Hewlett Packard Enterprise)

Spaceborne Computer intends to run a year-long high performance computing (HPC) experiment utilizing a commercial off-the-shelf (COTS) computer system on the International Space Station (ISS). This research has the potential to identify methods of using software to protect computers from the harsh environment of space without expensive, time-consuming or bulky protective shielding. This original experimental process is referred to as "HARDening with SOFTWARE". The Spaceborne Computer software architecture will be presented; the current status of the experiment will be shared; and a request from the audience for additional potential uses of HPC in space will be solicited.

Room: 605
5:15 pm - 7:00 pm

Machine Learning for Big Data: Integrated, Collaborative, Multi-Technological Solutions to Multi-Objective Problems

Marie-Christine Sawley (Intel Corporation), Sofia Vallecorsa (CERN)

The topic of this BoF is to explore the growing interaction between Machine Learning and Big Data. In particular, it will focus on how integrated collaborative platforms, multi-architectures frameworks, and cloud computing enable the solution of multi-objective problems (i.e. multivariate regression, sequence and multitask learning, multi-label classification).

The session is intended to bring together data scientists and researchers interested in the application of Big Data technology for Machine Learning in their respective fields but also technology experts, in order to highlight innovative strategies to boost collaboration and to solve more complex problems. The chosen format is a session of talks.

Room: 607
5:15 pm - 7:00 pm

A Comparison of Distributed Memory Fast Fourier Transform (FFT) Library Packages

Benson Muite (University of Tartu), Samar Aseeri (King Abdullah University of Science and Technology)

The FFT is used in many different applications but has poor scalability due to its high communication volume. It is used because of its low computational cost when accurate results are required. There are many methods to use for implementing Fourier Transforms, and it is not clear what will work best on a given distributed platform. To allow for comparisons, FFT developers and users will discuss design considerations and requirements of FFT libraries. It is an opportunity to learn about packages from developers, to be able to differentiate between the libraries, and to determine best FFT library design practices.

Room: 703
5:15 pm - 7:00 pm

High Performance Computing Education in US Data Science

Kelly Gaither (Texas Advanced Computing Center, University of Texas), Daniel Katz (University of Illinois; National Center for Supercomputing Applications, University of Illinois), Ann Stapleton (University of North Carolina, Wilmington), Gopalan Oppiliappan (Intel Corporation), Mark Speck (Chaminade University of Honolulu), Hui Zhang (University of Louisville), Rosalia Gomez (Texas Advanced Computing Center, University of Texas), Weijia Xu (Texas Advanced Computing Center, University of Texas)

A key property of Data Science is the adoption of new techniques and tools for conquering extremely large datasets using the latest computing infrastructure. However, challenges remain in integrating high performance computing (HPC) knowledge and hands-on practices into Data Science education. This BoF will consist of a panel of experts who will discuss challenges and needs of HPC education in Data Science programs and have a lively discussion with the audience to explore viable approaches. The ultimate goal is to bring HPC specialists, data scientists, and educators together to broaden HPC education and practice in Data Science.

Room: 709
5:15 pm - 7:00 pm

Building Efficient Clouds for HPC, Big Data, and Deep Learning Middleware and Applications

Xiaoyi Lu (Ohio State University), Dhabaleswar Panda (Ohio State University)

Cloud Computing platforms have become the desired environments for running HPC, Big Data, and Deep Learning workloads. The community is seeing the opportunities and challenges of designing high-performance HPC/Big Data/Deep Learning runtime over clouds. This BoF will involve all the speakers and the audience to identify the most critical challenges facing the community and come up with a roadmap for the next 5-10 years in building efficient clouds with virtual-machines/containers for running HPC/Big Data/Deep Learning workloads. In-depth overviews of virtualization system software, high-performance communication and I/O mechanisms, and example applications on HPC clouds will be discussed.

Room: 710-712
5:15 pm - 7:00 pm

Big Data and Exascale Computing (BDEC) Community Report

Mark Asch (University of Picardie, Total SA), Peter Beckman (Argonne National Laboratory), Jack Dongarra (University of Tennessee)

The emergence of large scale data analytics in a wide variety of scientific fields, and the explosive growth of data generated in edge environments by new instruments and IoT, has disrupted the landscape on which plans for exascale computing are developing. The international Big Data and Extreme-scale Computing (BDEC) workshop series is systematically mapping out the ways that the major issues associated with data intensive science interact with plans for exascale systems. This meeting will present their report on this effort and elicit community input on the development of plans for the convergence on a common software ecosystem for science.

Thursday, November 16th

Room: 201-203
12:15 pm - 1:15 pm

Slurm User Group Meeting

Danny Auble (SchedMD LLC), Morris Jette (SchedMD LLC)

Slurm is an open source workload manager used many on TOP500 systems and provides a rich set of features including topology aware optimized resource allocation, the ability to expand and shrink jobs on demand, the ability to power down idle nodes and restart them as needed, hierarchical bank accounts with fair-share job prioritization and many resource limits. The meeting will consist of three parts: The Slurm development team will present details about changes in the new version 17.11, describe the Slurm roadmap, and solicit user feedback. Everyone interested in Slurm use and/or development is encouraged to attend.

Room: 205-207
12:15 pm - 1:15 pm

HPC Education: Meeting of the SIGHPC Education Chapter

Fernanda Foertter (Oak Ridge National Laboratory), Steve Gordon (Ohio Supercomputer Center), Richard Coffey (Argonne National Laboratory)

This BoF will explore the SC17 #hpcconnects themes and bring together practitioners from traditional and emerging fields to explore educational needs in HPC. Current officers will also review past and present activities of the chapter.

Room: 210-212
12:15 pm - 1:15 pm

Charting the PMIx Roadmap

David Solt (IBM), Artem Polyakov (Mellanox Technologies), Ralph Castain (Intel Corporation)

The PMI Exascale (PMIx) community will be concluding its third year of existence this fall that included release of the PMIx v2.0 standard. We'll discuss what PMIx has accomplished over the past year and present a proposed roadmap for next year.

The PMIx community includes viewpoints from across the HPC runtime community. To that end, we solicit feedback and suggestions on the roadmap in advance of the session, and will include time for a lively discussion at the meeting.

So please join us at the BoF to plan the roadmap. New contributors are welcome!

Room: 301-302-303
12:15 pm - 1:15 pm

Charm++ and AMPI: Adaptive and Asynchronous Parallel Programming

Laxmikant Kalé (University of Illinois, Charmworks Inc), Sam White (Charmworks Inc, University of Illinois), Phil Miller (Charmworks Inc)

A community gathering about parallel programming using Charm++, Adaptive MPI, the many applications built on them, and associated tools. This session will cover recent advances in Charm++ and the experiences of application developers with Charm++. There will also be a discussion on the future directions of Charm++ and opportunities to learn more and form collaborations.

Charm++ is a production-grade many-tasking programming framework and runtime system for modern HPC systems. It offers high productivity and performance portability through features such as multicore and accelerator support, dynamic load balancing, fault tolerance, latency hiding, interoperability with MPI and OpenMP, and online job-resizing.

Room: 402-403-404
12:15 pm - 1:15 pm

Analyzing Parallel I/O

Julian Kunkel (German Climate Computing Center), Philip Carns (Argonne National Laboratory)

Parallel application I/O performance often fails to meet user expectations. In addition, subtle changes in access patterns may lead to significant changes in performance due to complex interactions between hardware and software. These challenges call for sophisticated tools to capture, analyze, understand, and tune application I/O.

In this BoF, we will highlight recent advances in monitoring and characterization tools to help address this problem. We will also encourage community discussion to compare best practices, identify gaps in measurement and analysis, and find ways to translate parallel I/O analysis into actionable outcomes for users, facility operators, and researchers.

Room: 405-406-407
12:15 pm - 1:15 pm

Sharing Research Data: Data Commons, Distributed Clouds, and Distributed Data Services

Robert Grossman (University of Chicago)

Data commons collocate data, storage, and computing infrastructure with core data services and commonly used tools and applications for managing, analyzing, and sharing data to create an interoperable resource for the research community. This session will discuss practical experiences designing, building and operating data commons for the research community. It will also discuss key services that data commons require, such as index services, metadata services, etc.

Room: 501-502
12:15 pm - 1:15 pm

Machine Learning for Parallel Performance Analytics

Felix Wolf (Technical University Darmstadt), Hans-Christian Hoppe (Intel Corporation)

Parallel performance analysis tools are a mature field, with several highly scalable and capable tool families available. The objective of automating the analysis process and creating insights from the sea of raw data has not been fully achieved yet. Modern techniques in machine learning and AI (such as deep learning networks) might provide the means to develop highly automated, easy to use performance tools. This BoF assembles experts from the performance tools community and discusses how AI/ML techniques could be taken up to achieve this goal and usher in the era of automatic parallel performance analytics.

Room: 503-504
12:15 pm - 1:15 pm

SAGE2 9th Annual International SC BOF: Scalable Amplified Group Environment for Global Collaboration

Maxine Brown (University of Illinois, Chicago), Luc Renambot (University of Illinois, Chicago), Jason Leigh (University of Hawaii, Manoa)

SAGE2 (Scalable Amplified Group Environment) is an innovative, open-source, web-based, user-centered platform for small groups

or large distributed teams to share and jointly discuss digital media content on tiled displays in order to glean insights and discoveries. SAGE2 treats tiled displays as a seamless ultra-high-resolution desktop, enabling users to juxtapose content from various sources. SAGE2, like its predecessor SAGE (Scalable Adaptive Graphics Environment), is the de facto operating system for managing Big Data on tiled walls, providing the scientific community with persistent visualization and collaboration services for global cyberinfrastructure.

Room: 507

12:15 pm - 1:15 pm

BeeGFS - Architecture, Implementation Examples, and Future Development

Frank Baetke (Hewlett Packard Enterprise), Sven Breuner (ThinkParQ GmbH)

BeeGFS, an open source parallel file system has been gaining more and more acceptance in HPC and associated areas like deep learning, life sciences, and financial services areas and is in use at several TOP500 systems. It is designed to combine key aspects as high scalability, flexibility, and usability with support for mixed ISAs, including ARM and OpenPOWER. At the BoF, an architectural overview, interesting implementations examples, features of upcoming version 7, and further roadmap details will be provided. The organizers encourage participation of all who are interested in high performance parallel storage and feedback from users and solution providers.

Room: 605

12:15 pm - 1:15 pm

Recruitment: How to Build Diverse Teams

Lorna Rivera (Georgia Institute of Technology), Toni Collis (University of Edinburgh)

Diverse teams are now generally accepted as a 'good idea'. However, a common refrain during discussions on improving diversity focuses on the lack of women applying for jobs. This session will address improving recruitment practices to reach more candidates from underrepresented groups. And will presentations from a panel of experts, followed by discussions on how to address the concerns of attendees, providing clearly defined steps for improving the recruitment process. This session invites participation from everyone in the HPC community, to help us gather knowledge from across the HPC workforce and provide clear guidance for steps to recruitment success.

Room: 703

12:15 pm - 1:15 pm

HPC Systems Monitoring Data in Action

Mike Showerman (National Center for Supercomputing Applications, University of Illinois), Ann Gentile (Sandia National Laboratories), Jim Brandt (Sandia National Laboratories)

New opportunities in the use of monitoring data include experimental measurement of science throughput, data driven approaches to systems architectural design decisions, real time analysis of system/application state and function, and long term trend analysis. We invite application users and developers, system architects, and system administrators to provide new perspectives. Panelists using extensive monitoring data will present new insights from their systems. They will interact with the audience to share experiences and foster collaborations on tools and designs. We will identify areas of overlap or potential joint activity to steer future analysis development and use of data.

Room: 708

12:15 pm - 1:15 pm

MPICH: A High-Performance Open-Source MPI Implementation

Pavan Balaji (Argonne National Laboratory), Abdelhalim Amer (Argonne National Laboratory), Yanfei Guo (Argonne National Laboratory), Neelima Bayyapu (Argonne National Laboratory), Kavitha Madhu (Argonne National Laboratory), Ken Raffenetti (Argonne National Laboratory)

MPICH is a widely used, open-source implementation of the MPI message passing standard. It has been ported to many platforms and used by several vendors and research groups as the basis for their own MPI implementations. This BoF session will provide a forum for users of MPICH as well as developers of MPI implementations derived from MPICH to discuss experiences and issues in using and porting MPICH. Future plans for MPICH will be discussed. Representatives from MPICH-derived implementations will provide brief updates on the status of their efforts. MPICH developers will also be present for an open forum discussion.

Room: 710-712

12:15 pm - 1:15 pm

The Ethernet Portfolio for HPC

John D'Ambrosia (Ethernet Alliance, Huawei)

Over the past 15 years, Ethernet has been heavily leveraged as the interconnect backbone for many HPC systems that have shown up on the Top500. Demand for increased bandwidth has driven the development of new Ethernet solutions operating at rates up to 400 Gb/s. The introduction of these standards will drive the performance of tomorrow's networks to never-before-seen levels and bring about the Next Ethernet Era. This session will provide the audience with an overview of the Ethernet solutions and technologies that will be leveraged for the next generation of computing, networking, and storage, and discuss their deployment.

Doctoral Showcase [\(back to top\)](#)

Tuesday, November 14th

Room: Four Seasons Ballroom
10:00 am - 5:15 pm

Exhibition of Doctoral Showcase Posters

Hierarchical Sparse Graph Computations on Multicore Platforms

Humayun Kabir (Pennsylvania State University)

k-core and k-truss are cohesive subgraphs used for visualization, community detection and centrality analysis. These cohesive subgraphs have hierarchical structures and can be computed exactly in polynomial time. We have developed PKC and PKT algorithms for k-core and k-truss decompositions respectively. PKC decreases atomic operations and uses less memory. PKT uses a well-designed data structure that updates trusses concurrently, is memory efficient and avoids using hash-table. On 24 threads, PKT is more than one magnitude faster than a state-of-art algorithm. Sparse matrix computations are useful for scientific computing and engineering. We have developed a multilevel data structure CSR-k to store a sparse matrix that enhances locality in memory access pattern and decreases work load imbalance. On a 32 core node, sparse matrix-vector multiplication using CSR-k is 2.41x faster than state-of-art pOSKI and sparse triangular solution using CSR-k is 2.18x faster than coloring method.

Runtime Solutions to Apply Non-Volatile Memories in Future Computer Systems

Hoda Aghaei Khouzani (University of Delaware)

Memory system is an indispensable part of any computer system from large supercomputers to embedded devices. Its performance, energy, capacity, and management algorithms must scale as computer systems scale to ensure the desired performance growth and the possibility of running new applications. Unfortunately, observable trends of current memory technologies from on-chip Static-RAM (SRAM) caches to Dynamic-RAM (DRAM) main memory and Hard Disk Drive (HDD) storage are failing to respond to these existing and emerging requirements, which lead researchers in both academia and industry to look for alternative solutions. Many promising non-volatile memory (NVM) technologies are introduced to replace current devices on various levels of memory hierarchy. Due to the extent of memory hierarchy subject and the diversity of NVM technologies, the main focus of this study is main memory. Specifically, it explores the feasibility of applying two of promising NVMs, namely, phase change memory (PCM) and domain wall memory (DWM).

Modeling and Comparison of Large-Scale Interconnect Designs

Md Atiqul Mollah (Florida State University)

Modern day high performance computing (HPC) clusters and data centers require a large number of computing and storage elements to be interconnected. Interconnect performance is one of the major contributing factors to the overall performance of such systems. Due to the massive scale of the network, interconnect designs are often evaluated and compared through models. My doctoral research is focused on modeling large-scale interconnections and applying such models to investigate the effect of different topology designs and routing scheme designs on the interconnect performance.

Performance Prediction Modeling of GPU Applications

Marcos Amaris (University of Sao Paulo)

Given the large number of GPU architectures and the many different possibilities to execute an algorithm, the prediction of application execution times over GPUs is a great challenge and is essential for efficient JMSs. The available GPU performance modeling solutions are very sensitive to applications and platform changes,

Here a summary of two main works is shown. In the first work, we present the comparison of a developed BSP-based model to three different ML techniques, this comparison was done with 9 well-known matrix/vector applications. In this work, we wanted to perform a fair comparison, for this reason, we decided that ML process would had the same features that the BSP-based model.

In the second work, we have compared among ML techniques. Here, a two step of extraction features are done. First a correlation analysis and after hierarchical clustering analysis. In this second work, 10 irregular CUDA kernels were used.

Nornir: A Power-Aware Runtime Support for Parallel Applications

Daniele De Sensi (University of Pisa)

Power consumption is becoming a key factor in designing applications and computing systems.

In this thesis we propose Nornir, a runtime system for parallel applications, which can be used by application users to enforce power consumption or performance constraints on their applications, by operating on several knobs like: number of used cores, clock frequency of the cores, etc... Nornir is fully customizable and can be used by autonomic strategies designers to plug in new reconfiguration strategies and to validate them against existing ones.

In this thesis we also designed different reconfiguration algorithms: i) an heuristic to enforce performance constraints; ii) an online learning algorithms to enforce both performance and power consumption constraints; iii) a "concurrency control" algorithm to optimize performance and power consumption of single-producer, single-consumer queues.

All these algorithms have been implemented and validated by using Nornir.

A Heterogeneous HPC Platform for Ill-Structured Spatial Join Processing

Danial Aghajarian (Georgia State University)

Given two layers of large polygonal datasets, detecting those pairs of cross-layer polygons which satisfy a join predicate, such as intersection, is one of the most computationally intensive primitive operations in the spatial domain applications. There are alternative solutions for this ill-structured problem in literature. However, none of them is designed to take advantage of heterogeneous clusters equipped with GPU accelerators to process big spatial data efficiently.

In this research, we propose a distributed heterogeneous HPC platform for spatial join processing based on two-step filter and refinement approach. This work includes two main parts. First, we introduce a set of novel GPU-suited data structures and algorithms. Proof of correctness and analysis is also provided for each algorithm. Second, by applying these new techniques, we propose several GPU-based spatial join systems for various operations such as ST_intersect and polygon overlay to improve the performance of current state of the art systems.

Designing and Building Efficient HPC Cloud with Modern Networking Technologies on Heterogeneous HPC Clusters

Jie Zhang (Ohio State University)

Cloud Computing platforms through virtualization technologies are widely used by many users due to their high availability and flexibility. However, running HPC applications on the cloud still suffers from fairly low performance, more specifically, the degraded I/O performance from the virtualized I/O devices. SR-IOV solves the problem by delivering near-native I/O performance. Nevertheless, SR-IOV lacks high-performance virtualization support, such as locality-aware and NUMA-aware support, and also prevents VM live migration. Moreover, the critical HPC resources among VMs, (e.g, SR-IOV enabled Virtual Functions), need to be managed and isolated to efficiently support running multiple concurrent MPI jobs on HPC clouds. Therefore, we propose a framework to design and build the efficient HPC clouds with modern networking technologies on heterogeneous HPC clusters. Through this framework, the HPC cloud can deliver near-native performance for MPI applications on different types of virtualization environments, support SR-IOV enabled VM migration, and provide the efficient resource sharing.

Productivity and Software Development Effort Estimation in HPC

Sandra Wienke (RWTH Aachen University)

Relying vast HPC investments on an informed decision making process is important when serving the ever increasing demands for computational power. Providing a quantitative metric to compare and evaluate HPC systems in procurement processes, I define a productivity model with predictive power that focuses on the number of simulation application runs and the total cost of ownership (TCO). As part of TCO, the software development costs determined by efforts to parallelize, port or tune simulation codes for novel HPC setups must be predicted. Since approaches from mainstream software engineering does not directly suite HPC needs, I set up a methodology to estimate software development effort in HPC. Proof of concepts are mostly based on real-world HPC setups at RWTH Aachen University.

Reducing Communication Costs in the Parallel Algebraic Multigrid

Amanda J. Bienz (University of Illinois)

Algebraic multigrid is an optimal sparse linear solver that suffers from poor parallel scalability due to large communication requirements. This costly communication can be improved through changes to both the algorithm and parallel implementation. The amount of data required to be communicated can be reduced through sparsification of the matrices, yielding slower convergence but significantly cheaper iterations. Furthermore, the data that must be communicated can be sent between processes with increased efficiency through the use of topology-aware methods. The number of messages injected into the network can be reduced greatly at the cost of additional intra-node communication, yielding cheaper communication on coarse levels of AMG.

Speeding Up GPU Graph Processing Using Structural Graph Properties

Merijn Elwin Verstraaten (University of Amsterdam)

Due to its wide applicability and flexibility, graph processing is an increasingly important part of data science. To scale complex graph analytics computations to large datasets it is becoming popular to utilise accelerator-based architectures, such as Graphical Processing Units (GPUs).

Mapping irregular graph algorithms to hardware designed for highly regular parallelism is a complex task. There are often multiple ways to parallelise the same operation on the GPU. Which of these parallelisation strategies is the fastest is dependent on the structure of the input graph. Performance differences can be an order of magnitude or more, and the optimal strategy varies from graph to graph.

The goal of my PhD research is to identify how structural properties impact the performance of different strategies and use this information to speed-up GPU graph processing by predicting the fastest parallelisation of an algorithm for a specific input graph.

Mitigating Variability in HPC Systems and Applications for Performance and Power Efficiency

Bilge Acun (University of Illinois)

Power consumption and process variability are two important, interconnected, challenges of future generation large-scale HPC data-centers. Current production petaflop supercomputers consume more than 10 megawatts of power that costs millions of dollars every year. As HPC moves towards exascale, power consumption is expected to become a major concern. Not solely dynamic behavior of HPC applications (such as irregular or imbalanced applications) but also dynamic behavior of HPC systems (such as thermal, power, frequency variations among processors) makes it challenging to optimize the performance and power efficiency of large scale applications. Smart and adaptive runtime systems have great potential to handle these challenges transparently from the application.

In my thesis, I first analyze frequency, temperature, and power variations in large-scale HPC systems using thousands of cores and different applications. After I identify the cause of these variations, I propose solutions to mitigate them to improve performance and power efficiency.

Using Runtime Energy Optimizations to Improve Energy Efficiency in High Performance Computing

Sridutt Bhalachandra (University of North Carolina)

Energy efficiency in high performance computing (HPC) will be critical to limit operating costs and carbon footprints in future supercomputing centers. With both hardware and software factors affecting energy usage there exists a need for dynamic power regulation. This dissertation highlights an adaptive runtime framework that can allow processors capable of per-core specific power control to reduce power with little performance impact by dynamically adapting to workload characteristics. Monitoring of performance and power regulation is done transparently within MPI runtime and no code changes are required in the underlying application. In the presence of workload imbalance, the runtime reduces the frequency on cores not on the critical path thereby reducing power without deteriorating performance. This is shown to reduce run-to-run performance variation and improve performance in certain scenarios. For applications plagued by memory related issues, new memory metrics are identified that facilitate lowering power without adversely impacting performance.

Bounded Asynchrony and Nested Parallelism for Scalable Graph Processing

Adam Fidel (Texas A&M University)

Processing large-scale graphs has become a critical component in a variety of fields, from scientific computing to social analytics. The irregular access pattern for graph workloads, coupled with complex graph structures and large data sizes makes efficiently executing parallel graph workloads challenging. In this dissertation, we develop two broad techniques for improving the performance of general parallel graph algorithms: bounded asynchrony and nested parallelism. Increasing asynchrony in a bounded manner allows one to avoid costly global synchronization at scale, while still avoiding the penalty of unbounded asynchrony including redundant work. Using this technique, we scale a BFS workload to 98,304 cores where traditional methods stop scaling. Additionally, asynchrony enables a new family of approximate algorithms for applications tolerant to fixed amounts of error. Representing graph algorithms in a nested parallel manner enables the full use of available parallelism inherent in graph algorithms, while efficiently managing communication through nested sections.

Computing Architectures Exploiting Optical Interconnect and Optical Memory Technologies

Pavlos Maniotis (Aristotle University of Thessaloniki)

The long-standing "Bandwidth" and "Memory-Wall" problems hinder exascale computing within the available energy boundaries. This work proposes new optical-systems for use in innovative Computing Architectures towards increasing bandwidth and reducing energy-consumption. The initial efforts were devoted towards elevating optical-memories from the so-far single-bit layouts to complete multi-bit optical-cache setups, demonstrating the first optical-cache blueprint at 16Gb/s. The optical-cache credentials are highlighted by proposing an optically enabled Chip-Multiprocessor architecture; the shared single-level optical-cache approach negates the need for complex cache-hierarchy and demonstrates 20% speed-up for PARSEC benchmarks. Extending optics utilization to complete HPC environments, OptoHPC-Sim simulator was developed for supporting system-scale utilization of optical-interconnect and electro-optical routing technologies. Using OptoHPC-Sim, an OptoBoard-based HPC-network is proposed demonstrating 190% mean throughput improvement and 83% mean packet-delay reduction compared to world's #3 HPC. Finally, all this experience led to the design of the first optical Ternary-CAM-row architecture at 20Gb/s for fast address look-up configurations.

Room: 210-212
10:30 am - 12:00 pm

Doctoral Showcase Session 1

Designing and Building Efficient HPC Cloud with Modern Networking Technologies on Heterogeneous HPC Clusters

Jie Zhang (Ohio State University)

Cloud Computing platforms through virtualization technologies are widely used by many users due to their high availability and flexibility. However, running HPC applications on the cloud still suffers from fairly low performance, more specifically, the degraded I/O performance from the virtualized I/O devices. SR-IOV solves the problem by delivering near-native I/O performance. Nevertheless, SR-IOV lacks high-performance virtualization support, such as locality-aware and NUMA-aware support, and also prevents VM live migration. Moreover, the critical HPC resources among VMs, (e.g, SR-IOV enabled Virtual Functions), need to be managed and isolated to efficiently support running multiple concurrent MPI jobs on HPC clouds. Therefore, we propose a framework to design and build the efficient HPC clouds with modern networking technologies on heterogeneous HPC clusters. Through this framework, the HPC cloud can deliver near-native performance for MPI applications on different types of virtualization environments, support SR-IOV enabled VM migration, and provide the efficient resource sharing.

Runtime Solutions to Apply Non-Volatile Memories in Future Computer Systems

Hoda Aghaei Khouzani (University of Delaware)

Memory system is an indispensable part of any computer system from large supercomputers to embedded devices. Its performance, energy, capacity, and management algorithms must scale as computer systems scale to ensure the desired performance growth and the possibility of running new applications. Unfortunately, observable trends of current memory technologies from on-chip Static-RAM (SRAM) caches to Dynamic-RAM (DRAM) main memory and Hard Disk Drive (HDD) storage are failing to respond to these existing and emerging requirements, which lead researchers in both academia and industry to look for alternative solutions. Many promising non-volatile memory (NVM) technologies are introduced to replace current devices on various levels of memory hierarchy. Due to the extent of memory hierarchy subject and the diversity of NVM technologies, the main focus of this study is main memory. Specifically, it explores the feasibility of applying two of promising NVMs, namely, phase change memory (PCM) and domain wall memory (DWM).

Reducing Communication Costs in the Parallel Algebraic Multigrid

Amanda J. Bienz (University of Illinois)

Algebraic multigrid is an optimal sparse linear solver that suffers from poor parallel scalability due to large communication requirements. This costly communication can be improved through changes to both the algorithm and parallel implementation. The amount of data required to be communicated can be reduced through sparsification of the matrices, yielding slower convergence but significantly cheaper iterations. Furthermore, the data that must be communicated can be sent between processes with increased efficiency through the use of topology-aware methods. The number of messages injected into the network can be reduced greatly at the cost of additional intra-node communication, yielding cheaper communication on coarse levels of AMG.

Modeling and Comparison of Large-Scale Interconnect Designs

Md Atiqul Mollah (Florida State University)

Modern day high performance computing (HPC) clusters and data centers require a large number of computing and storage elements to be interconnected. Interconnect performance is one of the major contributing factors to the overall performance of such systems. Due to the massive scale of the network, interconnect designs are often evaluated and compared through models. My doctoral research is focused on modeling large-scale interconnections and applying such models to investigate the effect of different topology designs and routing scheme designs on the interconnect performance.

Computing Architectures Exploiting Optical Interconnect and Optical Memory Technologies

Pavlos Maniotis (Aristotle University of Thessaloniki)

The long-standing "Bandwidth" and "Memory-Wall" problems hinder exascale computing within the available energy boundaries. This work proposes new optical-systems for use in innovative Computing Architectures towards increasing bandwidth and reducing energy-consumption. The initial efforts were devoted towards elevating optical-memories from the so-far single-bit layouts to complete multi-bit optical-cache setups, demonstrating the first optical-cache blueprint at 16Gb/s. The optical-cache credentials are highlighted by proposing an optically enabled Chip-Multiprocessor architecture; the shared single-level optical-cache approach negates the need for complex cache-hierarchy and demonstrates 20% speed-up for PARSEC benchmarks. Extending optics utilization to complete HPC environments, OptoHPC-Sim simulator was developed for supporting system-scale utilization of optical-interconnect and electro-optical routing technologies. Using OptoHPC-Sim, an OptoBoard-based HPC-network is proposed demonstrating 190% mean throughput improvement and 83% mean packet-delay reduction compared to world's #3 HPC. Finally, all this experience led to the design of the first optical Ternary-CAM-row architecture at 20Gb/s for fast address look-up configurations.

Room: 210-212
1:30 pm - 3:00 pm

Doctoral Showcase Session 2

Mitigating Variability in HPC Systems and Applications for Performance and Power Efficiency

Bilge Acun (University of Illinois)

Power consumption and process variability are two important, interconnected, challenges of future generation large-scale HPC data-centers. Current production petaflop supercomputers consume more than 10 megawatts of power that costs millions of dollars every year. As HPC moves towards exascale, power consumption is expected to become a major concern. Not solely dynamic behavior of HPC applications (such as irregular or imbalanced applications) but also dynamic behavior of HPC systems (such as thermal, power, frequency variations among processors) makes it challenging to optimize the performance and power efficiency of large scale applications. Smart and adaptive runtime systems have great potential to handle these challenges transparently from the application.

In my thesis, I first analyze frequency, temperature, and power variations in large-scale HPC systems using thousands of cores and different applications. After I identify the cause of these variations, I propose solutions to mitigate them to improve performance and power efficiency.

Nornir: A Power-Aware Runtime Support for Parallel Applications

Daniele De Sensi (University of Pisa)

Power consumption is becoming a key factor in designing applications and computing systems.

In this thesis we propose Nornir, a runtime system for parallel applications, which can be used by application users to enforce power consumption or performance constraints on their applications, by operating on several knobs like: number of used cores, clock frequency of the cores, etc... Nornir is fully customizable and can be used by autonomic strategies designers to plug in new reconfiguration strategies and to validate them against existing ones.

In this thesis we also designed different reconfiguration algorithms: i) an heuristic to enforce performance constraints; ii) an online learning algorithms to enforce both performance and power consumption constraints; iii) a "concurrency control" algorithm to optimize performance and power consumption of single-producer, single-consumer queues.

All these algorithms have been implemented and validated by using Nornir.

Using Runtime Energy Optimizations to Improve Energy Efficiency in High Performance Computing

Sridutt Bhalachandra (University of North Carolina)

Energy efficiency in high performance computing (HPC) will be critical to limit operating costs and carbon footprints in future supercomputing centers. With both hardware and software factors affecting energy usage there exists a need for dynamic power regulation. This dissertation highlights an adaptive runtime framework that can allow processors capable of per-core specific power control to reduce power with little performance impact by dynamically adapting to workload characteristics. Monitoring of performance and power regulation is done transparently within MPI runtime and no code changes are required in the underlying application. In the presence of workload imbalance, the runtime reduces the frequency on cores not on the critical path thereby reducing power without deteriorating performance. This is shown to reduce run-to-run performance variation and improve performance in certain scenarios. For applications plagued by memory related issues, new memory metrics are identified that facilitate lowering power without adversely impacting performance.

Bounded Asynchrony and Nested Parallelism for Scalable Graph Processing

Adam Fidel (Texas A&M University)

Processing large-scale graphs has become a critical component in a variety of fields, from scientific computing to social analytics. The irregular access pattern for graph workloads, coupled with complex graph structures and large data sizes makes efficiently executing parallel graph workloads challenging. In this dissertation, we develop two broad techniques for improving the performance of general parallel graph algorithms: bounded asynchrony and nested parallelism. Increasing asynchrony in a bounded manner allows one to avoid costly global synchronization at scale, while still avoiding the penalty of unbounded asynchrony including redundant work. Using this technique, we scale a BFS workload to 98,304 cores where traditional methods stop scaling. Additionally, asynchrony enables a new family of approximate algorithms for applications tolerant to fixed amounts of error. Representing graph algorithms in a nested parallel manner enables the full use of available parallelism inherent in graph algorithms, while efficiently managing communication through nested sections.

Room: 210-212
3:30 pm - 5:00 pm

Doctoral Showcase Session 3

Productivity and Software Development Effort Estimation in HPC

Sandra Wienke (RWTH Aachen University)

Relying vast HPC investments on an informed decision making process is important when serving the ever increasing demands for computational power. Providing a quantitative metric to compare and evaluate HPC systems in procurement processes, I define a productivity model with predictive power that focuses on the number of simulation application runs and the total cost of ownership (TCO). As part of TCO, the software development costs determined by efforts to parallelize, port or tune simulation codes for novel HPC setups must be predicted. Since approaches from mainstream software engineering does not directly suite HPC needs, I set up a methodology to estimate software development effort in HPC. Proof of concepts are mostly based on real-world HPC setups at RWTH Aachen University.

Performance Prediction Modeling of GPU Applications

Marcos Amarís (University of Sao Paulo)

Given the large number of GPU architectures and the many different possibilities to execute an algorithm, the prediction of application execution times over GPUs is a great challenge and is essential for efficient JMSs. The available GPU performance modeling solutions are very sensitive to applications and platform changes,

Here a summary of two main works is shown. In the first work, we present the comparison of a developed BSP-based model to three different ML techniques, this comparison was done with 9 well-known matrix/vector applications. In this work, we wanted to perform a fair comparison, for this reason, we decided that ML process would had the same features that the BSP-based model.

In the second work, we have compared among ML techniques. Here, a two step of extraction features are done. First a correlation analysis and after hierarchical clustering analysis. In this second work, 10 irregular CUDA kernels were used.

Speeding Up GPU Graph Processing Using Structural Graph Properties

Merijn Elwin Verstraaten (University of Amsterdam)

Due to its wide applicability and flexibility, graph processing is an increasingly important part of data science. To scale complex graph analytics computations to large datasets it is becoming popular to utilise accelerator-based architectures, such as Graphical Processing Units (GPUs).

Mapping irregular graph algorithms to hardware designed for highly regular parallelism is a complex task. There are often multiple ways to parallelise the same operation on the GPU. Which of these parallelisation strategies is the fastest is dependent on the structure of the input graph. Performance differences can be an order of magnitude or more, and the optimal strategy varies from graph to graph.

The goal of my PhD research is to identify how structural properties impact the performance of different strategies and use this information to speed-up GPU graph processing by predicting the fastest parallelisation of an algorithm for a specific input graph.

A Heterogeneous HPC Platform for Ill-Structured Spatial Join Processing

Danial Aghajarian (Georgia State University)

Given two layers of large polygonal datasets, detecting those pairs of cross-layer polygons which satisfy a join predicate, such as intersection, is one of the most computationally intensive primitive operations in the spatial domain applications. There are alternative solutions for this ill-structured problem in literature. However, none of them is designed to take advantage of heterogeneous clusters equipped with GPU accelerators to process big spatial data efficiently.

In this research, we propose a distributed heterogeneous HPC platform for spatial join processing based on two-step filter and refinement approach. This work includes two main parts. First, we introduce a set of novel GPU-suited data structures and algorithms. Proof of correctness and analysis is also provided for each algorithm. Second, by applying these new techniques, we propose several GPU-based spatial join systems for various operations such as ST_intersect and polygon overlay to improve the performance of current state of the art systems.

Hierarchical Sparse Graph Computations on Multicore Platforms

Humayun Kabir (Pennsylvania State University)

k-core and k-truss are cohesive subgraphs used for visualization, community detection and centrality analysis. These cohesive subgraphs have hierarchical structures and can be computed exactly in polynomial time. We have developed PKC and PKT algorithms for k-core and k-truss decompositions respectively. PKC decreases atomic operations and uses less memory. PKT uses a well-designed data structure that updates trusses concurrently, is memory efficient and avoids using hash-table. On 24 threads, PKT is more than one magnitude faster than a state-of-art algorithm. Sparse matrix computations are useful for scientific computing and engineering. We have developed a multilevel data structure CSR-k to store a sparse matrix that enhances locality in memory access pattern and decreases work load imbalance. On a 32 core node, sparse matrix-vector multiplication using CSR-k is 2.41x

faster than state-of-art pOSKI and sparse triangular solution using CSR-k is 2.18x faster than coloring method.

Room: Four Seasons Ballroom
5:15 pm - 7:00 pm

Doctoral Showcase Posters Reception

Hierarchical Sparse Graph Computations on Multicore Platforms

Humayun Kabir (Pennsylvania State University)

k-core and k-truss are cohesive subgraphs used for visualization, community detection and centrality analysis. These cohesive subgraphs have hierarchical structures and can be computed exactly in polynomial time. We have developed PKC and PKT algorithms for k-core and k-truss decompositions respectively. PKC decreases atomic operations and uses less memory. PKT uses a well-designed data structure that updates trusses concurrently, is memory efficient and avoids using hash-table. On 24 threads, PKT is more than one magnitude faster than a state-of-art algorithm. Sparse matrix computations are useful for scientific computing and engineering. We have developed a multilevel data structure CSR-k to store a sparse matrix that enhances locality in memory access pattern and decreases work load imbalance. On a 32 core node, sparse matrix-vector multiplication using CSR-k is 2.41x faster than state-of-art pOSKI and sparse triangular solution using CSR-k is 2.18x faster than coloring method.

Runtime Solutions to Apply Non-Volatile Memories in Future Computer Systems

Hoda Aghaei Khouzani (University of Delaware)

Memory system is an indispensable part of any computer system from large supercomputers to embedded devices. Its performance, energy, capacity, and management algorithms must scale as computer systems scale to ensure the desired performance growth and the possibility of running new applications. Unfortunately, observable trends of current memory technologies from on-chip Static-RAM (SRAM) caches to Dynamic-RAM (DRAM) main memory and Hard Disk Drive (HDD) storage are failing to respond to these existing and emerging requirements, which lead researchers in both academia and industry to look for alternative solutions. Many promising non-volatile memory (NVM) technologies are introduced to replace current devices on various levels of memory hierarchy. Due to the extent of memory hierarchy subject and the diversity of NVM technologies, the main focus of this study is main memory. Specifically, it explores the feasibility of applying two of promising NVMs, namely, phase change memory (PCM) and domain wall memory (DWM).

Modeling and Comparison of Large-Scale Interconnect Designs

Md Atiqul Mollah (Florida State University)

Modern day high performance computing (HPC) clusters and data centers require a large number of computing and storage elements to be interconnected. Interconnect performance is one of the major contributing factors to the overall performance of such systems. Due to the massive scale of the network, interconnect designs are often evaluated and compared through models. My doctoral research is focused on modeling large-scale interconnections and applying such models to investigate the effect of different topology designs and routing scheme designs on the interconnect performance.

Performance Prediction Modeling of GPU Applications

Marcos Amarís (University of Sao Paulo)

Given the large number of GPU architectures and the many different possibilities to execute an algorithm, the prediction of application execution times over GPUs is a great challenge and is essential for efficient JMSs. The available GPU performance modeling solutions are very sensitive to applications and platform changes,

Here a summary of two main works is shown. In the first work, we present the comparison of a developed BSP-based model to three different ML techniques, this comparison was done with 9 well-known matrix/vector applications. In this work, we wanted to perform a fair comparison, for this reason, we decided that ML process would had the same features that the BSP-based model.

In the second work, we have compared among ML techniques. Here, a two step of extraction features are done. First a correlation analysis and after hierarchical clustering analysis. In this second work, 10 irregular CUDA kernels were used.

Nornir: A Power-Aware Runtime Support for Parallel Applications

Daniele De Sensi (University of Pisa)

Power consumption is becoming a key factor in designing applications and computing systems.

In this thesis we propose Nornir, a runtime system for parallel applications, which can be used by application users to enforce power consumption or performance constraints on their applications, by operating on several knobs like: number of used cores, clock frequency of the cores, etc... Nornir is fully customizable and can be used by autonomic strategies designers to plug in new reconfiguration strategies and to validate them against existing ones.

In this thesis we also designed different reconfiguration algorithms: i) an heuristic to enforce performance constraints; ii) an online learning algorithms to enforce both performance and power consumption constraints; iii) a "concurrency control" algorithm to optimize performance and power consumption of single-producer, single-consumer queues.

All these algorithms have been implemented and validated by using Nornir.

A Heterogeneous HPC Platform for Ill-Structured Spatial Join Processing

Danial Aghajarian (Georgia State University)

Given two layers of large polygonal datasets, detecting those pairs of cross-layer polygons which satisfy a join predicate, such as intersection, is one of the most computationally intensive primitive operations in the spatial domain applications. There are alternative solutions for this ill-structured problem in literature. However, none of them is designed to take advantage of heterogeneous clusters equipped with GPU accelerators to process big spatial data efficiently.

In this research, we propose a distributed heterogeneous HPC platform for spatial join processing based on two-step filter and refinement approach. This work includes two main parts. First, we introduce a set of novel GPU-suited data structures and algorithms. Proof of correctness and analysis is also provided for each algorithm. Second, by applying these new techniques, we propose several GPU-based spatial join systems for various operations such as ST_intersect and polygon overlay to improve the performance of current state of the art systems.

Designing and Building Efficient HPC Cloud with Modern Networking Technologies on Heterogeneous HPC Clusters

Jie Zhang (Ohio State University)

Cloud Computing platforms through virtualization technologies are widely used by many users due to their high availability and flexibility. However, running HPC applications on the cloud still suffers from fairly low performance, more specifically, the degraded I/O performance from the virtualized I/O devices. SR-IOV solves the problem by delivering near-native I/O performance. Nevertheless, SR-IOV lacks high-performance virtualization support, such as locality-aware and NUMA-aware support, and also prevents VM live migration. Moreover, the critical HPC resources among VMs, (e.g, SR-IOV enabled Virtual Functions), need to be managed and isolated to efficiently support running multiple concurrent MPI jobs on HPC clouds. Therefore, we propose a framework to design and build the efficient HPC clouds with modern networking technologies on heterogeneous HPC clusters. Through this framework, the HPC cloud can deliver near-native performance for MPI applications on different types of virtualization environments, support SR-IOV enabled VM migration, and provide the efficient resource sharing.

Productivity and Software Development Effort Estimation in HPC

Sandra Wienke (RWTH Aachen University)

Relying vast HPC investments on an informed decision making process is important when serving the ever increasing demands for computational power. Providing a quantitative metric to compare and evaluate HPC systems in procurement processes, I define a productivity model with predictive power that focuses on the number of simulation application runs and the total cost of ownership (TCO). As part of TCO, the software development costs determined by efforts to parallelize, port or tune simulation codes for novel HPC setups must be predicted. Since approaches from mainstream software engineering does not directly suite HPC needs, I set up a methodology to estimate software development effort in HPC. Proof of concepts are mostly based on real-world HPC setups at RWTH Aachen University.

Reducing Communication Costs in the Parallel Algebraic Multigrid

Amanda J. Bienz (University of Illinois)

Algebraic multigrid is an optimal sparse linear solver that suffers from poor parallel scalability due to large communication requirements. This costly communication can be improved through changes to both the algorithm and parallel implementation. The amount of data required to be communicated can be reduced through sparsification of the matrices, yielding slower convergence but significantly cheaper iterations. Furthermore, the data that must be communicated can be sent between processes with increased efficiency through the use of topology-aware methods. The number of messages injected into the network can be reduced greatly at the cost of additional intra-node communication, yielding cheaper communication on coarse levels of AMG.

Speeding Up GPU Graph Processing Using Structural Graph Properties

Merijn Elwin Verstraaten (University of Amsterdam)

Due to its wide applicability and flexibility, graph processing is an increasingly important part of data science. To scale complex graph analytics computations to large datasets it is becoming popular to utilise accelerator-based architectures, such as Graphical Processing Units (GPUs).

Mapping irregular graph algorithms to hardware designed for highly regular parallelism is a complex task. There are often multiple ways to parallelise the same operation on the GPU. Which of these parallelisation strategies is the fastest is dependent on the structure of the input graph. Performance differences can be an order of magnitude or more, and the optimal strategy varies from

graph to graph.

The goal of my PhD research is to identify how structural properties impact the performance of different strategies and use this information to speed-up GPU graph processing by predicting the fastest parallelisation of an algorithm for a specific input graph.

Mitigating Variability in HPC Systems and Applications for Performance and Power Efficiency

Bilge Acun (University of Illinois)

Power consumption and process variability are two important, interconnected, challenges of future generation large-scale HPC data-centers. Current production petaflop supercomputers consume more than 10 megawatts of power that costs millions of dollars every year. As HPC moves towards exascale, power consumption is expected to become a major concern. Not solely dynamic behavior of HPC applications (such as irregular or imbalanced applications) but also dynamic behavior of HPC systems (such as thermal, power, frequency variations among processors) makes it challenging to optimize the performance and power efficiency of large scale applications. Smart and adaptive runtime systems have great potential to handle these challenges transparently from the application.

In my thesis, I first analyze frequency, temperature, and power variations in large-scale HPC systems using thousands of cores and different applications. After I identify the cause of these variations, I propose solutions to mitigate them to improve performance and power efficiency.

Using Runtime Energy Optimizations to Improve Energy Efficiency in High Performance Computing

Sridutt Bhalachandra (University of North Carolina)

Energy efficiency in high performance computing (HPC) will be critical to limit operating costs and carbon footprints in future supercomputing centers. With both hardware and software factors affecting energy usage there exists a need for dynamic power regulation. This dissertation highlights an adaptive runtime framework that can allow processors capable of per-core specific power control to reduce power with little performance impact by dynamically adapting to workload characteristics. Monitoring of performance and power regulation is done transparently within MPI runtime and no code changes are required in the underlying application. In the presence of workload imbalance, the runtime reduces the frequency on cores not on the critical path thereby reducing power without deteriorating performance. This is shown to reduce run-to-run performance variation and improve performance in certain scenarios. For applications plagued by memory related issues, new memory metrics are identified that facilitate lowering power without adversely impacting performance.

Bounded Asynchrony and Nested Parallelism for Scalable Graph Processing

Adam Fidel (Texas A&M University)

Processing large-scale graphs has become a critical component in a variety of fields, from scientific computing to social analytics. The irregular access pattern for graph workloads, coupled with complex graph structures and large data sizes makes efficiently executing parallel graph workloads challenging. In this dissertation, we develop two broad techniques for improving the performance of general parallel graph algorithms: bounded asynchrony and nested parallelism. Increasing asynchrony in a bounded manner allows one to avoid costly global synchronization at scale, while still avoiding the penalty of unbounded asynchrony including redundant work. Using this technique, we scale a BFS workload to 98,304 cores where traditional methods stop scaling. Additionally, asynchrony enables a new family of approximate algorithms for applications tolerant to fixed amounts of error. Representing graph algorithms in a nested parallel manner enables the full use of available parallelism inherent in graph algorithms, while efficiently managing communication through nested sections.

Computing Architectures Exploiting Optical Interconnect and Optical Memory Technologies

Pavlos Maniotis (Aristotle University of Thessaloniki)

The long-standing "Bandwidth" and "Memory-Wall" problems hinder exascale computing within the available energy boundaries. This work proposes new optical-systems for use in innovative Computing Architectures towards increasing bandwidth and reducing energy-consumption. The initial efforts were devoted towards elevating optical-memories from the so-far single-bit layouts to complete multi-bit optical-cache setups, demonstrating the first optical-cache blueprint at 16Gb/s. The optical-cache credentials are highlighted by proposing an optically enabled Chip-Multiprocessor architecture; the shared single-level optical-cache approach negates the need for complex cache-hierarchy and demonstrates 20% speed-up for PARSEC benchmarks. Extending optics utilization to complete HPC environments, OptoHPC-Sim simulator was developed for supporting system-scale utilization of optical-interconnect and electro-optical routing technologies. Using OptoHPC-Sim, an OptoBoard-based HPC-network is proposed demonstrating 190% mean throughput improvement and 83% mean packet-delay reduction compared to world's #3 HPC. Finally, all this experience led to the design of the first optical Ternary-CAM-row architecture at 20Gb/s for fast address look-up configurations.

Wednesday, November 15th

Room: Four Seasons Ballroom

Exhibition of Doctoral Showcase Posters

Hierarchical Sparse Graph Computations on Multicore Platforms

Humayun Kabir (Pennsylvania State University)

k-core and k-truss are cohesive subgraphs used for visualization, community detection and centrality analysis. These cohesive subgraphs have hierarchical structures and can be computed exactly in polynomial time. We have developed PKC and PKT algorithms for k-core and k-truss decompositions respectively. PKC decreases atomic operations and uses less memory. PKT uses a well-designed data structure that updates trusses concurrently, is memory efficient and avoids using hash-table. On 24 threads, PKT is more than one magnitude faster than a state-of-art algorithm. Sparse matrix computations are useful for scientific computing and engineering. We have developed a multilevel data structure CSR-k to store a sparse matrix that enhances locality in memory access pattern and decreases work load imbalance. On a 32 core node, sparse matrix-vector multiplication using CSR-k is 2.41x faster than state-of-art pOSKI and sparse triangular solution using CSR-k is 2.18x faster than coloring method.

Runtime Solutions to Apply Non-Volatile Memories in Future Computer Systems

Hoda Aghaei Khouzani (University of Delaware)

Memory system is an indispensable part of any computer system from large supercomputers to embedded devices. Its performance, energy, capacity, and management algorithms must scale as computer systems scale to ensure the desired performance growth and the possibility of running new applications. Unfortunately, observable trends of current memory technologies from on-chip Static-RAM (SRAM) caches to Dynamic-RAM (DRAM) main memory and Hard Disk Drive (HDD) storage are failing to respond to these existing and emerging requirements, which lead researchers in both academia and industry to look for alternative solutions. Many promising non-volatile memory (NVM) technologies are introduced to replace current devices on various levels of memory hierarchy. Due to the extent of memory hierarchy subject and the diversity of NVM technologies, the main focus of this study is main memory. Specifically, it explores the feasibility of applying two of promising NVMs, namely, phase change memory (PCM) and domain wall memory (DWM).

Modeling and Comparison of Large-Scale Interconnect Designs

Md Atiqul Mollah (Florida State University)

Modern day high performance computing (HPC) clusters and data centers require a large number of computing and storage elements to be interconnected. Interconnect performance is one of the major contributing factors to the overall performance of such systems. Due to the massive scale of the network, interconnect designs are often evaluated and compared through models. My doctoral research is focused on modeling large-scale interconnections and applying such models to investigate the effect of different topology designs and routing scheme designs on the interconnect performance.

Performance Prediction Modeling of GPU Applications

Marcos Amaris (University of Sao Paulo)

Given the large number of GPU architectures and the many different possibilities to execute an algorithm, the prediction of application execution times over GPUs is a great challenge and is essential for efficient JMSs. The available GPU performance modeling solutions are very sensitive to applications and platform changes,

Here a summary of two main works is shown. In the first work, we present the comparison of a developed BSP-based model to three different ML techniques, this comparison was done with 9 well-known matrix/vector applications. In this work, we wanted to perform a fair comparison, for this reason, we decided that ML process would had the same features that the BSP-based model.

In the second work, we have compared among ML techniques. Here, a two step of extraction features are done. First a correlation analysis and after hierarchical clustering analysis. In this second work, 10 irregular CUDA kernels were used.

Nornir: A Power-Aware Runtime Support for Parallel Applications

Daniele De Sensi (University of Pisa)

Power consumption is becoming a key factor in designing applications and computing systems.

In this thesis we propose Nornir, a runtime system for parallel applications, which can be used by application users to enforce power consumption or performance constraints on their applications, by operating on several knobs like: number of used cores, clock frequency of the cores, etc... Nornir is fully customizable and can be used by autonomic strategies designers to plug in new reconfiguration strategies and to validate them against existing ones.

In this thesis we also designed different reconfiguration algorithms: i) an heuristic to enforce performance constraints; ii) an online learning algorithms to enforce both performance and power consumption constraints; iii) a "concurrency control" algorithm to optimize performance and power consumption of single-producer, single-consumer queues.

All these algorithms have been implemented and validated by using Nornir.

A Heterogeneous HPC Platform for Ill-Structured Spatial Join Processing

Danial Aghajarian (Georgia State University)

Given two layers of large polygonal datasets, detecting those pairs of cross-layer polygons which satisfy a join predicate, such as intersection, is one of the most computationally intensive primitive operations in the spatial domain applications. There are alternative solutions for this ill-structured problem in literature. However, none of them is designed to take advantage of heterogeneous clusters equipped with GPU accelerators to process big spatial data efficiently.

In this research, we propose a distributed heterogeneous HPC platform for spatial join processing based on two-step filter and refinement approach. This work includes two main parts. First, we introduce a set of novel GPU-suited data structures and algorithms. Proof of correctness and analysis is also provided for each algorithm. Second, by applying these new techniques, we propose several GPU-based spatial join systems for various operations such as ST_intersect and polygon overlay to improve the performance of current state of the art systems.

Designing and Building Efficient HPC Cloud with Modern Networking Technologies on Heterogeneous HPC Clusters

Jie Zhang (Ohio State University)

Cloud Computing platforms through virtualization technologies are widely used by many users due to their high availability and flexibility. However, running HPC applications on the cloud still suffers from fairly low performance, more specifically, the degraded I/O performance from the virtualized I/O devices. SR-IOV solves the problem by delivering near-native I/O performance. Nevertheless, SR-IOV lacks high-performance virtualization support, such as locality-aware and NUMA-aware support, and also prevents VM live migration. Moreover, the critical HPC resources among VMs, (e.g, SR-IOV enabled Virtual Functions), need to be managed and isolated to efficiently support running multiple concurrent MPI jobs on HPC clouds. Therefore, we propose a framework to design and build the efficient HPC clouds with modern networking technologies on heterogeneous HPC clusters. Through this framework, the HPC cloud can deliver near-native performance for MPI applications on different types of virtualization environments, support SR-IOV enabled VM migration, and provide the efficient resource sharing.

Productivity and Software Development Effort Estimation in HPC

Sandra Wienke (RWTH Aachen University)

Relying vast HPC investments on an informed decision making process is important when serving the ever increasing demands for computational power. Providing a quantitative metric to compare and evaluate HPC systems in procurement processes, I define a productivity model with predictive power that focuses on the number of simulation application runs and the total cost of ownership (TCO). As part of TCO, the software development costs determined by efforts to parallelize, port or tune simulation codes for novel HPC setups must be predicted. Since approaches from mainstream software engineering does not directly suite HPC needs, I set up a methodology to estimate software development effort in HPC. Proof of concepts are mostly based on real-world HPC setups at RWTH Aachen University.

Reducing Communication Costs in the Parallel Algebraic Multigrid

Amanda J. Bienz (University of Illinois)

Algebraic multigrid is an optimal sparse linear solver that suffers from poor parallel scalability due to large communication requirements. This costly communication can be improved through changes to both the algorithm and parallel implementation. The amount of data required to be communicated can be reduced through sparsification of the matrices, yielding slower convergence but significantly cheaper iterations. Furthermore, the data that must be communicated can be sent between processes with increased efficiency through the use of topology-aware methods. The number of messages injected into the network can be reduced greatly at the cost of additional intra-node communication, yielding cheaper communication on coarse levels of AMG.

Speeding Up GPU Graph Processing Using Structural Graph Properties

Merijn Elwin Verstraaten (University of Amsterdam)

Due to its wide applicability and flexibility, graph processing is an increasingly important part of data science. To scale complex graph analytics computations to large datasets it is becoming popular to utilise accelerator-based architectures, such as Graphical Processing Units (GPUs).

Mapping irregular graph algorithms to hardware designed for highly regular parallelism is a complex task. There are often multiple ways to parallelise the same operation on the GPU. Which of these parallelisation strategies is the fastest is dependent on the structure of the input graph. Performance differences can be an order of magnitude or more, and the optimal strategy varies from graph to graph.

The goal of my PhD research is to identify how structural properties impact the performance of different strategies and use this information to speed-up GPU graph processing by predicting the fastest parallelisation of an algorithm for a specific input graph.

Mitigating Variability in HPC Systems and Applications for Performance and Power Efficiency

Bilge Acun (University of Illinois)

Power consumption and process variability are two important, interconnected, challenges of future generation large-scale HPC data-centers. Current production petaflop supercomputers consume more than 10 megawatts of power that costs millions of dollars every year. As HPC moves towards exascale, power consumption is expected to become a major concern. Not solely dynamic behavior of HPC applications (such as irregular or imbalanced applications) but also dynamic behavior of HPC systems (such as thermal, power, frequency variations among processors) makes it challenging to optimize the performance and power efficiency of large scale applications. Smart and adaptive runtime systems have great potential to handle these challenges transparently from the application.

In my thesis, I first analyze frequency, temperature, and power variations in large-scale HPC systems using thousands of cores and different applications. After I identify the cause of these variations, I propose solutions to mitigate them to improve performance and power efficiency.

Using Runtime Energy Optimizations to Improve Energy Efficiency in High Performance Computing

Sridutt Bhalachandra (University of North Carolina)

Energy efficiency in high performance computing (HPC) will be critical to limit operating costs and carbon footprints in future supercomputing centers. With both hardware and software factors affecting energy usage there exists a need for dynamic power regulation. This dissertation highlights an adaptive runtime framework that can allow processors capable of per-core specific power control to reduce power with little performance impact by dynamically adapting to workload characteristics. Monitoring of performance and power regulation is done transparently within MPI runtime and no code changes are required in the underlying application. In the presence of workload imbalance, the runtime reduces the frequency on cores not on the critical path thereby reducing power without deteriorating performance. This is shown to reduce run-to-run performance variation and improve performance in certain scenarios. For applications plagued by memory related issues, new memory metrics are identified that facilitate lowering power without adversely impacting performance.

Bounded Asynchrony and Nested Parallelism for Scalable Graph Processing

Adam Fidel (Texas A&M University)

Processing large-scale graphs has become a critical component in a variety of fields, from scientific computing to social analytics. The irregular access pattern for graph workloads, coupled with complex graph structures and large data sizes makes efficiently executing parallel graph workloads challenging. In this dissertation, we develop two broad techniques for improving the performance of general parallel graph algorithms: bounded asynchrony and nested parallelism. Increasing asynchrony in a bounded manner allows one to avoid costly global synchronization at scale, while still avoiding the penalty of unbounded asynchrony including redundant work. Using this technique, we scale a BFS workload to 98,304 cores where traditional methods stop scaling. Additionally, asynchrony enables a new family of approximate algorithms for applications tolerant to fixed amounts of error. Representing graph algorithms in a nested parallel manner enables the full use of available parallelism inherent in graph algorithms, while efficiently managing communication through nested sections.

Computing Architectures Exploiting Optical Interconnect and Optical Memory Technologies

Pavlos Maniotis (Aristotle University of Thessaloniki)

The long-standing “Bandwidth” and “Memory-Wall” problems hinder exascale computing within the available energy boundaries. This work proposes new optical-systems for use in innovative Computing Architectures towards increasing bandwidth and reducing energy-consumption. The initial efforts were devoted towards elevating optical-memories from the so-far single-bit layouts to complete multi-bit optical-cache setups, demonstrating the first optical-cache blueprint at 16Gb/s. The optical-cache credentials are highlighted by proposing an optically enabled Chip-Multiprocessor architecture; the shared single-level optical-cache approach negates the need for complex cache-hierarchy and demonstrates 20% speed-up for PARSEC benchmarks. Extending optics utilization to complete HPC environments, OptoHPC-Sim simulator was developed for supporting system-scale utilization of optical-interconnect and electro-optical routing technologies. Using OptoHPC-Sim, an OptoBoard-based HPC-network is proposed demonstrating 190% mean throughput improvement and 83% mean packet-delay reduction compared to world’s #3 HPC. Finally, all this experience led to the design of the first optical Ternary-CAM-row architecture at 20Gb/s for fast address look-up configurations.

Thursday, November 16th

Room: Four Seasons Ballroom

10:00 am - 3:00 pm

Exhibition of Doctoral Showcase Posters

Hierarchical Sparse Graph Computations on Multicore Platforms

Humayun Kabir (Pennsylvania State University)

k-core and k-truss are cohesive subgraphs used for visualization, community detection and centrality analysis. These cohesive subgraphs have hierarchical structures and can be computed exactly in polynomial time. We have developed PKC and PKT algorithms for k-core and k-truss decompositions respectively. PKC decreases atomic operations and uses less memory. PKT uses a well-designed data structure that updates trusses concurrently, is memory efficient and avoids using hash-table. On 24 threads, PKT is more than one magnitude faster than a state-of-art algorithm. Sparse matrix computations are useful for scientific computing and engineering. We have developed a multilevel data structure CSR-k to store a sparse matrix that enhances locality in memory access pattern and decreases work load imbalance. On a 32 core node, sparse matrix-vector multiplication using CSR-k is 2.41x faster than state-of-art pOSKI and sparse triangular solution using CSR-k is 2.18x faster than coloring method.

Runtime Solutions to Apply Non-Volatile Memories in Future Computer Systems

Hoda Aghaei Khouzani (University of Delaware)

Memory system is an indispensable part of any computer system from large supercomputers to embedded devices. Its performance, energy, capacity, and management algorithms must scale as computer systems scale to ensure the desired performance growth and the possibility of running new applications. Unfortunately, observable trends of current memory technologies from on-chip Static-RAM (SRAM) caches to Dynamic-RAM (DRAM) main memory and Hard Disk Drive (HDD) storage are failing to respond to these existing and emerging requirements, which lead researchers in both academia and industry to look for alternative solutions. Many promising non-volatile memory (NVM) technologies are introduced to replace current devices on various levels of memory hierarchy. Due to the extent of memory hierarchy subject and the diversity of NVM technologies, the main focus of this study is main memory. Specifically, it explores the feasibility of applying two of promising NVMs, namely, phase change memory (PCM) and domain wall memory (DWM).

Modeling and Comparison of Large-Scale Interconnect Designs

Md Atiqul Mollah (Florida State University)

Modern day high performance computing (HPC) clusters and data centers require a large number of computing and storage elements to be interconnected. Interconnect performance is one of the major contributing factors to the overall performance of such systems. Due to the massive scale of the network, interconnect designs are often evaluated and compared through models. My doctoral research is focused on modeling large-scale interconnections and applying such models to investigate the effect of different topology designs and routing scheme designs on the interconnect performance.

Performance Prediction Modeling of GPU Applications

Marcos Amarís (University of Sao Paulo)

Given the large number of GPU architectures and the many different possibilities to execute an algorithm, the prediction of application execution times over GPUs is a great challenge and is essential for efficient JMSs. The available GPU performance modeling solutions are very sensitive to applications and platform changes,

Here a summary of two main works is shown. In the first work, we present the comparison of a developed BSP-based model to three different ML techniques, this comparison was done with 9 well-known matrix/vector applications. In this work, we wanted to perform a fair comparison, for this reason, we decided that ML process would had the same features that the BSP-based model.

In the second work, we have compared among ML techniques. Here, a two step of extraction features are done. First a correlation analysis and after hierarchical clustering analysis. In this second work, 10 irregular CUDA kernels were used.

Nornir: A Power-Aware Runtime Support for Parallel Applications

Daniele De Sensi (University of Pisa)

Power consumption is becoming a key factor in designing applications and computing systems.

In this thesis we propose Nornir, a runtime system for parallel applications, which can be used by application users to enforce power consumption or performance constraints on their applications, by operating on several knobs like: number of used cores, clock frequency of the cores, etc... Nornir is fully customizable and can be used by autonomic strategies designers to plug in new reconfiguration strategies and to validate them against existing ones.

In this thesis we also designed different reconfiguration algorithms: i) an heuristic to enforce performance constraints; ii) an online learning algorithms to enforce both performance and power consumption constraints; iii) a "concurrency control" algorithm to optimize performance and power consumption of single-producer, single-consumer queues.

All these algorithms have been implemented and validated by using Nornir.

A Heterogeneous HPC Platform for III-Structured Spatial Join Processing

Danial Aghajarian (Georgia State University)

Given two layers of large polygonal datasets, detecting those pairs of cross-layer polygons which satisfy a join predicate, such as intersection, is one of the most computationally intensive primitive operations in the spatial domain applications. There are alternative solutions for this ill-structured problem in literature. However, none of them is designed to take advantage of heterogeneous clusters equipped with GPU accelerators to process big spatial data efficiently.

In this research, we propose a distributed heterogeneous HPC platform for spatial join processing based on two-step filter and refinement approach. This work includes two main parts. First, we introduce a set of novel GPU-suited data structures and algorithms. Proof of correctness and analysis is also provided for each algorithm. Second, by applying these new techniques, we propose several GPU-based spatial join systems for various operations such as ST_intersect and polygon overlay to improve the performance of current state of the art systems.

Designing and Building Efficient HPC Cloud with Modern Networking Technologies on Heterogeneous HPC Clusters

Jie Zhang (Ohio State University)

Cloud Computing platforms through virtualization technologies are widely used by many users due to their high availability and flexibility. However, running HPC applications on the cloud still suffers from fairly low performance, more specifically, the degraded I/O performance from the virtualized I/O devices. SR-IOV solves the problem by delivering near-native I/O performance. Nevertheless, SR-IOV lacks high-performance virtualization support, such as locality-aware and NUMA-aware support, and also prevents VM live migration. Moreover, the critical HPC resources among VMs, (e.g, SR-IOV enabled Virtual Functions), need to be managed and isolated to efficiently support running multiple concurrent MPI jobs on HPC clouds. Therefore, we propose a framework to design and build the efficient HPC clouds with modern networking technologies on heterogeneous HPC clusters. Through this framework, the HPC cloud can deliver near-native performance for MPI applications on different types of virtualization environments, support SR-IOV enabled VM migration, and provide the efficient resource sharing.

Productivity and Software Development Effort Estimation in HPC

Sandra Wienke (RWTH Aachen University)

Relying vast HPC investments on an informed decision making process is important when serving the ever increasing demands for computational power. Providing a quantitative metric to compare and evaluate HPC systems in procurement processes, I define a productivity model with predictive power that focuses on the number of simulation application runs and the total cost of ownership (TCO). As part of TCO, the software development costs determined by efforts to parallelize, port or tune simulation codes for novel HPC setups must be predicted. Since approaches from mainstream software engineering does not directly suite HPC needs, I set up a methodology to estimate software development effort in HPC. Proof of concepts are mostly based on real-world HPC setups at RWTH Aachen University.

Reducing Communication Costs in the Parallel Algebraic Multigrid

Amanda J. Bienz (University of Illinois)

Algebraic multigrid is an optimal sparse linear solver that suffers from poor parallel scalability due to large communication requirements. This costly communication can be improved through changes to both the algorithm and parallel implementation. The amount of data required to be communicated can be reduced through sparsification of the matrices, yielding slower convergence but significantly cheaper iterations. Furthermore, the data that must be communicated can be sent between processes with increased efficiency through the use of topology-aware methods. The number of messages injected into the network can be reduced greatly at the cost of additional intra-node communication, yielding cheaper communication on coarse levels of AMG.

Speeding Up GPU Graph Processing Using Structural Graph Properties

Merijn Elwin Verstraaten (University of Amsterdam)

Due to its wide applicability and flexibility, graph processing is an increasingly important part of data science. To scale complex graph analytics computations to large datasets it is becoming popular to utilise accelerator-based architectures, such as Graphical Processing Units (GPUs).

Mapping irregular graph algorithms to hardware designed for highly regular parallelism is a complex task. There are often multiple ways to parallelise the same operation on the GPU. Which of these parallelisation strategies is the fastest is dependent on the structure of the input graph. Performance differences can be an order of magnitude or more, and the optimal strategy varies from graph to graph.

The goal of my PhD research is to identify how structural properties impact the performance of different strategies and use this information to speed-up GPU graph processing by predicting the fastest parallelisation of an algorithm for a specific input graph.

Mitigating Variability in HPC Systems and Applications for Performance and Power Efficiency

Bilge Acun (University of Illinois)

Power consumption and process variability are two important, interconnected, challenges of future generation large-scale HPC

data-centers. Current production petaflop supercomputers consume more than 10 megawatts of power that costs millions of dollars every year. As HPC moves towards exascale, power consumption is expected to become a major concern. Not solely dynamic behavior of HPC applications (such as irregular or imbalanced applications) but also dynamic behavior of HPC systems (such as thermal, power, frequency variations among processors) makes it challenging to optimize the performance and power efficiency of large scale applications. Smart and adaptive runtime systems have great potential to handle these challenges transparently from the application.

In my thesis, I first analyze frequency, temperature, and power variations in large-scale HPC systems using thousands of cores and different applications. After I identify the cause of these variations, I propose solutions to mitigate them to improve performance and power efficiency.

Using Runtime Energy Optimizations to Improve Energy Efficiency in High Performance Computing

Sridutt Bhalachandra (University of North Carolina)

Energy efficiency in high performance computing (HPC) will be critical to limit operating costs and carbon footprints in future supercomputing centers. With both hardware and software factors affecting energy usage there exists a need for dynamic power regulation. This dissertation highlights an adaptive runtime framework that can allow processors capable of per-core specific power control to reduce power with little performance impact by dynamically adapting to workload characteristics. Monitoring of performance and power regulation is done transparently within MPI runtime and no code changes are required in the underlying application. In the presence of workload imbalance, the runtime reduces the frequency on cores not on the critical path thereby reducing power without deteriorating performance. This is shown to reduce run-to-run performance variation and improve performance in certain scenarios. For applications plagued by memory related issues, new memory metrics are identified that facilitate lowering power without adversely impacting performance.

Bounded Asynchrony and Nested Parallelism for Scalable Graph Processing

Adam Fidel (Texas A&M University)

Processing large-scale graphs has become a critical component in a variety of fields, from scientific computing to social analytics. The irregular access pattern for graph workloads, coupled with complex graph structures and large data sizes makes efficiently executing parallel graph workloads challenging. In this dissertation, we develop two broad techniques for improving the performance of general parallel graph algorithms: bounded asynchrony and nested parallelism. Increasing asynchrony in a bounded manner allows one to avoid costly global synchronization at scale, while still avoiding the penalty of unbounded asynchrony including redundant work. Using this technique, we scale a BFS workload to 98,304 cores where traditional methods stop scaling. Additionally, asynchrony enables a new family of approximate algorithms for applications tolerant to fixed amounts of error. Representing graph algorithms in a nested parallel manner enables the full use of available parallelism inherent in graph algorithms, while efficiently managing communication through nested sections.

Computing Architectures Exploiting Optical Interconnect and Optical Memory Technologies

Pavlos Maniotis (Aristotle University of Thessaloniki)

The long-standing "Bandwidth" and "Memory-Wall" problems hinder exascale computing within the available energy boundaries. This work proposes new optical-systems for use in innovative Computing Architectures towards increasing bandwidth and reducing energy-consumption. The initial efforts were devoted towards elevating optical-memories from the so-far single-bit layouts to complete multi-bit optical-cache setups, demonstrating the first optical-cache blueprint at 16Gb/s. The optical-cache credentials are highlighted by proposing an optically enabled Chip-Multiprocessor architecture; the shared single-level optical-cache approach negates the need for complex cache-hierarchy and demonstrates 20% speed-up for PARSEC benchmarks. Extending optics utilization to complete HPC environments, OptoHPC-Sim simulator was developed for supporting system-scale utilization of optical-interconnect and electro-optical routing technologies. Using OptoHPC-Sim, an OptoBoard-based HPC-network is proposed demonstrating 190% mean throughput improvement and 83% mean packet-delay reduction compared to world's #3 HPC. Finally, all this experience led to the design of the first optical Ternary-CAM-row architecture at 20Gb/s for fast address look-up configurations.

Early Career Program [\(back to top\)](#)

Monday, November 13th

Room: 711

8:30 am - 10:00 am

Forming Connections I: Connecting Sideways, with Ourselves and Our Peers

Program Overview

This event is exclusive to accepted participants of the Early Career program.

Peer Networking

This event is exclusive to accepted participants of the Early Career program.

Understanding and Appreciating our Professional Strengths and Weaknesses

This event is exclusive to accepted participants of the Early Career program.

Room: 711

10:30 am - 12:00 pm

Career Management

Thinking Strategically *Andrew Lumsdaine (Pacific Northwest National Laboratory, University of Washington)*

This event is open to any registrant of the SC17 conference.

Forming Strong Networks and Collaborations *Dorian Arnold (Emory University)*

This event is open to any registrant of the SC17 conference.

Performance, Advancement, and Promotions *Barney Maccabe (Oak Ridge National Laboratory)*

This event is open to any registrant of the SC conference.

Time Management *Jeanine Cook (Sandia National Laboratories)*

This event is open to any registrant of the SC17 conference.

Career Management Panel with Dorian Arnold (Emory University), Jeanine Cook (Sandia National Labs), Andrew Lumsdaine (Pacific Northwest National Lab and University of Washington) and Barney Maccabe (Oak Ridge National Lab)

This event is open to any registrant of the SC conference.

Room: Four Seasons Ballroom

12:00 pm - 1:30 pm

Luncheon

Luncheon

This event is exclusive to formally accepted Early Career Program participants.

Room: 711

1:30 pm - 3:00 pm

Research Execution

Funding Agencies *Kirk Cameron (Virginia Tech)*

This event is open to any registrant of the SC17 conference.

Writing Effective Proposals *Mary Hall (University of Utah)*

This event is open to any registrant of the SC17 conference.

Research Methods *Mike Heroux (Sandia National Laboratories)*

This event is open to any registrant of the SC17 conference.

Publishing *Torsten Hoefler (ETH Zurich)*

This event is open to any registrant of the SC17 conference.

Research Execution Panel with Kirk Cameron (Virginia Tech), Mary Hall (University of Utah), Mike Heroux (Sandia National Labs) and Torsten Hoefler (ETH Zürich)

This event is open to any registrant of the SC17 conference.

Room: Food Court - Lobby B

3:30 pm - 5:00 pm

Connections II: Connecting with Mentors

Speed Mentoring

This event is exclusive to accepted participants of the Early Career program.

Emerging Technologies [\(back to top\)](#)

Tuesday, November 14th

Room: 401

10:00 am - 6:00 pm

Emerging Technologies Showcase (Day 1)

Rook Distributed Storage System

Bassam Tabbara (Quantum Corporation), Dan Kerns (Quantum Corporation), Dave Suzuki (Quantum Corporation)

Rook is an open source distributed storage system which exposes file, block, and object storage on top of shared resource pools. Designed for clustered computing, hyperscale environments and enterprise HPC environments requiring connectivity to a variety of applications or codes with differing requirements of block, file or object interface to the storage.

Optimized to deliver low latency access to NVMe and new network devices, Rook has minimal physical hardware dependencies and can be deployed in dedicated storage clusters or converged clusters. It's self-managing, self-protecting, self-healing, and is designed to work without teams of engineers managing it, Rook scales from a single node to multi-PB clusters spread geographically.

Rook is based on the Ceph project that has over 10 years of production deployments in some of the largest storage clusters in the world. Rook technology takes on solving fast resource provisioning from containers, addresses gaps in troubleshooting and optimizing performance, and adapts to the latest NVMe-based technologies.

Towards a Composable Computer System

IHsin Chung (IBM), Brian Pan (H3 Platform Inc), Tony Hsu (Inventec Corporation)

The recent advancement of technology in both software and hardware enables us to revisit the design of the composable architecture in the system design. The composable system design provides capability to serve a variety of workloads. The system offers a dynamic co-design platform that allows experiments and measurements. This speeds up the system design and software evolution. It also decouples the lifecycles of components. The design consideration includes adopting available technology with the understanding of application characteristics. The design has the potential to be the infrastructure of cloud computing and the HPC architecture.

PARADISE: A ToolFlow to Model Emerging Technologies for the Post-CMOS Era in HPC

Dilip Vasudevan (Lawrence Berkeley National Laboratory), George Michelogiannakis (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), John Shalf (Lawrence Berkeley National Laboratory)

With the decline and eventual end of historical rates of lithographic scaling, we arrive at a crossroad where synergistic and holistic decisions are required to preserve Moore's law technology scaling. Numerous emerging technologies aim to extend digital electronics scaling of performance, energy efficiency, and computational power/density, ranging from devices (transistors), memories, 3D integration capabilities, specialized architectures, photonics, and others. The wide range of technology options creates the need for an integrated strategy to understand the impact of these emerging technologies on future large-scale digital systems for diverse application requirements and optimization metrics. To this end, we present PARADISE, a comprehensive modeling suite to evaluate different kinds of emerging technologies from the transistor level all the way up to the architecture. We will demonstrate how to model new devices and architectures at the low level, and then use the characterization results as inputs to the architectural level, and this way conduct experiments to illustrate the high-level impact of new devices. PARADISE is based on open-source tools and currently uses important HPC kernels to specialize hardware.

Silicon Photonic LIONS: All-to-All Interconnects for Energy-Efficient, Scalable, and Modular HPC Systems

Roberto Proietti (University of California, Davis), Yu Zhang (University of California, Davis), Xian Xiao (University of California, Davis), Gengchen Liu (University of California, Davis), S.J. Ben Yoo (University of California, Davis)

This project pursues extremely energy-efficient and high-throughput photonic interconnected computing systems enabled by scalable and modular integration of 2.5D or 3D silicon photonic (SiP) interposers with all-to-all, contention-less, and arbitration-free wavelength routing by cyclic Arrayed-Waveguide-Grating-Routers (AWGRs) [LIONS: Low-Latency Interconnect Optical Network Switches]. First, this paper introduces the working principle of the all-to-all photonic interconnect technology and show how it can be used to implement a multi-chip computing board. Second, the paper evaluates performance of the system under various workloads in terms of execution time and energy per transaction via simulations. Third, the latest results in terms of device fabrication and testing of SiP AWGR photonic routers are presented. Finally, the paper discusses modular and hierarchical scaling of the all-to-all photonic interconnects utilizing AWGRs with practical numbers of port counts.

Scalable Heterogeneous Programmable Reconfigurable Computing

Andrei Vassiliev (Scientific Concepts International Corporation)

Multi-FPGA acceleration has already shown several orders of magnitude speedups in today's analytics, scientific, financial, image processing and machine learning applications. However, the scalability of FPGA-based HPC demanded by the future avalanche of data and network traffic is still in its infancy. OpenCL memory model holds the promise of straightforward and pragmatic scalability of FPGA computing beyond single FPGA within the framework of today's OpenCL FPGA compilers. Scientific Concepts International develops novel Smart Cell Interconnect (SCI(tm)) optimized for OpenCL global memory, streaming data accesses and network packets encapsulated into switched cells. Recent advances in Open Source tools based on polyhedral model and IR enable development of source-to-source coarse grain code and data partitioning of the HPC workloads written initially in OpenCL and followed by C/C++, SYCL. Evolution of the multi-FPGA partitioning tools and SCI interconnect IP will enable true scalability of the computing fabric of arrays and clusters of FPGAs. Cloud FPGA computing, fog computing at the source of the generated data as well as fusion of networking, security, and computing are addressed by our Heterogeneous Programmable Reconfigurable Computing (HPRC(tm)) architecture, SCI(tm) interconnect and partitioner tools.

DOME Hot-Water Cooled MicroDataCenter

Ronald Peter Luijten (IBM)

Our DOME microDataCenter is a complete datacenter-in-a-box, consisting of compute, storage, communication, power and cooling in the densest known packaged result. It is 20-fold denser than current datacenter technology, effectively compacting an entire server rack into a single 2U rack-unit with the same delivered aggregate compute performance. The DOME microDataCenter uses commodity parts and standards and has no moving parts. It was designed from ground up by ASTRON and IBM research, targeting SKA, and is being brought to market by ila-microservers which was incorporated Jan 2017. It recently won the HPC Data Center Innovation Award from IDC HPC (now hyperion research). ILA plans to start shipping a qualified product early 2018 (Orders can be placed this year for pre-production evaluation systems). The salient features are 1) very energy efficient, 2) high reliability (no fans or rotating drives), 3) Hot-water cooling allowing to capture and resell up to 80% of the electricity costs, 4) high density. The 2U rack-unit contains two 32-way carriers from yielding 64 servers with up to 1536 cores, 1.5Tbyte memory and 64TB storage with a TDP of 1500W. We will run various applications live including HPL and the our new BlueBild imager.

Posit Math Unit (PMU) – A New Approach Toward Exascale Computing

Anantha P. Kinnal (Calligo Technologies, Posit Research), Theodore Omtzigt (Stillwater Supercomputing Inc, Posit Research), Siew Hoon Leong (National Supercomputing Center Singapore, Posit Research), John L. Gustafson (National University of Singapore, Posit Research)

A new data type called a posit is designed as a replacement for IEEE Standard 754 floating-point numbers (floats). Unlike earlier forms of universal number (unum) arithmetic, posits do not require interval arithmetic or variable size operands; like floats, they round if an answer is inexact. However, they provide compelling advantages over floats - larger dynamic range, higher accuracy, identical results across systems, simpler hardware, simpler exception handling. Posits never overflow to infinity or underflow to zero, and "Not-a-Number" (NaN) indicates an action instead of a bit pattern.

Posit math unit (PMU) takes less circuitry than IEEE float FPU. Using lower power and smaller silicon, posit operations per second (POPS) supported by a chip can be significantly higher than FLOPS using similar hardware resources. GPU accelerators and Deep Learning processors, can do more per watt and per dollar with posits, yet deliver superior answer quality.

Series of benchmarks compares floats and posits for accuracy produced for a set precision. Low precision posits provide a better solution than "approximate computing" that try to tolerate decreased answer quality. High precision posits provide more correct decimals than floats of the same size; in some cases, 32-bit posit may safely replace 64-bit float.

Introducing DPU - Data-Storage Processing Unit – Placing Intelligence in Storage

Yafei Yang (Shenzhen DAPU Microelectronics Company), Weijun Li (Shenzhen DAPU Microelectronics Company), Qing Yang (Shenzhen DAPU Microelectronics Company, University of Rhode Island)

It has come to a historical point where we should reconsider storage controller's architecture to keep pace with the explosive growth of big data and fast emergence of new storage media technologies. We introduce a storage controller with built-in intelligence, referred to as DPU for Data-storage Processing Unit. DPU manages, controls, analyzes, and classifies data at the place where they are stored. The idea is to place sufficient intelligence closest to the storage devices that are experiencing revolutionary changes with the emergence of storage class memories. Machine learning logics are a major part of DPU that learns I/O behaviors inside the storage to optimize adaptively I/O performance, data reliability, and availability for a variety of applications. By learning production I/Os, page writes are clustered before they are programmed to flash blocks, resulting in 10x reduction of flash block erasures in flash SSD, dramatically extending the life time of flash memories. DPU efficiently implements in-situ data intensive processing tasks such as string/text search and data conversions, which alleviates I/O bus pressure. Our results have shown over 10x performance improvements for simple string search and data serialization/deserialization applications compared with traditional systems.

Wednesday, November 15th

Room: 401
10:00 am - 6:00 pm

Emerging Technologies Showcase (Day 2)

Rook Distributed Storage System

Bassam Tabbara (Quantum Corporation), Dan Kerns (Quantum Corporation), Dave Suzuki (Quantum Corporation)

Rook is an open source distributed storage system which exposes file, block, and object storage on top of shared resource pools. Designed for clustered computing, hyperscale environments and enterprise HPC environments requiring connectivity to a variety of applications or codes with differing requirements of block, file or object interface to the storage.

Optimized to deliver low latency access to NVMe and new network devices, Rook has minimal physical hardware dependencies and can be deployed in dedicated storage clusters or converged clusters. It's self-managing, self-protecting, self-healing, and is designed to work without teams of engineers managing it, Rook scales from a single node to multi-PB clusters spread geographically.

Rook is based on the Ceph project that has over 10 years of production deployments in some of the largest storage clusters in the world. Rook technology takes on solving fast resource provisioning from containers, addresses gaps in troubleshooting and optimizing performance, and adapts to the latest NVMe-based technologies.

Towards a Composable Computer System

IHsin Chung (IBM), Brian Pan (H3 Platform Inc), Tony Hsu (Inventec Corporation)

The recent advancement of technology in both software and hardware enables us to revisit the design of the composable architecture in the system design. The composable system design provides capability to serve a variety of workloads. The system offers a dynamic co-design platform that allows experiments and measurements. This speeds up the system design and software evolution. It also decouples the lifecycles of components. The design consideration includes adopting available technology with the understanding of application characteristics. The design has the potential to be the infrastructure of cloud computing and the HPC architecture.

PARADISE: A ToolFlow to Model Emerging Technologies for the Post-CMOS Era in HPC

Dilip Vasudevan (Lawrence Berkeley National Laboratory), George Michelogiannakis (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), John Shalf (Lawrence Berkeley National Laboratory)

With the decline and eventual end of historical rates of lithographic scaling, we arrive at a crossroad where synergistic and holistic decisions are required to preserve Moore's law technology scaling. Numerous emerging technologies aim to extend digital electronics scaling of performance, energy efficiency, and computational power/density, ranging from devices (transistors), memories, 3D integration capabilities, specialized architectures, photonics, and others. The wide range of technology options creates the need for an integrated strategy to understand the impact of these emerging technologies on future large-scale digital systems for diverse application requirements and optimization metrics. To this end, we present PARADISE, a comprehensive modeling suite to evaluate different kinds of emerging technologies from the transistor level all the way up to the architecture. We will demonstrate how to model new devices and architectures at the low level, and then use the characterization results as inputs to the architectural level, and this way conduct experiments to illustrate the high-level impact of new devices. PARADISE is based on open-source tools and currently uses important HPC kernels to specialize hardware.

Silicon Photonic LIONS: All-to-All Interconnects for Energy-Efficient, Scalable, and Modular HPC Systems

Roberto Proietti (University of California, Davis), Yu Zhang (University of California, Davis), Xian Xiao (University of California, Davis), Gengchen Liu (University of California, Davis), S.J. Ben Yoo (University of California, Davis)

This project pursues extremely energy-efficient and high-throughput photonic interconnected computing systems enabled by scalable and modular integration of 2.5D or 3D silicon photonic (SiP) interposers with all-to-all, contention-less, and arbitration-free wavelength routing by cyclic Arrayed-Waveguide-Grating-Routers (AWGRs) [LIONS: Low-Latency Interconnect Optical Network Switches]. First, this paper introduces the working principle of the all-to-all photonic interconnect technology and show how it can be used to implement a multi-chip computing board. Second, the paper evaluates performance of the system under various workloads in terms of execution time and energy per transaction via simulations. Third, the latest results in terms of device fabrication and testing of SiP AWGR photonic routers are presented. Finally, the paper discusses modular and hierarchical scaling of the all-to-all photonic interconnects utilizing AWGRs with practical numbers of port counts.

Scalable Heterogeneous Programmable Reconfigurable Computing

Andrei Vassiliev (Scientific Concepts International Corporation)

Multi-FPGA acceleration has already shown several orders of magnitude speedups in today's analytics, scientific, financial, image

processing and machine learning applications. However, the scalability of FPGA-based HPC demanded by the future avalanche of data and network traffic is still in its infancy. OpenCL memory model holds the promise of straightforward and pragmatic scalability of FPGA computing beyond single FPGA within the framework of today's OpenCL FPGA compilers. Scientific Concepts International develops novel Smart Cell Interconnect (SCI(tm)) optimized for OpenCL global memory, streaming data accesses and network packets encapsulated into switched cells. Recent advances in Open Source tools based on polyhedral model and IR enable development of source-to-source coarse grain code and data partitioning of the HPC workloads written initially in OpenCL and followed by C/C++, SYCL. Evolution of the multi-FPGA partitioning tools and SCI interconnect IP will enable true scalability of the computing fabric of arrays and clusters of FPGAs. Cloud FPGA computing, fog computing at the source of the generated data as well as fusion of networking, security, and computing are addressed by our Heterogeneous Programmable Reconfigurable Computing (HPRC(tm)) architecture, SCI(tm) interconnect and partitioner tools.

DOME Hot-Water Cooled MicroDataCenter

Ronald Peter Luijten (IBM)

Our DOME microDataCenter is a complete datacenter-in-a-box, consisting of compute, storage, communication, power and cooling in the densest known packaged result. It is 20-fold denser than current datacenter technology, effectively compacting an entire server rack into a single 2U rack-unit with the same delivered aggregate compute performance. The DOME microDataCenter uses commodity parts and standards and has no moving parts. It was designed from ground up by ASTRON and IBM research, targeting SKA, and is being brought to market by ila-microservers which was incorporated Jan 2017. It recently won the HPC Data Center Innovation Award from IDC HPC (now hyperion research). ILA plans to start shipping a qualified product early 2018 (Orders can be placed this year for pre-production evaluation systems). The salient features are 1) very energy efficient, 2) high reliability (no fans or rotating drives), 3) Hot-water cooling allowing to capture and resell up to 80% of the electricity costs, 4) high density. The 2U rack-unit contains two 32-way carriers from yielding 64 servers with up to 1536 cores, 1.5Tbyte memory and 64TB storage with a TDP of 1500W. We will run various applications live including HPL and the our new BlueBild imager.

Posit Math Unit (PMU) – A New Approach Toward Exascale Computing

Anantha P. Kinnal (Calligo Technologies, Posit Research), Theodore Omtzigt (Stillwater Supercomputing Inc, Posit Research), Siew Hoon Leong (National Supercomputing Center Singapore, Posit Research), John L. Gustafson (National University of Singapore, Posit Research)

A new data type called a posit is designed as a replacement for IEEE Standard 754 floating-point numbers (floats). Unlike earlier forms of universal number (unum) arithmetic, posits do not require interval arithmetic or variable size operands; like floats, they round if an answer is inexact. However, they provide compelling advantages over floats - larger dynamic range, higher accuracy, identical results across systems, simpler hardware, simpler exception handling. Posits never overflow to infinity or underflow to zero, and "Not-a-Number" (NaN) indicates an action instead of a bit pattern.

Posit math unit (PMU) takes less circuitry than IEEE float FPU. Using lower power and smaller silicon, posit operations per second (POPS) supported by a chip can be significantly higher than FLOPS using similar hardware resources. GPU accelerators and Deep Learning processors, can do more per watt and per dollar with posits, yet deliver superior answer quality.

Series of benchmarks compares floats and posits for accuracy produced for a set precision. Low precision posits provide a better solution than "approximate computing" that try to tolerate decreased answer quality. High precision posits provide more correct decimals than floats of the same size; in some cases, 32-bit posit may safely replace 64-bit float.

Introducing DPU - Data-Storage Processing Unit – Placing Intelligence in Storage

Yafei Yang (Shenzhen DAPU Microelectronics Company), Weijun Li (Shenzhen DAPU Microelectronics Company), Qing Yang (Shenzhen DAPU Microelectronics Company, University of Rhode Island)

It has come to a historical point where we should reconsider storage controller's architecture to keep pace with the explosive growth of big data and fast emergence of new storage media technologies. We introduce a storage controller with built-in intelligence, referred to as DPU for Data-storage Processing Unit. DPU manages, controls, analyzes, and classifies data at the place where they are stored. The idea is to place sufficient intelligence closest to the storage devices that are experiencing revolutionary changes with the emergence of storage class memories. Machine learning logics are a major part of DPU that learns I/O behaviors inside the storage to optimize adaptively I/O performance, data reliability, and availability for a variety of applications. By learning production I/Os, page writes are clustered before they are programmed to flash blocks, resulting in 10x reduction of flash block erasures in flash SSD, dramatically extending the life time of flash memories. DPU efficiently implements in-situ data intensive processing tasks such as string/text search and data conversions, which alleviates I/O bus pressure. Our results have shown over 10x performance improvements for simple string search and data serialization/deserialization applications compared with traditional systems.

Thursday, November 16th

Room: 401

10:00 am - 6:00 pm

Emerging Technologies Showcase (Day 3)

Rook Distributed Storage System

Bassam Tabbara (Quantum Corporation), Dan Kerns (Quantum Corporation), Dave Suzuki (Quantum Corporation)

Rook is an open source distributed storage system which exposes file, block, and object storage on top of shared resource pools. Designed for clustered computing, hyperscale environments and enterprise HPC environments requiring connectivity to a variety of applications or codes with differing requirements of block, file or object interface to the storage.

Optimized to deliver low latency access to NVMe and new network devices, Rook has minimal physical hardware dependencies and can be deployed in dedicated storage clusters or converged clusters. It's self-managing, self-protecting, self-healing, and is designed to work without teams of engineers managing it, Rook scales from a single node to multi-PB clusters spread geographically.

Rook is based on the Ceph project that has over 10 years of production deployments in some of the largest storage clusters in the world. Rook technology takes on solving fast resource provisioning from containers, addresses gaps in troubleshooting and optimizing performance, and adapts to the latest NVMe-based technologies.

Towards a Composable Computer System

IHsin Chung (IBM), Brian Pan (H3 Platform Inc), Tony Hsu (Inventec Corporation)

The recent advancement of technology in both software and hardware enables us to revisit the design of the composable architecture in the system design. The composable system design provides capability to serve a variety of workloads. The system offers a dynamic co-design platform that allows experiments and measurements. This speeds up the system design and software evolution. It also decouples the lifecycles of components. The design consideration includes adopting available technology with the understanding of application characteristics. The design has the potential to be the infrastructure of cloud computing and the HPC architecture.

PARADISE: A ToolFlow to Model Emerging Technologies for the Post-CMOS Era in HPC

Dilip Vasudevan (Lawrence Berkeley National Laboratory), George Michelogiannakis (Lawrence Berkeley National Laboratory), David Donofrio (Lawrence Berkeley National Laboratory), John Shalf (Lawrence Berkeley National Laboratory)

With the decline and eventual end of historical rates of lithographic scaling, we arrive at a crossroad where synergistic and holistic decisions are required to preserve Moore's law technology scaling. Numerous emerging technologies aim to extend digital electronics scaling of performance, energy efficiency, and computational power/density, ranging from devices (transistors), memories, 3D integration capabilities, specialized architectures, photonics, and others. The wide range of technology options creates the need for an integrated strategy to understand the impact of these emerging technologies on future large-scale digital systems for diverse application requirements and optimization metrics. To this end, we present PARADISE, a comprehensive modeling suite to evaluate different kinds of emerging technologies from the transistor level all the way up to the architecture. We will demonstrate how to model new devices and architectures at the low level, and then use the characterization results as inputs to the architectural level, and this way conduct experiments to illustrate the high-level impact of new devices. PARADISE is based on open-source tools and currently uses important HPC kernels to specialize hardware.

Silicon Photonic LIONS: All-to-All Interconnects for Energy-Efficient, Scalable, and Modular HPC Systems

Roberto Proietti (University of California, Davis), Yu Zhang (University of California, Davis), Xian Xiao (University of California, Davis), Gengchen Liu (University of California, Davis), S.J. Ben Yoo (University of California, Davis)

This project pursues extremely energy-efficient and high-throughput photonic interconnected computing systems enabled by scalable and modular integration of 2.5D or 3D silicon photonic (SiP) interposers with all-to-all, contention-less, and arbitration-free wavelength routing by cyclic Arrayed-Waveguide-Grating-Routers (AWGRs) [LIONS: Low-Latency Interconnect Optical Network Switches]. First, this paper introduces the working principle of the all-to-all photonic interconnect technology and show how it can be used to implement a multi-chip computing board. Second, the paper evaluates performance of the system under various workloads in terms of execution time and energy per transaction via simulations. Third, the latest results in terms of device fabrication and testing of SiP AWGR photonic routers are presented. Finally, the paper discusses modular and hierarchical scaling of the all-to-all photonic interconnects utilizing AWGRs with practical numbers of port counts.

Scalable Heterogeneous Programmable Reconfigurable Computing

Andrei Vassiliev (Scientific Concepts International Corporation)

Multi-FPGA acceleration has already shown several orders of magnitude speedups in today's analytics, scientific, financial, image processing and machine learning applications. However, the scalability of FPGA-based HPC demanded by the future avalanche of data and network traffic is still in its infancy. OpenCL memory model holds the promise of straightforward and pragmatic scalability of FPGA computing beyond single FPGA within the framework of today's OpenCL FPGA compilers. Scientific Concepts International develops novel Smart Cell Interconnect (SCI(tm)) optimized for OpenCL global memory, streaming data accesses and network packets encapsulated into switched cells. Recent advances in Open Source tools based on polyhedral model and IR enable development of source-to-source coarse grain code and data partitioning of the HPC workloads written initially in OpenCL

and followed by C/C++, SYCL. Evolution of the multi-FPGA partitioning tools and SCI interconnect IP will enable true scalability of the computing fabric of arrays and clusters of FPGAs. Cloud FPGA computing, fog computing at the source of the generated data as well as fusion of networking, security, and computing are addressed by our Heterogeneous Programmable Reconfigurable Computing (HPRC(tm)) architecture, SCI(tm) interconnect and partitioner tools.

DOME Hot-Water Cooled MicroDataCenter

Ronald Peter Luijten (IBM)

Our DOME microDataCenter is a complete datacenter-in-a-box, consisting of compute, storage, communication, power and cooling in the densest known packaged result. It is 20-fold denser than current datacenter technology, effectively compacting an entire server rack into a single 2U rack-unit with the same delivered aggregate compute performance. The DOME microDataCenter uses commodity parts and standards and has no moving parts. It was designed from ground up by ASTRON and IBM research, targeting SKA, and is being brought to market by ila-microservers which was incorporated Jan 2017. It recently won the HPC Data Center Innovation Award from IDC HPC (now hyperion research). ILA plans to start shipping a qualified product early 2018 (Orders can be placed this year for pre-production evaluation systems). The salient features are 1) very energy efficient, 2) high reliability (no fans or rotating drives), 3) Hot-water cooling allowing to capture and resell up to 80% of the electricity costs, 4) high density. The 2U rack-unit contains two 32-way carriers from yielding 64 servers with up to 1536 cores, 1.5Tbyte memory and 64TB storage with a TDP of 1500W. We will run various applications live including HPL and the our new BlueBild imager.

Posit Math Unit (PMU) – A New Approach Toward Exascale Computing

Anantha P. Kinnal (Calligo Technologies, Posit Research), Theodore Omtzigt (Stillwater Supercomputing Inc, Posit Research), Siew Hoon Leong (National Supercomputing Center Singapore, Posit Research), John L. Gustafson (National University of Singapore, Posit Research)

A new data type called a posit is designed as a replacement for IEEE Standard 754 floating-point numbers (floats). Unlike earlier forms of universal number (unum) arithmetic, posits do not require interval arithmetic or variable size operands; like floats, they round if an answer is inexact. However, they provide compelling advantages over floats - larger dynamic range, higher accuracy, identical results across systems, simpler hardware, simpler exception handling. Posits never overflow to infinity or underflow to zero, and “Not-a-Number” (NaN) indicates an action instead of a bit pattern.

Posit math unit (PMU) takes less circuitry than IEEE float FPU. Using lower power and smaller silicon, posit operations per second (POPS) supported by a chip can be significantly higher than FLOPS using similar hardware resources. GPU accelerators and Deep Learning processors, can do more per watt and per dollar with posits, yet deliver superior answer quality.

Series of benchmarks compares floats and posits for accuracy produced for a set precision. Low precision posits provide a better solution than “approximate computing” that try to tolerate decreased answer quality. High precision posits provide more correct decimals than floats of the same size; in some cases, 32-bit posit may safely replace 64-bit float.

Introducing DPU - Data-Storage Processing Unit – Placing Intelligence in Storage

Yafei Yang (Shenzhen DAPU Microelectronics Company), Weijun Li (Shenzhen DAPU Microelectronics Company), Qing Yang (Shenzhen DAPU Microelectronics Company, University of Rhode Island)

It has come to a historical point where we should reconsider storage controller's architecture to keep pace with the explosive growth of big data and fast emergence of new storage media technologies. We introduce a storage controller with built-in intelligence, referred to as DPU for Data-storage Processing Unit. DPU manages, controls, analyzes, and classifies data at the place where they are stored. The idea is to place sufficient intelligence closest to the storage devices that are experiencing revolutionary changes with the emergence of storage class memories. Machine learning logics are a major part of DPU that learns I/O behaviors inside the storage to optimize adaptively I/O performance, data reliability, and availability for a variety of applications. By learning production I/Os, page writes are clustered before they are programmed to flash blocks, resulting in 10x reduction of flash block erasures in flash SSD, dramatically extending the life time of flash memories. DPU efficiently implements in-situ data intensive processing tasks such as string/text search and data conversions, which alleviates I/O bus pressure. Our results have shown over 10x performance improvements for simple string search and data serialization/deserialization applications compared with traditional systems.

Exhibitor Forum [\(back to top\)](#)

Tuesday, November 14th

Room: 501-502
1:30 pm - 3:00 pm

Systems

Post-K Supercomputer with Fujitsu's Original CPU, Powered by ARM ISA

Toshiyuki Shimizu (Fujitsu Ltd)

Fujitsu has provided supercomputers for forty years; most recently the K computer in 2011, winning 1st place twice in the TOP500, the Fujitsu-original PRIMEHPC FX10 in 2012 and the PRIMEHPC FX100 in 2015. Fujitsu also delivers x86 clusters for wide application areas and customer requirements.

RIKEN and Fujitsu are now developing the “post-K” supercomputer as Japan’s new national supercomputer project. The post-K inherits the K computer’s architecture and targets up to 100 times higher application performance, utilizing a Fujitsu CPU with the ARMv8-A and SVE ISA to widen application opportunities and contribute to the ARM ecosystem for HPC applications, as well as improve science and society. At SC17, Fujitsu will provide updates on post-K development.

Project Aurora – Unveiling NEC’s Brand New Vector Supercomputer

Shintaro MOMOSE (NEC Corporation)

NEC unveils the details of the next-generation vector supercomputer (project code: Aurora), which is a successor system to the latest model SX-ACE. This brand-new system is slated for its product release in 2018. In order to make available both high sustained performance and productivity, NEC develops an innovative vector processor offering the world’s highest memory bandwidth per processor. It is built into a standard PCIe card as Vector Engine, which allows the user to capitalize on a high sustained performance for a spectrum of applications in the de facto standard x86/Linux environment. There is no need to struggle for special programming techniques in harnessing its processing capabilities. This card device enables easy programming in standard Fortran/C/C++ languages, for which the NEC compiler will automatically carry out vectorization and parallelization. The user can benefit from a higher sustained performance with minimum effort by utilizing the Vector Engine hardware.

Innovative Alternate Architectures for Exascale Computing: ThunderX2 and Beyond

Shah Varun (Cavium Inc)

Exascale initiatives around the world will enable 50-100 times faster performance of HPC workloads than on today’s highest performing supercomputers. This breakthrough will open the flood gates to get answers to complex questions in the field of drug discovery, personalized health care, weather forecasting, clean-energy and other scientific and commercial segments by modeling and simulating problems in innovative ways with faster time to solution. Cavium has been leading the industry in ARM servers for the HPC market since the launch of the first generation of ThunderX. This talk will address how Cavium is providing differentiated design point for HPC Applications with its latest ThunderX2 ARMv8 processor and how it plans to tackle the exascale challenge to deliver breakthrough performance to solve the world’s most demanding compute requirements leveraging processor portfolio and compelling roadmap that include system innovations such as Gen Z.

Room: 503-504
1:30 pm - 3:00 pm

Communications and Software

OpenCAPI: High Performance, Host-Agnostic, Coherent Accelerator Interface

Jeff Stuecheli (IBM), Steve Fields (IBM)

Open Coherent Accelerator Processor Interface (OpenCAPI) is a new industry standard architecture agnostic device interface. OpenCAPI enables the development of host-agnostic devices which can coherently connect to any host platform which supports the OpenCAPI standard, providing device capability to coherently cache host memory to facilitate accelerator execution, perform DMA and atomics to host memory, send messages and interrupts to the host, and act as a host memory home agent. OpenCAPI utilizes high-frequency differential signaling technology while providing high bandwidth in addition to low latency needed by advanced accelerators. OpenCAPI encapsulates the serializing cache access and address translation constructs in high speed host silicon technology to minimize overhead and design complexity in attached silicon such as FPGAs and foundry ASICs. Finally, OpenCAPI architecturally ties together transaction layer, link layer, and physical layer attributes to optimally align to high SERDES ratios and enable high bandwidth, highly parallel exploitation of attached silicon.

OpenMP: Enabling HPC for Twenty Years

Michael Klemm (OpenMP Architecture Review Board)

OpenMP® is twenty this year. Since its inauguration in 1997, it has become the standard programming model for multi-threading in HPC applications and drove many scientific discoveries by helping domain scientists implement multi-threading in their simulation codes. The OpenMP API is based on directives to augment code written in C/C++ and Fortran with parallelization hints to the compiler.

In its 20 years of existence, OpenMP has evolved from a purely multi-threaded model to a programming model that supports modern task-based programming as well heterogeneous programming for offload devices such as GPUs. With its extensions it also supports data-parallel programming for single-instruction multiple-data architectures.

This presentation reviews the history of OpenMP, the current features and provides an outlook for the upcoming OpenMP API specification, OpenMP Technical Report 6: Version 5.0 Preview 2, which will be released during SC'17. The final version of OpenMP version 5.0 is planned for SC18.

Understanding Gen-Z Technology – A High Performance Interconnect for the Data-Centric Future

Michael Krause (Gen-Z Consortium)

Gen-Z is a high-bandwidth, low-latency fabric with separate media and memory controllers that can be realized inside or beyond traditional chassis limits. It treats all components as memory (memory-semantic communications), and moves data with minimal overhead and latency, taking full advantage of emerging persistent memory (memory accessed over the data bus at memory speeds). Gen-Z can also handle compute elements, such as GPUs, FPGAs, and ASIC or coprocessor-based accelerators, with no need for extra copy operations, special DMA channels, or complex error recovery schemes. Separate controllers allow scaling of processing, media, and memory with an end result of higher throughput and lower complexity for big data solutions in such applications as data analytics, deep packet inspection, artificial intelligence, machine learning, and video and image processing. Learn about this new approach and how it promotes innovation during this presentation.

Room: 501-502

3:30 pm - 5:00 pm

Data Analytics

Exploiting HPC for Big Data, Analytics and AI

Rangan Sukumar (Cray Inc)

In this presentation, we will cover the use of HPC technologies for Big Data problems. We will share the design philosophy behind the architecture of Cray analytic systems and discuss results from benchmarks on three kinds of workloads: graph analytics, matrix factorization and deep learning training. These results will demonstrate how the combination of the network interconnect and application of HPC best practices enables the ability to process 1000x bigger graph datasets up to 100x faster than competing tools on commodity hardware, provides a 2-26x speed-up on matrix factorization workloads compared to cloud-friendly Apache Spark and promises over 90% scaling efficiency on deep learning workloads (i.e. potentially reducing training times from days to hours). We will end by presenting success stories of organizations that have leveraged HPC thinking both in the enterprise and scientific computing sectors.

From HPC-as-a-Service to Deep Learning-as-a-Service

Patrice Calegari (Bull)

High Performance Computing (HPC) and Deep Learning (DL) share many characteristics: intensive computing, large datasets, and need for easy access to clustered resources by end-users. We predict that DL usage generalization will boost HPC market growth. Conversely, HPC community experience and the large installed base of HPC infrastructures will boost DL growth. Both HPC and DL are evolving together to the as-a-service model by reusing and adapting matured HPC and Cloud concepts such as massive scalability, resource managers, containers, orchestration mechanisms, GPU Computing, batch schedulers, HPC-as-a-Service software, HTTP RESTful-APIs, and web user interfaces. Data scientists need a high-level DL API and a web user interface that both hide HPC systems' complexity. We will illustrate our move from HPCaaS to DLaaS by showing how we manage DL training tasks and frameworks on standard HPC clusters through a web user interface. Pros and cons of possible architecture choices will be discussed.

Genomic Computations at Scale with Serverless, and Docker Swarm

Dmitri Zimine (Brocade Communications Systems, StackStorm Inc)

This talk is a story of bio-tech meeting DevOps to produce genomic computations, economically, and at scale.

Genomic computation is growing in demand as it comes to the mainstream practices of bio-technology, agriculture, and personal medicine. It also explodes the demand for compute resources. In fact, with inexpensive next-gen sequencing, some labs sequence over 1,000,000 billion bases per year. Genetic data banks are growing over 10x annually. How to compute the genomic data at massive scale, and do it in a cost-efficient way? In the presentation, we describe and demonstrate a serverless solution built with Docker, Docker Swarm, StackStorm and other tools from the DevOps toolchain on AWS. The solution offers a new take on creating and computing a bio-informatic pipelines that can run at high scale and at optimal cost.

Room: 503-504
3:30 pm - 5:00 pm

Programming Tools

Compiler-Assisted Software Testing for Heterogeneous Computing Systems

Michael Wolfe (NVIDIA, Portland Group)

When optimizing or porting HPC applications, it's common to see numerical differences. This is even more common on heterogeneous systems where execution transitions between CPU and GPU. Finding the exact point where the computations diverge is challenging and labor intensive. We discuss a new PGI Compiler-Assisted Software Testing (PCAST) feature to address this problem. PCAST will allow a programmer to save intermediate and final results of an existing application via API runtime calls or directives that create a reference file. The same runtime calls or directives in the optimized or ported application compare the intermediate and final results to the reference file. For OpenACC applications on a heterogeneous system, dynamic results checking is supported by executing each parallel construct on the CPU as well as the GPU and comparing the results after each construct. We'll cover these and other potential applications that can simplify HPC application development and tuning.

Approaches to Debugging Mixed-Language HPC Apps

Jasmit Singh (Rogue Wave Software)

The HPC community is embracing the advantages of mixed-language development environments, presenting challenges for debugging and testing when application execution and data flow cross languages. How can we take advantage of the unique features offered by different languages while minimizing the impact on bug reproduction, root-cause analysis, and solution?

This interactive session walks through the current mixed-language HPC landscape to describe the problems with testing these types of applications and best-practice solutions using TotalView for HPC. You will learn how these architectures make it easy to "steer" computation between modules of different languages, to accelerate prototyping and development, and how advanced testing techniques provide visibility into the call stack and data for efficient debugging. The session will include presentation material and demonstrations using mixed Python and C/C++ applications.

Khronos SYCL: Tomorrow's Heterogeneous C++ and C Today

Michael Wong (Khronos Group Inc., Codeplay Software Ltd)

SYCL is a Khronos specification for heterogeneous computing built on top of OpenCL and C++. The SYCL 1.2 specification was published on May 2015, and the current SYCL 2.2 specification has been published on February 2016. Behind these two specifications, there has been an important community effort ongoing for more than five years. Now the specifications is available and Codeplay is releasing their ComputeCpp Community Edition as well as optimized version commercially supported for many cpu+cpu/dsp/fpga combinations.

This talk will demonstrate how SYCL can be used today to support your exascale C++ effort, leading to ISO-ready C++ code that can support numerical computations on any compute node+gpu combinations, while staying with the current and future C++ standard direction for executors and heterogeneous computing.

Wednesday, November 15th

Room: 501-502
1:30 pm - 3:00 pm

HPC Trends

The Silver Lining of the Cloud is the EDGE

Revathi Narayanan (Micron Technology Inc)

The next big wave of Edge computing is driven by the number of intelligent devices that proliferate across consumers, manufacturing, vehicles, households and remote installations. By 2020, the explosive Edge Computing growth to 50 billion devices will generate 44 zettabytes of data annually (44 trillion GB). Edge computing requires real-time analysis for critical decision making,

needing high performance processing power at the edge. For real-time analysis, processing must take place away from the cloud; at the device level or distributed across the network. This will drive increased demands on compute capability and memory and storage solutions to support edge processing. Micron, as a leading provider of both memory and storage solutions, is a key enabler of Edge Compute growth. Micron will present trends and associated requirements across Edge Compute applications.

ARM's Road to Exascale

Eric Van Hensbergen (ARM Ltd)

ARM's business model is based on licensing intellectual property (IP) to partners. At the same time, it has built a software ecosystem that is leveraged by, contributed to and maintained by the partnership. The efforts of ARM and its partners in the domain of high-end systems are creating a diverse ecosystem of components for high-performance computing (HPC) and high-performance data analytics (HPDA). The key enabling technologies include the ARMv8 architecture including the scalable vector extension (SVE); optimized operating systems, compilers, math libraries, development environment, debugging and profiling tools; and reliability, availability and serviceability (RAS) support. This talk will focus on ARM's journey towards exascale through co-design with partners in US, European, and Japanese programs; cover the current state of HPC relevant developments within the ARM hardware and software ecosystem; and give an overview of some existing and future ARM system deployments.

Efficient Managing and Monitoring of InfiniBand HPC Clusters

Tor Skeie (Fabriscale Technologies AS, Simula Research Laboratory)

The Fabriscale Monitoring System (FMS) is a cluster interconnect monitoring software that provides visual insight into the status of your InfiniBand cluster. In this presentation we present the key features of the FMS and show you how to get an overview of performance and drill-down into statistics, alerts and key metrics. With FMS monitoring of the cluster is automated and alarms are only raised when the operator's attention is required. The operator will be pointed to where the problem has occurred, supported by relevant metrics and statistics. This saves time, leads to faster error recovery, less strain on operators and reduced downtime for your cluster.

The FMS integrates with job schedulers, e.g. Slurm, MOAB HPC Suite, and PBS Works, to leverage scheduling information to present performance information as a function of workload. Potential network bottlenecks can be identified per job, and utilization for a job can be specified per port.

Room: 503-504

1:30 pm - 3:00 pm

Interconnects

Interconnect Your Future with Mellanox "Smart" Interconnect

Gerald Lotto (Mellanox Technologies)

Pushing the frontiers of science and technology will require extreme-scale computing with machines that are 500-to-1,000 times more capable than today's supercomputers. We will discuss the new era of data-centric computing which enables data processing everywhere; including the network. The next generation of intelligent interconnect devices, such as Mellanox/s Quantum™ HDR 200gbps InfiniBand Switching with SHARP (Scalable Hierarchical Aggregation and Reduction Protocol) and ConnectX-6 200Gb/s VPI Host Channel Adapters enable in-network co-processing, enabling communication framework processing and user algorithm processing on the interconnect fabric increases system performance by an order of magnitude or more. We will explore the benefits of the next generation capabilities of these smart-interconnect devices, off-load architecture and in-network co-processing as we prepare to move toward the next milestones at and beyond exascale computing.

Protecting against Hyper Scale Network Attacks with Bump-in-the-Wire 100G filtering

Carolyn Raab (Corsa Technology)

This presentation will cover how difficult it is to protect high capacity network links (100G) against increasing frequency and size of cyber attacks. The status quo of using gateway routers or firewalls can only manage to a point, especially in HPC environments where large data transfers are so prevalent. I will introduce a new concept of using a bump-in-the-wire programmable filter that can be added to networks transparently without needing to change any of the network's architecture. Sometimes the simplest ideas can be the most useful. I would like to use this forum to be as interactive as possible and invite participants to bring their network security challenges to the session for discussion. Corsa's lead architect and CTO will be attending so the conversation can go as technically deep as needed.

A Networked-FPGA Platform Offering Flexible Ethernet Switching from Layer 1 All the Way to Full SDN via P4

Matthew Knight (Metamako LP)

Metamako devices provide a flexible and compelling platform for multi-layer Ethernet switching. This session will cover the benefits of Layer 1 switching including ultra-low-latency Ethernet fan-out, software-defined port patching, and media conversion (including

DWDM).

It will also cover the devices' ability to leverage their integrated FPGAs to implement SDN via SDNet and P4.

Room: 501-502
3:30 pm - 5:00 pm

FPGAs in HPC

Efficiently Accelerating HPC Workloads with FPGAs

Michael Strickland (Intel Corporation)

With the new classes of emerging workloads (e.g. ML, Big Data) there are increasing opportunities in the data center for FPGA algorithm, networking, and data access acceleration. For example, Microsoft has announced that they will accelerate Bing search, machine learning, and networking with FPGAs. Key advances in FPGA hardware architecture and integration with Xeons are also opening up new and broader opportunities for Cloud and Enterprise with significant performance increases in areas such as networking, search, machine learning scoring, and data analytics.

This presentation will provide an overview on the role and the impact of FPGAs in HPC and demonstrate how FPGAs are efficiently accelerating key HPC workloads, such as genomics, machine learning, video analytics and big data analytics.

Enabling FPGAs for the Software Developers

Bernhard Friebe (Intel Corporation)

FPGAs play a critical part in heterogeneous compute platforms as flexible, reprogrammable, multi-function accelerators. They enable hardware performance with the programmability of software. The industry trend towards software-defined hardware challenges not just the traditional architectures - compute, memory, network resources - but also the programming model of heterogeneous compute platforms.

Traditionally, the FPGA programming model is narrowly tailored and hardware-centric. As FPGAs become part of heterogeneous compute platforms and users expect the hardware to be "software-defined", they must be accessible not just by hardware developers but by software developers which require the programming model of FPGAs to evolve dramatically. This presentation focuses on outlining a highly evolved, software-centric programming model which will enable FPGAs for software developers through a comprehensive solutions stack including FPGA-optimized libraries, compilers, tools, frameworks, SDK integration, and an FPGA-enabled ecosystem. We'll also show a real-world example using Machine Learning Inference acceleration on FPGAs.

Cooling Hot FPGAs: A Thermals First Approach

Ron Huizen (BittWare Inc), Jeff Milrod (BittWare Inc)

Today's largest FPGAs have impressive performance potential—with Intel's Stratix 10 boasting 9.2 TFLOPS and Xilinx' UltraScale+ VU13P offering 3.8M logic elements. This, in addition to direct access to high-speed memory and networking, is making FPGAs a top choice for high-performance computing in datacenters.

However, the potential continuous power draw of these large FPGAs is now 200 watts and up, meaning significant heat that must be removed. While it's easy to simply add liquid cooling or fans if you're targeting a desktop, datacenter servers prefer compact, passive solutions.

The solution is a thermals first approach. This isn't about a particular technology (although we'll cover the latest options) but a combination of elements, including physical board design, flow simulation informed by a proper understanding of airflow in servers, and more. This presentation will be informative for anyone seeking to learn about the latest large FPGAs in production servers.

Room: 503-504
3:30 pm - 5:00 pm

Storage

Spectra Logic Delivers a New Paradigm in Tape Library Deployment

Matt Starr (Spectra Logic Corporation)

There are often many challenges associated with migrating from one storage hardware technology to another. Tape libraries, capable of managing supercomputing workloads, are large and complex and can be difficult to integrate into one's data center. IT staff may be faced with significant acquisition costs, extensive executive sign-off and budget approvals, large footprint requirements in the data center, and multiple support contracts for existing and new technology.

Spectra now offers a TFinity Research Platform that allows customers to test performance, reliability and functionality of its Spectra TFinity Tape Library. The platform provides direct access to Spectra's executives and engineers, while reducing deployment time and costs, and mitigating risk associated with installation of new technologies. Attend today's discussion to learn more about this cost-effective, high-performance research storage solution, as well as how the current tape technology roadmap makes tape an extremely viable storage medium.

Burst Buffers: Flash in the Pan?

James Coomer (DataDirect Networks)

Burst Buffers have been associated with next-gen, exascale, leading-edge HPC scratch systems for a few years now. For the performance and capacity requirements of large scale IO solutions the economics still mandates a cache approach rather than all-flash. The implementation of "burst buffers" is becoming mainstream as solid state inevitably works its way into HPC storage albeit in a number of guises: classic exascale burst buffers through to more modest controller caching and a number of alternatives in between.

This Exhibitor Forum will discuss the relative merits of flash implementations in HPC environments, with respect to economics, usability, performance and manageability and encourages organizations offering flash caching solutions, and implementers of flash caches to contribute to the assessment of this dynamically evolving area of HPC. Specific application verticals which are primarily benefitting from flash in HPC will be discussed and extreme acceleration examples and use cases will also be shared.

Building End-to-End NVMe over Fabric Infrastructure for HPC

Rick Kumar (Sanmina Corporation)

The goals of non-volatile memory (NVM) storage for high performance computing (HPC) generally involve minimizing storage latency, especially when the NVM storage use case is local and dedicated SSDs for burst buffers or parallel file systems, or pooled and shared SSDs for parallel file system or object storage acceleration. Common high-performance SSD interconnection technologies include Serial Attached SCSI (SAS) and PCI Express® (PCIe®), yet even PCIe alone is no longer sufficient to meet the performance requirements of present and future HPC architectures or Storage Class Memory (SCM) technologies.

During June 2016, The NVMe Express® (NVMe®) organization published the NVMe over Fabrics 1.0 (NVMeoF) specification defining a common architecture supporting the NVMe block storage protocol over a range of networking fabrics. Commercial off the shelf (COTS) products are becoming available enabling end-to-end NVMeoF infrastructures and one reference architecture will be presented demonstrating next-generation latency, throughput, and bandwidth for high-performance storage.

Thursday, November 16th

Room: 503-504

10:30 am - 12:00 pm

System Cooling

Future of the Thermal Management – Thercon-LHP Water-Free Solutions for HPC Cooling

Yury F. Maydanik (Institute of Thermal Physics Ural Branch, Thercon-LHP)

Loop Heat Pipe is a two-phase heat-transfer device for providing an effective thermal link between a heat source and a heat sink in cooling systems for HPC. Vaporization and condensation of working fluid circulating inside the loop of tubes provides heat transfer and capillary forces serve as a driving mechanism. Capillary structure acts as a pump without moving parts providing high reliability.

LHP heat transfer technology enables a range of water free cooling solutions for different electronic components like CPU, memory, GPU. Developed for high heat load applications in particular for HPC, they are easily adaptable and work without electric energy consumption.

LHP conception allows a wide variety of different design, which essentially extends the application fields and performance capabilities. LHP leave behind conventional cooling systems due to higher heat-transfer ability, compactness and flexibility. Technology improves productivity, lowers energy consumption, and provides high component density

Chip-to-Atmosphere: Providing Safe and Effective Cooling for High-Density, High-Performance Data Center Environments

Geoff Lyon (CoolIT Systems Inc), David Meadows (Stulz Air Technology)

To keep up with their customers' ever-increasing demands, high-tech manufacturers are now introducing high-performance processor chips and servers that require liquid cooling. Liquid-cooled chips include the Intel Xeon 150 chip (and the forthcoming Xeon 215W – 350 W), and the Nvidia GeForce GTX 1080 graphics card. As companies demand more from their IT footprints, liquid cooling in the data center is rapidly becoming a necessity for high-density heat loads. But while the transition from air cooling to liquid cooling may be inevitable, it does not have to be exceptionally painful for enterprise IT companies. In this presentation, CoolIT

and STULZ will look at the benefits of liquid cooling. In particular, we'll highlight the Chip-To-Atmosphere method as an example of an effective cooling solution for high-capacity IT footprints.

Benefits of Having Sensors in Your Water Cooled HPC

Mads Melchior (Grundfos), Lars Mejsner (Grundfos)

Grundfos, world leading manufacturer of pumps and controls, also manufactures sensors and transmitters for: flow, pressure, differential pressure and temperature.

In this session, we will talk about benefits of having sensors in your water cooled HPC system. Included are: integration levels of sensors, a selected number of theoretical and customer cases, and the functionality of Grundfos MEMS and Vortex technology.

Exhibits [\(back to top\)](#)

Tuesday, November 14th

Room: Exhibit Halls A, B, E and F
10:00 am - 6:00 pm

Exhibits

Exhibits

Come visit with over 300 exhibitors that range from industry to academia and research. There is sure to be something for everyone.

Wednesday, November 15th

Room: Exhibit Halls A, B, E and F
10:00 am - 6:00 pm

Exhibits

Exhibits

Come visit with over 300 exhibitors that range from industry to academia and research. There is sure to be something for everyone.

Room: Exhibit Halls A, B, E and F
4:00 pm - 6:00 pm

Family Day

Family Day

On Wednesday, November 15, from 4:00pm to 6:00pm, families are invited to walk through the Exhibit Hall. Everyone will be required to have a badge and children (12 years and older) must be accompanied by an adult at all times. If you need a badge for a family member, go to the Attendee Registration area.

Thursday, November 16th

Room: Exhibit Halls A, B, E and F
10:00 am - 3:00 pm

Exhibits

Exhibits

Come visit with over 300 exhibitors that range from industry to academia and research. There is sure to be something for everyone.

HPC Connects Plenary [\(back to top\)](#)

Monday, November 13th

Room: Mile High Ballroom
5:30 pm - 6:30 pm

HPC Connects Plenary: The Century of the City

Moderator: *Charlie Catlett (Argonne National Laboratory)*

HPC Connects Plenary: The Century of the City

Pete Beckman (Argonne National Laboratory), Debra Lam (Georgia Institute of Technology), Michael Mattmiller (City of Seattle)

The 21st Century is frequently referenced as the “Century of the City,” reflecting the unprecedented global migration into urban areas, and their central role in both driving and being affected by factors ranging from economics to health to energy to climate change. The notion of a “smart” city is one that recognizes the use and influence of technology in cities. For instance, the use of technology to enable cities to more effectively harness their roles as centers of opportunity, while address the challenges of providing equity of opportunity across a heterogenous population.

High Performance Computing (HPC) is already playing a key role in helping cities to pursue these objectives. Intelligent devices enabled with HPC “at the edge” have potential to enable real-time optimization in areas such as energy generation and delivery or the flow of goods and services, and to allow urban infrastructures to adapt autonomously to changes and events such as severe storms or traffic congestion.

With the accelerating volume of data about cities, HPC is helping cities to optimize their services, from making food safety inspections more effective to identifying children most at risk for lead poisoning. HPC is supporting the creation of computational models of urban sectors such as transportation, energy demand, or economics. And increasingly there is opportunity—and need—to develop multi-scale, coupled modeling systems that harness new data sources and measurement techniques and that capture the interdependencies between these sectors to provide a more holistic modeling capability for urban designers, planners, and developers.

This plenary panel will discuss emerging needs and opportunities suggesting an increasing role for HPC in cities, with perspectives from city government, planning and design, and embedded urban HPC systems.

HPC Impact Showcase [\(back to top\)](#)

Tuesday, November 14th

Room: 501-502

10:30 am - 12:00 pm

HPC Impact Showcase: Healthcare and Manufacturing

Impact of the DOE and NCI Partnership on Precision Oncology

Rick Stevens (Argonne National Laboratory)

The Cancer Moonshot was established in 2016 with the goal of doubling the rate of progress in cancer research. A major component is the strategy to use modeling, simulation, and machine learning to advance our understanding of cancer biology and to integrate what we know into predictive models that can guide research and therapeutic developments. In 2015, the U.S. Department of Energy (DOE) formed a partnership with the National Cancer Institute (NCI) to jointly develop advanced computing solutions for cancer by bringing together researchers from four DOE laboratories (Argonne, Los Alamos, Livermore, and Oak Ridge) with the Frederick National Laboratory for Cancer Research (FNLCR). This integrated team has launched three pilot projects, each addressing a major challenge problem on the forefront of precision oncology: (1) provide better understanding and eventually develop new drugs for the RAS oncogene family of cancers which impact 30% of cancers; (2) develop models that can predict tumor response to drugs to enable physicians to more precisely target an individual patient's tumor; and (3) analyze electronic medical records of millions of cancer patients to streamline the introduction of new precision oncology therapies. The CANDLE (CANcer Distributed Learning Environment) project aims to develop an end-to-end computational environment to bring deep learning to these key problems. This talk introduces the cancer moonshot, describes the three overarching cancer problems, and provides a roadmap of the CANDLE project. We will discuss the impact of deep learning, advanced modeling, and simulation on cancer research and future approaches to treating patients.

Using HPC to Impact US Manufacturing through the HPC4Mfg Program

Lori Diachin (Lawrence Livermore National Laboratory)

Modeling and simulation on high performance computing (HPC) systems can be used to accelerate innovation, save energy, and reduce costs in manufacturing processes. We present an overview of DOE's HPC4Mfg program, which aims to increase adoption of HPC in manufacturing settings run by both large and small companies. The program has funded over 40 projects to date, and we give several case studies that showcase the impact HPC can have on various a wide range of manufacturing sectors, from consumer products and steel manufacturing to clean energy technology development.

Development of High Performance Rubber Materials for Tires Using K Computer

Masato Naito (Sumitomo Rubber Industries Ltd)

The rapid development of high-performance rubber materials has become increasingly essential for the development of high-performance, high-quality tires that combine vehicle safety with environmental friendliness. The rubber compounds that make up tires contain many different types of materials, from their polymer framework to reinforcing agents such as silica and carbon black, additives such as oil and resin, as well as crosslinking agents. The interrelationships between these materials are extremely complex, and it is the interaction between them that manifest in a rubber's performance.

We have recently developed large-scale molecular simulation of rubber materials to analyze and elucidate the relationship between various factors at the molecular level and the rubber performance. This new simulation achieves realistic and highly precise simulations of rubber materials at the molecular level through coordinated utilization of such world-class research facilities as the SPring-8*, J-PARC** and the K computer. Applying this simulation to the observation of rubber failure has enabled us to observe, for the first time, the entire breakdown process from its origin in the formation of voids at the molecular level to the development of these voids into cracks.

In this presentation, we will explain some examples of tire development using our simulation technology, including examples of developing tires with high wear resistance using the large-scale molecular simulation.

* SPring-8 is large scale synchrotron radiation facility. (Location: Sayo, Hyogo, Japan) ** J-PARC is a proton accelerator and experimental facility. (Location: Tokai, Ibaraki, Japan)

Wednesday, November 15th

Room: 503-504

10:30 am - 11:30 am

HPC Impact Showcase: Computational Modeling

Making HPC Consumable: Helping Wet-Lab Chemists Access the Power of Computational Methods

Kirk Jordan (IBM, Hartree Centre)

The use of HPC and simulation to support experimental science can greatly increase laboratory efficiency and provide additional insights into interactions not easily described by traditional methods. However, obtaining meaningful insight from simulations often involves significant investments in computational and human resources. This presents a major barrier to the widespread adoption of computational methods as a driver for laboratory exploration.

To overcome this consumability problem non-experts should be able to obtain the results of cutting-edge computational science models executing on cutting-edge high performance systems as easily as they use a piece of wet-lab equipment. Such “computational appliances” would encapsulate the state-of-the-art computational know-how, and automate the generation of final results from a few input parameters. The user’s interaction with these appliances would focus on the scientific functions they provide rather than the underlying technology they run on.

In this presentation we describe the current status of a collaboration between the Hartree Centre (a collaboration between IBM Research and the Science and Technologies Facilities Council of the U.K.) and our industrial partners that is driven by interactions with our partners’ wet-lab scientists. As a result of these engagements we have developed in-silico counterparts to three laboratory experiments they commonly perform as part of their R&D activity. One goal of this project is to demonstrate this technology in an operational industrial environment by providing the experimentalists with access to production computational appliances which they will use to augment their work.

Gaining Insights into the Properties of Materials Using Atomistic Simulations on Large-Scale HPC Platforms

Danny Perez (Los Alamos National Laboratory)

Many of the macroscopic properties of materials are rooted in the details of their structure at the atomic scale. For example, the properties of real materials are often strikingly different from those predicted by assuming a perfectly crystalline state. Indeed, perhaps contrary to intuition, nano or micro-scale features such as point defects, dislocations, or grain boundaries, often dictate the performance of materials. In order to optimize desirable properties or avoid catastrophic failure, it is hence crucial to be able to perform simulation of materials with full atomistic resolution in both space and time. One of the most powerful methods to do so is Molecular Dynamics (MD), i.e., the direct integration of atomic equations of motion. MD is extremely powerful but also computationally intensive, due to the need to resolve the motion of each individual atom. Leveraging HPC resources is therefore critical and a large fraction of the computing budget of national supercomputing centers is currently spent on such calculations.

Through different examples, I will show how massively-parallel HPC platforms provide unique opportunities to access the time and length scales required to make accurate predictions of the behavior of materials. Doing so, I will pay special attention to recently-developed techniques that leverage parallelism to extend the simulation timescales that are amenable to direct MD simulations into the milliseconds, thereby approaching experimentally relevant timescales.

Room: 501-502

10:30 am - 12:00 pm

HPC Impact Showcase: Energy and Climate

HPC Powers Wind Energy

Earl P.N. Duque (Intelligent Light)

Electrical power generation via large scale wind turbine farms continues to grow worldwide as a cost-effective renewable energy resource. The proper placement of wind turbines within a given resource area must consider the optimal power capture and the wake interference effects from turbine to turbine. Flow simulations at the scale afforded by HPC systems will enable wind farm developers to account for these effects and reduce the total cost of energy.

Prof. Dimitri Mavriplis and his students at the University of Wyoming have been developing W2A2KE3D code, a high order finite element solver instrumented with VisIt/Libsim for in situ data processing. It has been applied to various aerodynamic applications such as airplanes, helicopters and now wind turbines. This HPC Impact Showcase will highlight how the Cheyenne supercomputer located at National Center for Atmospheric Research / University Corporation for Atmospheric Research (NCAR/UCAR) Computational and Information Systems Lab (CISL) was used to simulate the Lillgrund Wind Plant. This work represents the most highly resolved simulation of a wind farm to date. It includes the detailed unsteady fluid structure interactions occurring within each turbine and has the capability for atmospheric turbulent inflow conditions. These simulations required over 1.1 billion degrees of freedom to perform the simulation on over 32,000 compute cores.

Supercomputing for Everyone: Meeting the Growing Needs of Businesses

Elie Hachem (Mines ParisTech), Romain Klein (Transvalor SA, Aeromines)

As companies across diverse industries recognize the huge benefits that high-performance computing (HPC) can unleash, the appetite for HPC services shows no sign of slowing. Universities and research teams, so often on the boundaries of innovation, are uniquely positioned to uncover and drive forward new applications for these solutions.

Aeromines is a scientific computing cloud platform for computational fluid dynamics developed at Mines Paristech and marketed by Transvalor to industrial clients in several domains: Aerospace, Defense, Automotive, Construction, Energy, etc.

We present this unique project based upon cluster virtualization technology from IBM that brings students, researchers and industries together in a secure, flexible and collaborative high-performance computing environment – reducing the costs of complex computational modeling, providing unlimited storage and total security and shortening the time taken to test new ideas by 91%.

Additionally, this session will cover the advantages of this innovative solution and highlight the benefits of this collaboration including: • A flexible hybrid-cloud infrastructure tailored for HPC providing on-premise performance and the ability to easily burst to the cloud as required. • A complete PaaS solution to easily host the Aeromines software stack with enterprise grade scheduling • Security, with isolated resources to either a shared cluster or dedicated clusters for clients with highly confidential data • Flexibility with the ability to grow or shrink resources depending on client demands.

Thursday, November 16th

Room: 501-502

10:30 am - 12:00 pm

HPC Impact Showcase: Defense Systems

Accelerating Innovation of Defense Systems with Computational Prototypes and High Performance Computing

Douglass E. Post (US Department of Defense HPC Modernization Program)

The US Department of Defense High Performance Computing Program (DoD HPCMP) provides the DoD computational science and engineering community three key elements of an HPC ecosystem: 1) five computer centers with a total computing capacity of 31 PetaFLOPS, 2) DREN, a high-bandwidth, low-latency network to connect the customers with the computers, and 3) application software support. The exponential growth in computing power over the last 60 years now gives DoD engineers the ability to accurately predict the performance of full-scale weapon systems (e.g., a full-size ship or airplane). To realize this capability, the CREATE program was launched in 2007 to develop and deploy eleven physics-based HPC engineering tools for the design and analysis of ships (e.g. aircraft carriers, destroyers,...), air vehicles (fighters, bombers, submarine surveillance airplanes,...), ground vehicles (trucks, personnel carriers, tanks,...), and radio frequency radars and antennas, together with the ability to generate the meshes and geometries needed for analysis. These tools enable engineers to develop virtual prototypes of weapon systems. By analysis of those prototypes with high-fidelity, physics-based tools, design defects and performance shortfalls can be identified and fixed before metal has been cut. This reduces rework and the risks, costs, and time of acquiring and maintaining major DoD weapon systems. The tools are being used by more than 160 DoD engineering organizations (government, industry, and academia) and approximately 1,400 users. Illustrative examples of these applications and their impacts will be described during the presentation.

Accelerating Defense Innovation of US Naval Vessels with Computational Prototypes and High Performance Computing

Wesley M. Wilson (US Naval Surface Warfare Center)

Navy Enhanced Sierra Mechanics (NESM) builds on and enhances the DOE Sandia National Laboratory Sierra Mechanics software suite to enable assessment of ship structure and component response to weapons loading using physics-based, massively parallel, high performance computational tools. NESM is used to predict both ship shock response and structural damage, including severe cases. The NESM toolkit supports an evolving Navy process to reduce risk and cost associated with full ship shock testing. The toolkit is positioned to improve initial ship design processes by providing shock and associated equipment vulnerability assessment prior to final arrangement and installations decisions. The tightly coupled multi-physics capabilities include: structural dynamics, solid mechanics, fluid dynamics, fluid-structure interaction, and shock physics. The HPC capability enables engineers to address full-scale naval vessels including next generation aircraft carriers and submarines.

NavyFOAM, based on OpenFOAM, has been tailored specifically to simulate the air-sea interface, seaway loading, propulsor-hull interactions, and other hydrodynamic effects important for naval vessels. It has been extensively validated for underwater vehicles and surface ships based upon experimental data and ship trials information. It has been applied to many different naval systems to gain physical insights into complex flow phenomena and to guide and support physical model testing. NavyFOAM reduces design risk by leveraging HPC resources with high-fidelity physics prediction capabilities to address complex ship design questions and characterize complex ship performance earlier in the design cycle.

Examples of how NavyFOAM and NESM support workflows to enhance physical understanding and reduce physical testing requirements will be presented.

Accelerating Defense Innovation of Military Aircraft with Computational Prototypes and High Performance Computing

Robert L. Meakin (US Department of Defense HPC Modernization Program)

The CREATE Air Vehicles Project has been developing and deploying three tools: 1) Kestrel, a high-fidelity, full-vehicle, multi-physics, analysis tool for fixed-wing aircraft, 2) Helios, a similar capability for rotary-wing aircraft, and 3) CREATE Genesis-DESIGN, a conceptual design tool. This talk focuses on the use of Kestrel. Kestrel is a multi-physics code that includes the unique ability to calculate the combined effects of sub-sonic, transition, and supersonic aerodynamics, structural dynamics, flight control through control surfaces, and propulsion systems (gas-turbine engines). Kestrel has been used to predict flight performance prior to construction, evaluate planned operational scenarios, perform flight certifications, and plan and rehearse test campaigns. Kestrel has been applied to the analysis of over 30 fixed-wing DoD aviation systems including store separation, A-10, F-18E, F-15, B-52, E-2D, P-3, and many others. In all cases, the combination of supercomputing hardware with high-fidelity, multi-physics engineering codes have been critical to application success. Examples of a subset of these applications will be presented.

Invited Talk [\(back to top\)](#)

Tuesday, November 14th

Room: Mile High Ballroom

10:30 am - 12:00 pm

Invited Talks 1

The U.S. D.O.E. Exascale Computing Project – Goals and Challenges

Paul Messina (Argonne National Laboratory)

In 2016, the U.S. Department of Energy established the Exascale Computing Project (ECP) – a joint project of the DOE Office of Science (DOE-SC) and the DOE National Nuclear Security Administration (NNSA) – that will result in a broadly usable exascale ecosystem and prepare mission critical applications to take advantage of that ecosystem.

This ambitious project aims to create an exascale ecosystem that will:

- Enable classical simulation and modeling applications to tackle problems that are currently out of reach,
- Enable new types applications to utilize exascale systems, including ones that use machine learning, deep learning, and large-scale data analytics,
- Support widely used programming models as well as new ones that promise to be more effective on exascale architectures or for applications with new computational patterns, and
- Be suitable for applications that have lower performance requirements currently, thus providing an on ramp to exascale should their future problems require it.

Balancing evolution with innovation is challenging, especially since the ecosystem must be ready to support critical mission needs of DOE, other Federal agencies, and industry, when the first DOE exascale systems are delivered in 2021. The software ecosystem needs to evolve both to support new functionality demanded by applications and to use new hardware features efficiently. We are utilizing a co-design approach that uses over two dozen applications to guide the development of supporting software and R&D on hardware technologies as well as feedback from the latter to influence application development. To ensure that the software technologies developed by the ECP will integrate into the facilities' software stack, we also work closely with the supercomputing facilities that will host the exascale systems through development and innovation cycles. The project will also conduct R&D on tools and methods to boost productivity and portability between systems.

Taking the Nanoscale to the Exascale

Theresa Lynn Windus (Iowa State University, Ames Laboratory)

Chemistry is a central science in that, at its core, chemistry is the understanding of the way atoms and molecules interact. Chemistry influences many areas such as physics, biology, pharmaceuticals, materials, and chemical engineering. Computational chemistry, while a relatively young part of the chemistry field, has become a third pillar in the science next to experimentation and theory. The demand for understanding and prediction of experimental results through computation has never been higher, and the physical models that underlie the chemistry are continually improving and evolving to meet new scientific challenges. All of these pieces together, and the changing ecosystem of hardware and software, make the development of computational chemistry codes a larger challenge than we have ever experienced.

This talk will focus on the challenges that computational chemistry faces in taking the equations that model the very small (molecules and the reactions they undergo) to efficient and scalable implementations on the very large computers of today and tomorrow. In particular, how do we take advantage of the newest architectures while preparing for the next generation of computers? How do we increase programmer productivity while ensuring excellent performance, efficiency and portability across multiple platforms? How do we take advantage of the work of mathematicians, computer scientists and other computational scientists to enable our science, while ensuring maintainability and usability of the software? How do we ensure that the algorithms that we develop are making wise use of the computational resources? How do help the next generation of computational chemists to be ready for the complex computing environments that they will face? While not claiming to have answers to all (or any!) of these questions, we will explore some possible solutions and their implications as we go forward and face the current petascale and the future exascale challenges.

Room: Mile High Ballroom

1:30 pm - 3:00 pm

Test of Time Award Special Lecture

Why Iteration Space Tiling?

Michael Wolfe (Portland Group)

Tiling subdivides the multidimensional iteration space of a nested loop into blocks or tiles. Tiles become a natural candidate as the unit of work for parallel task scheduling, where scheduling and synchronization is done at tile granularity, reducing overhead at some loss of potential parallelism. Tiling allows separation of concerns between local optimizations (vectorization, cache footprint) and global optimization (parallelization, scheduling for locality). Tiling is well-known and has been included in many compilers and code transformation systems. The talk will explore the basic contribution of the SC1989 paper to the current state of iteration space tiling.

Room: Mile High Ballroom

3:30 pm - 5:00 pm

Invited Talks 2

Thirty Years of the Gordon Bell Prize

Gordon Bell (Microsoft)

The Gordon Bell Prizes have chronicled the important innovations and transitions of HPC beginning with the first award in 1987. Gustafson, Montry, and Benner demonstrated Amdahl's Law was not impenetrable. By 1994, MPI provided both a model and standard for rapid adoption, Beowulf became the recipe for building clusters, and a Thinking Machines cluster significantly outperformed the fastest Cray shared memory multiprocessor. The prize has recognized every gain in parallelism from widely distributed workstations to the Sunway 10 million core processor in 2016. The overlap from the Seymour Cray recipe of a shared memory, multi-vector processor (aka Fortran Computer) to today's multicomputer turned out to show up and be incredibly fast. Hopefully, the prize has helped the rapid adoption of new paradigms, e.g. GPUs and possibly FPGAs. The prize also recognized the value of specialized hardware. Finally, the prize recognizes the tremendous effort and creativity required with algorithms, computational science, and computer science to exploit the increasingly parallelism afforded by the computers as seen in the increase in number of authors.

TOP500 - Past, Present, Future

Erich Strohmaier (Lawrence Berkeley National Laboratory), Jack Dongarra (University of Tennessee, Oak Ridge National Laboratory), Horst Simon (Lawrence Berkeley National Laboratory), Martin Meuer (Prometeus GmbH)

The TOP500 list (www.top500.org) has served as the defining yardstick for supercomputing performance since 1993. Published twice a year, it compiles the world's 500 largest installations and some of their main characteristics. Systems are ranked according to their performance of the Linpack benchmark, which solves a dense system of linear equations. Over time, the data collected for the list has enabled the early identification and quantification of many important technological and architectural trends related to high-performance computing. We briefly describe the project's origins, the principles guiding data collection, and what has made the list so successful during the 25 year long transition from giga- to tera- to petascale computing. We also examine the list's limitations. The TOP500's simplicity has invited many criticisms, and we consider several complementary or competing projects that have tried to address these concerns. Finally, we explore several emerging trends and reflect on the list's potential usefulness for guiding large-scale HPC into the exascale era.

Wednesday, November 15th

Room: Mile High Ballroom

10:30 am - 12:00 pm

Invited Talks 3

Harp-DAAL: A Next Generation Platform for High Performance Machine Learning on HPC-Cloud

Judy Qiu (Indiana University)

Scientific discovery via advances in simulation and data analytics is an ongoing national priority. A corresponding challenge is to sustain the research, development, and deployment of the high performance infrastructure needed to enable those discoveries. Early cloud data centers are evolving with new technologies to better support massive data analytics. Analysis of Big Data use cases identifies the need for HPC technologies in the Apache Big Data Stack (HPC-ABDS). Deep learning, using GPU clusters, is a clear example, but many Machine Learning algorithms also need iteration and HPC communication and optimizations.

Our research has concentrated on runtime and data management to support HPC-ABDS. This is illustrated by our open source software Harp, a plug-in for native Apache Hadoop, which has a convenient science interface, high performance communication, and can invoke Intel's Data Analytics Acceleration Library (DAAL). We tested this on both a complex Latent Dirichlet Allocation topic

model and on subgraph mining algorithms using Intel's Xeon and Xeon Phi architectures. Other tests show that Harp can run Kmeans, Graph Layout, and Multi-Dimensional Scaling algorithms with realistic application datasets over 4096 cores on the IU Big Red II Supercomputer while achieving linear speedup.

We are building a scalable parallel Machine Learning library that includes routines in Apache Mahout, MLlib, and others built in an NSF funded collaboration. This already has 20 (12 using DAAL) library members and is being tested while we add more functionality. Our results show that data-centric parallelism extends our understanding of distributed and parallel computation to further advancements in handling big model data and speed of convergence. This finding demonstrates the effectiveness of using HPC machines for Big Data problems. We will continue to collaborate with academia, industry, and national centers in exploring the computational capabilities and their applications.

Citius, Altius, Fortius!

Hans-Joachim Bungartz (Technical University Munich)

What makes HPC actually high-performing? Besides the (typically mentioned) large-scale computational problems and the resulting involvement of large-scale computing infrastructure, it is probably the pursuit of efficiency at all levels, in order to obtain the high performance desired. Here, both performance and efficiency can have completely different meanings: the classical complexity-driven one ($O(N \log N)$ being worse than $O(N)$); the node-level one (80% of the peak performance being considered as fabulous); the parallel / scalability one (going for a good exploitation of massively parallel systems); the energy-aware one (looking for "cool" algorithms to avoid the need for power stations); the communication-avoiding one (taking into account the relatively increasing communication cost – some say "Flops are free"); or an emphasis on time-to-solution (focusing on "when do I have the result?" – typically the crucial issue for those who want to get their problems solved, i.e. computed). While many of these points of view are rather mono-dimensional (e.g., 80% of the peak performance, definitely, is a nice technical result, but it just tells me that I use my weapon well – not at all whether I use the right weapon), the problem is multi-faceted, and it involves modeling, algorithmics, implementation, and other software issues. The resulting complexity implies that education has to be considered in a very fundamental way, too.

The talk will start with a brief general discussion of performance and efficiency in the HPC context, then have a look at ongoing activities on software for exascale computing (with a focus on DFG's respective Priority Program SPPEXA including its multi-national dimension) as well as HPC education, and conclude with a few examples from our own research.

Room: Mile High Ballroom
3:30 pm - 5:00 pm

Invited Talks 4

Artificial Intelligence and The Virtuous Cycle of Compute

Pradeep Dubey (Intel Corporation)

Traditionally, there has been a division of labor between computers and humans where all forms of number crunching and bit manipulation are left to computers, whereas intelligent decision-making is left to us humans. We are now at the cusp of a major transformation that can disrupt this balance. This disruption is triggered by an unprecedented convergence of massive compute with massive data, and some recent algorithmic advances. This confluence has the potential to spur a virtuous cycle of compute. It can significantly impact how we do computing and what computing can do for us. In this talk, I will discuss some of the application-level opportunities and system-level challenges that lie at the heart of this intersection of traditional high-performance computing with emerging data-intensive computing.

Inference and Control in Routing Games

Alexandre Bayen (Lawrence Berkeley National Laboratory; University of California, Berkeley)

This talk presents inference, control, and game-theoretic algorithms developed to improve traffic flow in transportation networks, implemented on HPC platforms. First, traffic estimation algorithms using crowdsourced mobile data are presented. These rely on applications of convex optimization to inverse modeling problems involving partial differential equations (PDEs). The implementation of these algorithms on mobile phones increased the accuracy of traffic information. Second, the talk presents algorithms to control transportation infrastructure assets (metering lights, traffic lights in the arterial networks, variable speed limits, etc.). These algorithms rely on adjoint-based optimization of PDEs in discretized form. Finally, we investigate disruptions in demand due to the rapid expansion of the use of "selfish routing" apps. These disruptions cause congestion and make traditional approaches of traffic management less effective. Game theoretic approaches to demand modeling are presented. These models encompass heterogeneous users (some using routing information, some not) that share the same network and compete for the same commodity (capacity). Results will be presented for static loading, based on Nash-Stackelberg games, and in the context of repeated games, to account for the fact that routing algorithms learn the dynamics of the system over time when users change their behavior. HPC implementations on the NERSC cluster at LBNL will be used to demonstrate the ability to scale up algorithms for the entire LA Basin or the City of Chicago, using a parallel version of the Frank-Wolfe algorithm.

Thursday, November 16th

Room: Mile High Ballroom
8:30 am - 10:00 am

Plenary Invited Talk

Lessons on Integrating and Utilizing 10 Million Cores: Experience of Sunway TaihuLight

Haohuan Fu (Tsinghua University)

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 will first introduce and discuss design philosophy about the approach to integrate these 10 million cores, at both the processor and the system level. Based on such a system design, we will then talk about the efforts and the challenges as we see in the process of utilizing the 10 million cores to push forward the frontiers of science in domains such as climate, seismology, material, bioinformatics, and big data analytics.

Molecular Simulation at the Mesoscale

Rommie Amaro (University of California, San Diego)

Advances in structural, chemical, and biophysical data acquisition (e.g., protein structures via X-ray crystallography and near atomic cryo-EM, isothermal calorimetry, etc.), coupled with the continued exponential growth in computing power and advances in the underlying algorithms now make the application of computational methods are opening a new era for the simulation of biological systems at the molecular level, and at scales never before reached. We are developing new capabilities for multi-scale dynamic simulations that cross spatial scales from the molecular (angstrom) to cellular ultrastructure (near micron), and temporal scales from the picoseconds of macromolecular dynamics to the physiologically important time scales of organelles and cells (milliseconds to seconds). Our efforts are driven by the outstanding and persistent advances in peta- and exa-scale computing and availability of multi-modal biological datasets, as well as by gaps in current abilities to connect across scales where it is already clear that new approaches will result in novel fundamental understanding of biological phenomena.

Room: Mile High Ballroom
10:30 am - 12:00 pm

Invited Talks 5

Cyber-Physical System and Industrial Applications of Large-Scale Graph Analysis and Optimization Problems

Katsuki Fujisawa (Kyushu University, National Institute of Advanced Industrial Science and Technology)

Abstract: In this talk, we present our ongoing research project. The objective of this project is to develop advanced computing and optimization infrastructures for extremely large-scale graphs on post peta-scale supercomputers. We explain our challenge to the Graph500 benchmark that are designed to measure the performance of a computer system for applications that require irregular memory and network access patterns. In 2014 to 2017, our project team has been a winner at the eighth, and 10th to 14th Graph500 benchmark.

We commenced our research project for developing the Urban OS (Operating System) for a large-scale city in 2013. The Urban OS, which is regarded as one of the emerging applications of the cyber-physical system (CPS), gathers big data sets of the distribution of people and transportation movements by utilizing sensor technologies and storing them in the cloud storage system. 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 briefly explain our ongoing research project for realizing the Urban OS.

Computing with Physics: Analog Computation and Neural Network Classification with a Dot Product Engine

Catherine Graves (Hewlett Packard)

General-purpose digital systems for computing have benefited from favorable scaling for decades, but are now hitting a wall in energy efficiency. There is consequently a growing interest in more efficient computing systems that may be specialized in function but worth the tradeoffs of lower energy consumption and higher performance in certain demanding applications such as artificial neural networks. In particular, several highly efficient architectures with mixed analog implementations utilizing emerging nonvolatile technologies such as memristors have recently been proposed to accelerate neural network computations.

In this talk, I will present our work implementing a prototype hardware accelerator dot product engine (DPE) for vector-matrix multiplication (VMM). VMM is a bottleneck for many applications, particularly in neural networks, and can be performed in the analog domain using Ohm's law for multiplication and Kirchoff's current law for summation. The DPE performs VMM in a single step by applying a vector of voltages to the rows and reading the currents on the columns of a memristor crossbar, which stores real-

valued matrix elements as device conductances. We have demonstrated high-precision analog tuning and control of memristor cells across a 128x64 crossbar array and evaluated the resulting VMM accuracy in our DPE prototype. We also performed single-layer neural network inference in the DPE for the 10k MNIST handwritten digit test patterns and assessed the performance comparison to a digital approach. Finally, I will discuss forecasted computational efficiency of scaled and integrated DPEs on chip (>100 TOPS/W), other computational applications of the DPE, and our work on generalized accelerator architectures based on DPE for broader algorithms and more complex functions.

Keynote & Plenary Talk [\(back to top\)](#)

Tuesday, November 14th

Room: Mile High Ballroom
8:30 am - 10:00 am

SC17 Keynote Address

Philip Diamond (Square Kilometre Array), Rosie Bolton (Square Kilometre Array)

Professor Diamond, accompanied by Dr. Rosie Bolton, SKA Regional Centre Project Scientist, will take SC17 attendees around the globe and out into the deepest reaches of the observable universe as they describe the SKA's international partnership that will map and study the entire sky in greater detail than ever before.

Navigating SC17 [\(back to top\)](#)

Monday, November 13th

Room: 705-707
4:45 pm - 5:30 pm

Navigating SC17

Navigating SC17

The SC17 Conference welcomes first-time attendees. We hope to attract new and diverse groups of HPC professionals and students to the conference each year with the goal of sparking new conversations, new connections, and new ideas. However, we understand that it can be difficult to navigate the conference and experience all that SC17 has to offer.

We are hosting sessions on Monday afternoon and Tuesday morning to help first time attendees navigate the SC17 conference.

Tuesday, November 14th

Room: 705-707
7:30 am - 8:15 am

Navigating SC17

Navigating SC17

The SC17 Conference welcomes first-time attendees. We hope to attract new and diverse groups of HPC professionals and students to the conference each year with the goal of sparking new conversations, new connections, and new ideas. However, we understand that it can be difficult to navigate the conference and experience all that SC17 has to offer.

We are hosting sessions on Monday afternoon and Tuesday morning to help first time attendees navigate the SC17 conference.

Panel [\(back to top\)](#)

Tuesday, November 14th

Room: 201-203

10:30 am - 12:00 pm

How Serious Are We About the Convergence Between HPC and Big Data?

Moderator: Hatem Ltaief (King Abdullah University of Science and Technology)

Panelist: Sadaf Alam (Swiss National Supercomputing Centre), Katie Antypas (National Energy Research Scientific Computing Center), Barry Bolding (Cray Inc), David Keyes (King Abdullah University of Science and Technology), Bastian Koller (High Performance Computing Center Stuttgart), Steve Oberlin (Nvidia Corporation), Mark Seager (Intel Corporation)

The possible convergence between the third and fourth paradigms confronts the scientific community with both a daunting challenge and a unique opportunity. The challenge resides in the requirement to support both heterogeneous workloads with the same hardware architecture. The opportunity lies in creating a common software stack to accommodate the requirements of scientific simulations and big data applications productively while maximizing performance and throughput.

With the hardware technology scaling and the sheer volume of scientific data to process, the scientific community faces fresh opportunities. Come, discuss and share with experts panelists how the community needs now to address the following questions moving forward:

Is the HPC / Big Data convergence a forced marriage or of mutual consent?

What are the current and future challenges / limitations to reach this convergence?

Does the HPC / Big Data convergence necessarily mean compromise and sacrifice in order to gain the expected synergisms?

Room: 201-203

1:30 pm - 3:00 pm

The ARM Software Ecosystem: Are We There Yet?

Moderator: Chris J. Newburn (Nvidia Corporation)

Panelist: David Abdurachmanov (CERN), Larry Kaplan (Cray Inc), Simon McIntosh-Smith (University of Bristol), Matt McLean (University of Michigan), Shinji Sumimoto (Fujitsu Ltd), Eric Van Hensbergen (ARM Ltd), Verónica Vergara Larrea (Oak Ridge National Laboratory)

It appears that ARM deployments are picking up steam in the HPC and Big Data spaces. But what are the conditions for real business impact with ARM in those spaces, and when will they be met? Is it time to jump in, or are the ingredients for success still coming together? This panel takes a critical look at system plans, existing and future deployments. It explores efforts to evaluate software ecosystem dependences and to promote application readiness. It highlights technical challenges and remaining opens in that space. Providers of infrastructure, practitioners and end users share experiences, aspirations and concerns about how the ARM SW ecosystem is ripening. The audience can expect to be educated about substantive progress, key gaps and who's working to close them, areas where there's a lack of consensus and where data is being gathered to come to greater agreement.

Room: 201-203

3:30 pm - 5:00 pm

Reproducibility and Uncertainty in High Performance Computing

Moderator: Victoria Stodden (University of Illinois)

Panelist: Bruce Childers (University of Pittsburgh), Matthew Krafczyk (National Center for Supercomputing Applications, University of Illinois), Miriam Leiser (Northeastern University), Michela Taufer (University of Delaware), Andreas Schreiber (German Aerospace Center)

This panel intends to discuss current and next steps for research on reproducibility in high performance computing. Reproducibility is a relatively new area of research and we propose discussing several novel focus areas including: • Understanding next steps for defining and implementing reproducibility standards, • Understanding the relationship between reproducibility and uncertainty quantification, • Understanding infrastructure needs for improved reproducibility, • Understanding challenges in all these approaches.

We plan to engage the audience in a robust discussion of these topics. Each of the panelists has expertise in at least one focus

area mentioned.

Wednesday, November 15th

Room: 201-203

10:30 am - 12:00 pm

Best Practices for Architecting Performance and Capacity in the Burst Buffer Era

Moderator: Addison Snell (*Intersect360 Research*)

Panelist: Pamela Hill (*National Center for Atmospheric Research*), Chris Zimmer (*Oak Ridge National Laboratory*), Dirk Pleiter (*Juelich Supercomputing Center*), Sadaf R. Alam (*Swiss National Supercomputing Centre*), James Coomer (*DataDirect Networks*)

As supercomputing sites prepare for exascale – or are creating plans to take their environments to the next scale of performance – Burst Buffers and new methods for innovative deployments of Flash are rapidly becoming an expectation, and in some cases a mandatory requirement in large-scale HPC procurements. How will this game-changing technology disrupt the way high-performance compute, file and storage systems are architected? Incorporating a Burst Buffer inherently changes how you select, procure and aggregate compute/storage/networking components in order to achieve performance and capacity goals. This cache-centric approach promises to eliminate the performance roadblocks of today's parallel file systems and dramatically lower costs by changing IO behavior and providing significant cost, power, space and cooling savings. This session will remove ambiguity around cache-centric approaches by exploring the burst buffer storage system design for use by large-scale HPC systems.

Room: 201-203

1:30 pm - 3:00 pm

Supercomputing in the Shadow of Giants: Perspectives and Insights from Supercomputing Leaders Outside the “Big 5” Regions and Organizations

Moderator: Jay Lofstead (*Sandia National Laboratories*)

Panelist: Mark Dietrich (*Compute Canada*), Antonio Tedu A. Gomes (*National Laboratory for Scientific Computing, Brazil*), Marek Michalewicz (*University of Warsaw*), Happy Sithole (*Center for High Performance Computing in South Africa*)

The Big 5 regions (China, Japan, America (DOE, XSEDE, and DOD), Europe (PRACE and EGI)) are well funded and relatively resource rich. Researchers in these regions can experiment with scales and diverse resources few others worldwide can readily use. Outside of these areas, significant high quality work is being done around the world. Internationally, creative leads have found ways to achieve high impact work both helping their regions as well as contributing to the global research community. In this discussion the panelists will share their experiences with the audience to help the audience members achieve similar goals.

Room: 201-203

3:30 pm - 5:00 pm

Blurring the Lines: High-End Computing and Data Science

Moderator: Sandy Landsberg (*US Department of Defense HPC Modernization Program*)

Panelist: Francine Berman (*Rensselaer Polytechnic Institute*), Steve Conway (*Hyperion Research*), Robert Grossman (*University of Chicago*), Satoshi Matsuoka (*Tokyo Institute of Technology*), Rick Stevens (*Argonne National Laboratory*), Michela Taufer (*University of Delaware*)

High-End Computing (HEC) encompasses both massive computational and big data capability to solve computational problems of significant importance that are beyond the capability of small- to medium-scale systems. Data science includes large-scale data analytics and visualization across multiple scales of data from a multitude of sources. Increasingly on-demand and real-time data intensive computing, enabling real-time analysis of simulations, data-intensive experiments and streaming observations, is pushing the boundaries of computing and resulting in a convergence of traditional HEC and newer cloud computing environments. This panel will explore challenges and opportunities at the intersection of high-end computing and data science. • Which markets will drive the adoption of HEC for Data Science? What new applications could arise from this convergence? What game-changers will this enable? • What impact will this have on conventional workflows, architectures and new memory paradigms (supercomputers versus shared cloud computing environments), software tools and workforce development?

Thursday, November 16th

Room: 201-203

10:30 am - 12:00 pm

Common Big Data Challenges in Bio, Geo, Climate, and Social Sciences

Moderator: Sushil Prasad (Georgia State University)

Panelist: Joel Saltz (Stony Brook University), Vipin Kumar (University of Minnesota), Chaitanya Baru (National Science Foundation), E. Lynn Utery (US Geological Survey), Daniel Reed (University of Iowa), Pete Beckman (Argonne National Laboratory), Sangram Ganguly (NASA Ames Research Center)

Spatiotemporal data, whether captured through remote sensors, ground and ocean sensors, social media and handhelds, traffic-related sensors and cameras, medical imaging, or large scale simulations have always been “big.” A common thread among all these big collections of datasets is that they are spatial and temporal. Processing and analyzing these datasets requires high-performance computing infrastructures. Despite these commonalities, leading big data communities of bio, geo, climate and social sciences, are highly fragmented and work in silos, resulting in solutions that are difficult to discover, integrate, and cross-fertilize. This panel aims to bring together the aforementioned, diverse yet overlapping communities with substantive big data and compute problems under SC umbrella to facilitate dialogue to reduce the impedances.

Panel Questions:

- HPC and Spatial-temporal Computing - two ships passing in the night? - Does HPC offer a mechanism to facilitate cross-fertilization? - Impedances to large-scale adoption of Spatial Computation and Analytics?

Room: 201-203

1:30 pm - 3:00 pm

Virtualization Ecosystems – Supporting Increasingly Complex Scientific Applications

Moderator: Kenton McHenry (National Center for Supercomputing Applications, University of Illinois)

Panelist: Mike Conway (Renaissance Computing Institute), Niall Gaffney (University of Texas), Christine Kirkpatrick (San Diego Supercomputer Center), Rob Kooper (University of Illinois), Mahidhar Tatineni (San Diego Supercomputer Center), Jaroslaw Nabrzyski (University of Notre Dame)

Computational Science has rapidly grown over the past decades due to adoption of “commodity” technologies starting with Beowulf clusters of the 90’s, through building large computational resources based on commercially available servers. This trend led the scientific community to adopt commercial software practices such as virtualization and provisioning systems. As a result scientific applications are becoming increasingly less tied to specific hardware, increasingly diverse, and increasingly complex with many interconnected components and dependencies. Addressing problems requiring data/model synthesis or data sharing and analysis as services, considerations in this space typically take the form of managing heterogeneous sub-components, differing and at times conflicting dependencies, as well as diverse and changing deployment considerations. In this panel we will discuss the growing movement to address these challenges: changes in the type of support provided, conventions that would facilitate such activities, and tools that can be leveraged to simplify these types of applications.

Room: 201-203

3:30 pm - 5:00 pm

Post Moore Supercomputing

Moderator: Jeffrey S. Vetter (Oak Ridge National Laboratory)

Panelist: George Michelogiannakis (Lawrence Berkeley National Laboratory), Murray Thom (D-Wave Systems Inc), Karen Bergman (Columbia University), Max Shulaker (Massachusetts Institute of Technology), Bill Harrod (US Department of Energy)

With the approaching end of MOSFET technology scaling, an array of emerging technologies promises to preserve digital computing performance. Continuing progress of supercomputing beyond the scaling limits of Moore’s Law is likely to require a comprehensive re-thinking of technologies, from innovative materials and devices, circuits, system architectures, programming systems, system software, and applications. Our goal is to explore how to prepare supercomputing for this “Post Moore’s Law” era. This panel brings together experts from many of those emerging technologies to discuss their potential impact over the next two decades. This panel will draw from the exciting results and speakers from the PMES. The goal of this panel is to share results from the PMES workshop as well as the panelist’s observations, in order to help the community set expectations for these diverse emerging technologies. Our panel discussion will include brief introductions by each panelist, followed by a guided discussion.

Friday, November 17th

Room: 201-203

8:30 am - 10:00 am

Silent Errors in HPC Systems

Moderator: Krishna Kant (Temple University)

Panelist: Devesh Tiwari (Northeastern University), Sriram Krishnamoorthy (Pacific Northwest National Laboratory), Mattan Erez (University of Texas), Taieb Znati (University of Pittsburgh), Dorian Arnold (Emory University)

This panel will explore silent errors in HPC applications that are expected to increase significantly as the semiconductor technology reaches its feature size limits. The panel will address the following questions:

Can we characterize silent errors in a way that they can be distinguished from other types of errors? Specifically, when do we say whether or not a program has suffered from silent errors?

What solid evidence is there that silent errors are happening in real systems, and how they are affecting the correctness or running time of programs?

How does the propagation of errors and their detectability depend on application characteristics? How does one restructure a program (including use of minimal verification code) to enforce bounds on error impact and propagation?

Is it possible to develop general compiler, middleware, and hardware techniques to detect or mask their impact without significant performance degradation and increase in cost and energy?

Room: 205-207

8:30 am - 10:00 am

HPC Software: Is “Cool Stuff” Really Incompatible with Sustainability?

Moderator: *Dairsie Latimer (Red Oak Consulting)*

Panelist: *Paul Selwood (Met Office, UK), Mike Dewar (Numerical Algorithms Group), Lorena Barba (George Washington University), Rebecca Hartman-Baker (Lawrence Berkeley National Laboratory), Guy Lonsdale (Scapos)*

The HPC community faces a software conundrum. On the one hand, we need to maintain a long-running code base supported by long-term language standards backed by reliable, world-class compilers. On the other hand, there is a need to exploit new hardware features, evolving language ideas and constructs to support algorithmic advances. The traditional languages are slow to react and we often end up in a “Language du Jour” situation.

The result is that many sites now have large swathes of, sometimes operational, code written in a variety of languages with variable compiler and vendor support, relying on difficult to find skills. Other sites rigidly enforce the use of the ‘house language’ and only allow gradual and established changes in language standards. This leads to its own problems.

This panel will explore the spectrum between these two extremes and seek to determine if there is a better way forward.

Room: 201-203

10:30 am - 12:00 pm

Energy Efficiency Gains From Software: Retrospectives and Perspectives

Moderator: *Daniel Reed (University of Iowa)*

Panelist: *Satoshi Matsuoka (Tokyo Institute of Technology), Thomas Schulthess (Swiss National Supercomputing Centre), John Shalf (Lawrence Berkeley National Laboratory), William Gropp (National Center for Supercomputing Applications, University of Illinois)*

We have already achieved major gains in energy-efficiency for both the datacenter and HPC equipment. For example, the PUE of the Swiss Supercomputer (CSCS) datacenter prior to 2012 was 1.8, but the current PUE is about 1.25; a factor of ~1.5 improvement. HPC system improvements have also been very strong, as evidenced by FLOPS/Watt performance on the Green500 List. While we have seen gains from data center and HPC system efficiency, there are also energy-efficiency gains to be had from software- application performance improvements, for example. This panel will explore what HPC software capabilities were most helpful over the past years in improving HPC system energy efficiency? It will then look forward; asking in what layers of the software stack should a priority be put on introducing energy-awareness; e.g., runtime, scheduling, applications? What is needed moving forward? Who is responsible for that forward momentum?

Paper [\(back to top\)](#)

Tuesday, November 14th

Room: 301-302-303
10:30 am - 12:00 pm

HPC Application Development Tools

Egeria: A Framework for Auto-Construction of HPC Advising Tools through Multi-Layered Natural Language Processing

Authors: Hui Guan (North Carolina State University), Xipeng Shen (North Carolina State University), Hamid Krim (North Carolina State University)

Vendors often provide some detailed programming guides to assist programmers in developing high performance programs. However, these guides are frequently hundreds of pages long, making it difficult for general programmers to master and memorize all the rules and guidelines and properly apply them to a specific problem instance.

In this work, we develop a framework named Egeria to alleviate the difficulty. Through Egeria, one can easily construct an advising tool for a certain high performance computing (HPC) domain (e.g., GPU programming). Egeria is made possible through a distinctive multi-layered design that leverages the properties of HPC domains and overcomes the weaknesses of existing Natural Language Processing (NLP) techniques. Experiments on CUDA and OpenCL programming demonstrate the usefulness of Egeria for HPC both qualitatively and quantitatively.

DataRaceBench: A Benchmark Suite for Systematic Evaluation of Data Race Detection Tools

Authors: Chunhua Liao (Lawrence Livermore National Laboratory), Pei-Hung Lin (Lawrence Livermore National Laboratory), Joshua Asplund (Lawrence Livermore National Laboratory), Markus Schordan (Lawrence Livermore National Laboratory), Ian Karlin (Lawrence Livermore National Laboratory)

Data races in multi-threaded parallel applications are notoriously damaging while extremely difficult to detect. Many tools have been developed to help programmers find data races. However, there is no dedicated OpenMP benchmark suite to systematically evaluate data race detection tools for their strengths and limitations.

We present DataRaceBench, an open-source benchmark suite designed to systematically and quantitatively evaluate the effectiveness of data race detection tools. We focus on data race detection in programs written in OpenMP, the popular parallel programming model for multi-threaded applications. In particular, DataRaceBench includes a set of microbenchmark programs with or without data races.

We also define several metrics to represent effectiveness and efficiency of data race detection tools. We evaluate four different data race detection tools: Helgrind, ThreadSanitizer, Archer, and Intel Inspector. The results show that DataRaceBench is effective to provide comparable, quantitative results and discover strengths and weaknesses of the tools evaluated.

BP

Optimizing the Query Performance of Block Index Through Data Analysis and I/O Modeling

Authors: Tzuhsien Wu (National Tsing Hua University, Taiwan), Jerry Chou (National Tsing Hua University, Taiwan), Hao Shyng (National Tsing Hua University, Taiwan), Bin Dong (Lawrence Berkeley National Laboratory), Scott Klasky (University of Tennessee), Kesheng Wu (Lawrence Berkeley National Laboratory)

Indexing technique has become an efficient tool to enable scientists to directly access the most relevant data records. But, the time and space requirements of building and storing indexes are expensive in the traditional approaches, such as R-tree and bitmaps.

Recently, we started to address this issue by proposing the idea of "block index", and our previous work has shown promising results from comparing it against other well-known solutions, including ADIOS, SciDB, and FastBit. In this work, we further improve the technique from both theoretical and implementation perspectives. Driven by an extensive effort in characterizing scientific datasets and modeling I/O systems, we presented a theoretical model to analyze its query performance with respect to a given block size configuration. We also introduce three optimization techniques to achieve a 2.3x query time reduction compared to the original implementation.

Room: 402-403-404
10:30 am - 12:00 pm

Deep Learning

Deep Learning at 15PF: Supervised and Semi-Supervised Classification for Scientific Data

Authors: Thorsten Kurth (Lawrence Berkeley National Laboratory), Jian Zhang (Stanford University), Nadathur Satish (Intel Corporation), Evan Racah (Lawrence Berkeley National Laboratory), Ioannis Mitliagkas (Stanford University), Md Mostofa Patwary (Intel Corporation), Tareq Malas (Lawrence Berkeley National Laboratory), Narayanan Sundaram (Intel Corporation), Wahid Bhimji (Lawrence Berkeley National Laboratory), Mikhail E. Smorkalov (Intel Corporation), Jack Deslippe (Lawrence Berkeley National Laboratory), Mikhail Shiryayev (Intel Corporation), Srinivas Sridharan (Intel Corporation), Mr Prabhat (Lawrence Berkeley National Laboratory), Pradeep Dubey (Intel Corporation)

This paper presents the first, 15-PetaFLOP Deep Learning system for solving scientific pattern classification problems on contemporary HPC architectures. We develop supervised convolutional architectures for discriminating signals in high-energy physics data as well as semi-supervised architectures for localizing and classifying extreme weather in climate data. Our Intelcaffe-based implementation obtains ~2TFLOP/s on a single Cori Phase-II Xeon-Phi node. We use a hybrid strategy employing synchronous node-groups, while using asynchronous communication across groups. We use this strategy to scale training of a single model to ~9600 Xeon-Phi nodes; obtaining peak performance of 11.73-15.07 PFLOP/s and sustained performance of 11.41-13.27 PFLOP/s. At scale, our HEP architecture produces state-of-the-art classification accuracy on a dataset with 10M images, exceeding that achieved by selections on high-level physics-motivated features. Our semi-supervised architecture successfully extracts weather patterns in a 15TB climate dataset. Our results demonstrate that Deep Learning can be optimized and scaled effectively on many-core, HPC systems.

Understanding Error Propagation in Deep Learning Neural Network (DNN) Accelerators and Applications

Authors: Guanpeng Li (University of British Columbia), Siva Hari (Nvidia Corporation), Michael Sullivan (Nvidia Corporation), Timothy Tsai (Nvidia Corporation), Karthik Pattabiraman (University of British Columbia), Joel Emer (Nvidia Corporation), Stephen Keckler (Nvidia Corporation)

Specialized hardware accelerators have been proposed to accelerate the execution of DNN algorithms for high-performance and energy efficiency. Recently, they have been deployed in datacenters (potentially for business-critical or industrial applications) and safety-critical systems such as self-driving cars. Soft errors caused by high-energy particles have been increasing in hardware systems, and these can lead to catastrophic failures in DNN systems. Traditional methods for building resilient systems, e.g., Triple Modular Redundancy (TMR), are agnostic of DNN. Hence, these approaches incur high overheads, which makes them challenging to deploy. In this paper, we experimentally evaluate the resilience characteristics of DNN systems (i.e., DNN software running on specialized accelerators). We find that the error resilience of a DNN system depends on the data types, values, data reuses, and the types of layers in the design. Based on our observations, we propose two efficient protection techniques for DNN systems.

Scaling Deep Learning on GPU and Knights Landing Clusters

Authors: Yang You (University of California, Berkeley), Aydin Buluc (Lawrence Berkeley National Laboratory; University of California, Berkeley), James Demmel (University of California, Berkeley)

Training neural networks has become a big bottleneck. For example, training ImageNet dataset on one Nvidia K20 GPU needs 21 days. To speed up the training process, the current deep learning systems heavily rely on the hardware accelerators. However, these accelerators have limited on-chip memory compared with CPUs.

We use both self-host Intel Knights Landing (KNL) clusters and multi-GPU clusters as our target platforms. From the algorithm aspect, we focus on Elastic Averaging SGD (EASGD) to design algorithms for HPC clusters.

We redesign four efficient algorithms for HPC systems to improve EASGD's poor scaling on clusters. Async EASGD, Async MEASGD, and Hogwild EASGD are faster than existing counterpart methods (i.e. Async SGD, Async MSGD, and Hogwild SGD) in all comparisons. Our Sync EASGD achieves 5.3X speedup over original EASGD on the same platform. We achieve 91.5% weak scaling efficiency on 4253 KNL cores, which is higher than the state-of-the-art implementation.

Room: 405-406-407
10:30 am - 12:00 pm

Cutting Edge File Systems

LocoFS: A Loosely-Coupled Metadata Service for Distributed File Systems

Authors: Siyang Li (Tsinghua University), Youyou Lu (Tsinghua University), Jiwu Shu (Tsinghua University), Yang Hu (University of Texas, Dallas), Tao Li (University of Florida)

Key-Value stores provide scalable metadata service for distributed file systems. However, the metadata's organization itself, which is organized using a directory tree structure, does not fit the key-value access pattern, thereby limiting the performance. To address this issue, we propose a distributed file system with a loosely-coupled metadata service, LocoFS, to bridge the performance gap between file system metadata and key-value stores. LocoFS is designed to decouple the dependencies between different kinds of metadata with two techniques. First, LocoFS decouples the directory content and structure, which organizes file and directory index nodes in a flat space while reversely indexing the directory entries. Second, it decouples the file metadata to further improve the key-value access performance. Evaluations show that LocoFS with eight nodes boosts the metadata throughput by 5 times, which approaches 93% throughput of a single-node key-value store, compared to 18% in the state-of-the-art IndexFS.

TagIt: An Integrated Indexing and Search Service for File Systems

Authors: Hyogi Sim (Virginia Tech, Oak Ridge National Laboratory), Youngjae Kim (Sogang University), Sudharshan S. Vazhkudai (Oak Ridge National Laboratory), Geoffroy R. Vallee (Oak Ridge National Laboratory), Seung-Hwan Lim (Oak Ridge National Laboratory), Ali R. Butt (Virginia Tech)

Data services such as search, discovery, and management in scalable distributed environments have traditionally been decoupled from the underlying file systems, and are often deployed using external databases and indexing services. However, modern data production rates, looming data movement costs, and the lack of metadata, entail revisiting the decoupled file system-data services design philosophy.

In this paper, we present TagIt, a scalable data management service framework aimed at scientific datasets, which is tightly integrated into a shared-nothing distributed file system. A key feature of TagIt is a scalable, distributed metadata indexing framework, using which we implement a flexible tagging capability to support data discovery. The tags can also be associated with an active operator, for pre-processing, filtering, or automatic metadata extraction, which we seamlessly offload to file servers in a load-aware fashion. Our evaluation shows that TagIt can expedite data search by up to 10x over the extant decoupled approach.

A Configurable Rule-Based Classful Token Bucket Filter Network Request Scheduler for the Lustre File System

Authors: Yingjin Qian (DataDirect Networks), Xi Li (DataDirect Networks), Shuichi Ihara (DataDirect Networks), Lingfang Zeng (Johannes Gutenberg University Mainz), Jürgen Kaiser (Johannes Gutenberg University Mainz), Tim Süß (Johannes Gutenberg University Mainz), André Brinkmann (Johannes Gutenberg University Mainz)

HPC file systems today work in a best-effort manner where individual applications can flood the file system with requests, effectively leading to a denial of service for all other tasks. This paper presents a Token Bucket Filter (TBF) policy for the Lustre file system. The TBF enforces RPC rate limitations based on (potentially complex) Quality of Service (QoS) rules. The QoS rules are enforced in Lustre's Object Storage Servers, where each request is assigned to an automatically created QoS class.

The proposed QoS implementation for Lustre enables various features for each class including the support for high-priority and real-time requests even under heavy load and the utilization of spare bandwidth by less important tasks under light load. The framework also enables dependent rules to change a job's RPC even at very small timescales. Furthermore, we propose a Global Rate Limiting (GRL) algorithm to enforce system-wide RPC rate limitations.

Room: 301-302-303

3:30 pm - 5:00 pm

Multiphysics

A Framework for Scalable Biophysics-Based Image Analysis

Authors: Amir Gholami (University of Texas), Andreas Mang (University of Texas), Klaudius Scheufele (University of Stuttgart), Christos Davatzikos (University of Pennsylvania), Miriam Mehl (University of Stuttgart), George Biros (University of Texas)

We present SIBIA (Scalable Integrated Biophysics-based Image Analysis), a framework for coupling biophysical models with medical image analysis. It provides solvers for an image-driven inverse brain tumor growth model and an image registration problem, the combination of which can eventually help in diagnosis and prognosis of brain tumors. The two main computational kernels of SIBIA are a Fast Fourier Transformation (FFT) implemented in the library AccFFT to discretize differential operators, and a cubic interpolation kernel for semi-Lagrangian based advection. We present efficiency and scalability results for the computational kernels, the inverse tumor solver and image registration on two x86 systems. We showcase results that demonstrate that our solver can be used to solve registration problems of unprecedented scale, 4096^3 resulting in 200 billion unknowns---a problem size that is 64x larger than the state-of-the-art. For problem sizes of clinical interest, SIBIA is about 8x faster than the state-of-the-art.

BSP

Galactos: Computing the 3-pt Anisotropic Correlation for 2 Billion Galaxies

Authors: Brian Friesen (Lawrence Berkeley National Laboratory), Md. Mostofa Ali Patwary (Intel Corporation), Brian Austin (Lawrence Berkeley National Laboratory), Nadathur Satish (Intel Corporation), Zachary Slepian (Lawrence Berkeley National Laboratory), Narayanan Sundaram (Intel Corporation), Deborah Bard (Lawrence Berkeley National Laboratory), Daniel Eisenstein (Harvard University), Jack Deslippe (Lawrence Berkeley National Laboratory), Pradeep Dubey (Intel Corporation), Mr Prabhat (Lawrence Berkeley National Laboratory)

The nature of dark energy and the complete theory of gravity are two central questions currently facing cosmology. A vital tool for addressing them is the 3-point correlation function (3PCF), which probes deviations from a spatially random distribution of galaxies. However, the 3PCF's formidable computational expense has prevented its application to astronomical surveys comprising millions to billions of galaxies. We present Galactos, a high-performance implementation of a novel, $O(N^2)$ algorithm that uses a load-balanced k-d tree and spherical harmonic expansions to compute the anisotropic 3PCF. Our implementation is optimized for the Intel Xeon Phi architecture, exploiting SIMD parallelism, instruction and thread concurrency, and significant L1 and L2 cache reuse, reaching 39% of peak performance on a single node. Galactos scales to the full Cori system, achieving 9.8 PFLOPS (peak) across 9636 nodes, making the 3PCF easily computable for all galaxies in the observable universe.

Extreme Scale Multi-Physics Simulations of the Tsunamigenic 2004 Sumatra Megathrust Earthquake

Authors: Carsten Uphoff (Technical University Munich), Sebastian Rettenberger (Technical University Munich), Michael Bader (Technical University Munich), Stephanie Wollherr (Ludwig Maximilian University of Munich), Thomas Ulrich (Ludwig Maximilian University of Munich), Elizabeth H. Madden (Ludwig Maximilian University of Munich), Alice-Agnes Gabriel (Ludwig Maximilian University of Munich)

We present a high-resolution simulation of the 2004 Sumatra-Andaman earthquake, including non-linear frictional failure on a megathrust-splay fault system. Our method exploits unstructured meshes capturing the complicated geometries in subduction zones that are crucial to understand large earthquakes and tsunami generation. These up-to-date largest and longest dynamic rupture simulations enable analysis of dynamic source effects on the seafloor displacements.

To tackle the extreme size of this scenario an end-to-end optimization of the simulation code SeisSol was necessary. We implemented a new cache-aware wave propagation scheme and optimized the dynamic rupture kernels using code generation. We established a novel clustered local-time-stepping scheme for dynamic rupture. In total, we achieved a speed-up of 13.6 compared to the previous implementation. For the Sumatra scenario with 221 million elements this reduced the time-to-solution to 13.9 hours on 86,016 Haswell cores. Furthermore, we used asynchronous output to overlap I/O and compute time.

BP

Room: 402-403-404

3:30 pm - 5:00 pm

Compilation Techniques

Sympiler: Transforming Sparse Matrix Codes by Decoupling Symbolic Analysis

Authors: Kazem Cheshmi (Rutgers University), Shoaib Kamil (Adobe Research), Michelle Mills Strout (University of Arizona), Maryam Mehri Dehnavi (Rutgers University)

Sympiler is a domain-specific code generator that optimizes sparse matrix computations by decoupling the symbolic analysis phase from the numerical manipulation stage in sparse codes. The computation patterns in sparse numerical methods are guided by the input sparsity structure and the sparse algorithm itself. In many real-world simulations, the sparsity pattern changes little or not at all. Sympiler takes advantage of these properties to symbolically analyze sparse codes at compile-time and to apply inspector-guided transformations that enable applying low-level transformations to sparse codes. As a result, the Sympiler-generated code outperforms highly-optimized matrix factorization codes from commonly-used specialized libraries, obtaining average speedups over Eigen and CHOLMOD of 3.8× and 1.5× respectively.

Control Replication: Compiling Implicit Parallelism to Efficient SPMD with Logical Regions

Authors: Elliott Slaughter (Stanford University, SLAC National Accelerator Laboratory), Wonchan Lee (Stanford University), Sean Treichler (Stanford University, Nvidia Corporation), Wen Zhang (Stanford University), Michael Bauer (Nvidia Corporation), Galen Shipman (Los Alamos National Laboratory), Patrick McCormick (Los Alamos National Laboratory), Alex Aiken (Stanford University)

We present control replication, a technique for generating high-performance and scalable SPMD code from implicitly parallel programs. In contrast to traditional parallel programming models that require the programmer to explicitly manage threads and the communication and synchronization between them, implicitly parallel programs have sequential execution semantics and by their nature avoid the pitfalls of explicitly parallel programming. However, without optimizations to distribute control overhead, scalability is often poor.

Performance on distributed-memory machines is especially sensitive to communication and synchronization in the program. Thus optimizations for these machines require an intimate understanding of a program's memory accesses. Control replication achieves particularly effective and predictable results by leveraging language support for first-class data partitioning in the source programming model. We evaluate an implementation of control replication for Regent and show that it achieves up to 99% parallel efficiency at 1024 nodes with absolute performance comparable to hand-written MPI(+X) codes.

Optimizing Geometric Multigrid Method Computation Using a DSL Approach

Authors: Vinay Vasista (Indian Institute of Science), Kumudha Narasimhan (Indian Institute of Science), Siddharth Bhat (International Institute of Information Technology, Hyderabad), Uday Bondhugula (Indian Institute of Science)

The Geometric Multigrid (GMG) method is widely used in numerical analysis to accelerate the convergence of partial differential equations solvers using a hierarchy of grid discretizations. Multiple grid sizes and recursive expression of multigrid cycles make the task of program optimization tedious. A high-level language that aids domain experts for GMG and with good optimization support is thus valuable.

We demonstrate how high performance can be achieved along with enhanced programmability for GMG, with new language/optimization support in the PolyMage DSL framework. We compare our approach with (a) hand-optimized code, (b) hand-optimized code optimized in conjunction with polyhedral techniques, and (c) the existing PolyMage optimizer adapted to multigrid. We use benchmarks varying in multigrid cycle structure and smoothing steps for evaluation. On a 24-core Intel Xeon Haswell multicore system, our automatically optimized codes achieve a mean improvement of 3.2x over straightforward parallelization, and

1.31x over the PolyMage optimizer.

Room: 405-406-407
3:30 pm - 5:00 pm

Distributed Computing and Clouds

Efficient Process Mapping in Geo-Distributed Cloud Data Centers

Authors: Amelie Chi Zhou (Shenzhen University), Yifan Gong (TuSimple), Bingsheng He (National University of Singapore), Jidong Zhai (Tsinghua University)

Recently, various applications including data analytics and machine learning have been developed for geo-distributed cloud data centers. For those applications, the ways to map parallel processes to physical nodes (i.e., "process mapping") could significantly impact the performance of the applications because of non-uniform communication cost in such geo-distributed environments. While process mapping has been widely studied in grid/cluster environments, few of the existing studies have considered the problem in geo-distributed cloud environments. In this paper, we propose a novel model to formulate the geo-distributed process mapping problem and develop a new method to efficiently find the near optimal solution. Our algorithm considers both the network communication performance of geo-distributed data centers as well as the communication matrix of the target application. Evaluation results with real experiments on Amazon EC2 and simulations demonstrate that our proposal achieves significant performance improvement (50% on average) compared to the state-of-the-art algorithms.

Topology-Aware GPU Scheduling for Learning Workloads in Cloud Environments

Authors: Marcelo Carneiro do Amaral (Barcelona Supercomputing Center), Jorda Polo (Barcelona Supercomputing Center), David Carrera (Barcelona Supercomputing Center), Seetharami Seelam (IBM), Malgorzata Steinder (IBM)

Recent advances in hardware, such as systems with multiple GPUs and their availability in the cloud, are enabling deep learning in various domains including health care, autonomous vehicles, and Internet of Things. Multi-GPU systems exhibit complex connectivity among GPUs and between GPUs and CPUs. Workload schedulers must consider hardware topology and workload communication requirements in order to allocate CPU and GPU resources for optimal execution time and improved utilization in shared cloud environments.

This paper presents a new topology-aware workload placement strategy to schedule deep learning jobs on multi-GPU systems. The placement strategy is evaluated with a prototype on a Power8 machine with Tesla P100 cards, showing speedups of up to $\approx 1.30\times$ compared to state-of-the-art strategies; the proposed algorithm achieves this result by allocating GPUs that satisfy workload requirements while preventing interference. Additionally, a large-scale simulation shows that the proposed strategy provides higher resource utilization and performance in cloud systems.

Probabilistic Guarantees of Execution Duration for Amazon Spot Instances

Authors: Rich Wolski (University of California, Santa Barbara), John Brevik (California State University, Long Beach), Ryan Chard (Argonne National Laboratory), Kyle Chard (University of Chicago)

In this paper we propose DrAFTS -- a methodology for implementing probabilistic guarantees of instance reliability in the Amazon Spot tier. Amazon offers "unreliable" virtual machine instances (ones that may be terminated at any time) at a potentially large discount relative to "reliable" On-demand and Reserved instances. Our method predicts the "bid values" that users can specify to provision Spot instances which ensure at least a fixed duration of execution with a given probability. We illustrate the method and test its validity using Spot pricing data post facto, both randomly and using real-world workload traces. We also test the efficacy of the method experimentally by using it to launch Spot instances and then observing the instance termination rate. Our results indicate that it is possible to obtain the same level of reliability from unreliable instances that the Amazon service level agreement guarantees for reliable instances with a greatly reduced cost.

Wednesday, November 15th

Room: 301-302-303
10:30 am - 12:00 pm

Hierarchical Memory Usage

Understanding Object-Level Memory Access Patterns Across the Spectrum

Authors: Xu Ji (Tsinghua University, Qatar Computing Research Institute), Chao Wang (Oak Ridge National Laboratory), Nosayba El-Sayed (Massachusetts Institute of Technology, Qatar Computing Research Institute), Xiaosong Ma (Qatar Computing Research Institute), Youngjae Kim (Sogang University), Sudharshan S. Vazhkudai (Oak Ridge National Laboratory), Wei Xue (Tsinghua University), Daniel Sanchez (Massachusetts Institute of Technology)

Memory accesses limit the performance and scalability of countless applications. Many design and optimization efforts would benefit from an in-depth understanding of memory access behavior, which existing access tracing and profiling methods do not offer.

In this paper, we adopt a holistic memory access profiling approach to enable a better understanding of program-system memory interactions. We developed a two-pass tool adopting fast online and slow offline profiling, with which we profiled at the variable/object level a collection of 38 representative applications spanning major domains (HPC, personal computing, data analytics, AI, graph processing, and datacenter workloads), at varying problem sizes. We performed detailed result analysis and code examination. Our findings provide new insights into application memory behavior, including insights on per-object access patterns, adoption of data structures, and memory-access changes at different problem-sizes. We find that scientific computation applications exhibit distinct behaviors compared to datacenter workloads, motivating separate memory system design/optimizations.

Exploring and Analyzing the Real Impact of Modern On-Package Memory on HPC Scientific Kernels

Authors: Ang Li (Pacific Northwest National Laboratory), Weifeng Liu (University of Copenhagen, Norwegian University of Science and Technology), Mads R. B. Kristensen (University of Copenhagen), Brian Vinter (University of Copenhagen), Hao Wang (Virginia Tech), Kaixi Hou (Virginia Tech), Andres Marquez (Pacific Northwest National Laboratory), Shuaiwen Leon Song (Pacific Northwest National Laboratory)

High-bandwidth On-Package Memory (OPM) innovates the conventional memory hierarchy by augmenting a new on-package layer between the classic on-chip cache and off-chip DRAM. Despite the adaptation since Haswell, the performance and power impact of OPM on HPC scientific applications, especially essential scientific kernels, is still unknown.

In this paper, we fill this gap by conducting a comprehensive evaluation for a wide spectrum of major scientific kernels with a large amount of representative inputs, including dense, sparse and medium, on the two types of Intel OPM: eDRAM on multicore Broadwell architecture and MCDRAM on manycore Knights Landing architecture. 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.

BP

Large-Scale Adaptive Mesh Simulations Through Non-Volatile Byte-Addressable Memory

Authors: Bao Nguyen (Washington State University, Vancouver), Hua Tan (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

Octree-based mesh adaptation has enabled simulations of complex physical phenomena. Existing meshing algorithms were proposed with the assumption that computer memory is volatile. Consequently, for failure recovery, in-core algorithms need to save memory states with slow file I/Os. Out-of-core algorithms store octants on disks for persistence. However, neither of them was designed to leverage unique characteristics of non-volatile byte-addressable memory (NVBM). In this paper, we propose a novel data structure Persistent Merged octree (PM-octree) for both meshing and in-memory storage of persistent octrees using NVBM. It is a multi-version structure and can recover from failures using its earlier version stored in NVBM. In addition, we design a feature-directed sampling approach to help dynamically transform PM-octree for reducing NVBM-induced memory write latency. PM-octree has been successfully integrated with Gerris software for simulation of fluid dynamics. Our results demonstrate that PM-octree scales up to 2.1 billion elements octants on a Cray XT3.

Room: 402-403-404
10:30 am - 12:00 pm

GPUs and Communication

GPU Triggered Networking for Intra-Kernel Communications

Authors: Michael LeBeane (University of Texas, Advanced Micro Devices Inc), Khaled Hamidouche (Advanced Micro Devices Inc), Brad Benton (Advanced Micro Devices Inc), Mauricio Breternitz (University Institute of Lisbon), Steven K. Reinhardt (Microsoft), Lizy K. John (University of Texas)

GPUs are widespread across clusters of compute nodes due to their attractive performance for data parallel codes. However, communicating between GPUs across the cluster is cumbersome when compared to CPU networking implementations. A number of recent works have enabled GPUs to more naturally access the network, but suffer from performance problems, require hidden CPU helper threads, or restrict communications to kernel boundaries.

In this paper, we propose GPU Triggered Networking, a novel, GPU-centric networking approach which leverages the best of CPUs and GPUs. In this model, CPUs create and stage network messages and GPUs trigger the network interface when data is ready to send. GPU Triggered Networking decouples these two operations, thereby removing the CPU from the critical path. We illustrate how this approach can provide up to 25% speedup compared to standard GPU networking across microbenchmarks, a Jacobi stencil, an important MPI collective operation, and machine-learning workloads.

Gravel: Fine-Grain GPU-Initiated Network Messages

Authors: Marc S. Orr (University of Wisconsin), Shuai Che (Advanced Micro Devices Inc), Bradford M. Beckmann (Advanced Micro

Devices Inc), Mark Oskin (Advanced Micro Devices Inc, University of Washington), Steven K. Reinhardt (Microsoft), David A. Wood (University of Wisconsin, Advanced Micro Devices Inc)

Distributed systems incorporate GPUs because they provide massive parallelism in an energy-efficient manner. Unfortunately, existing programming models make it difficult to route a GPU-initiated network message. The traditional coprocessor model forces programmers to manually route messages through the host CPU. Other models allow GPU-initiated communication, but are inefficient for small messages.

To enable fine-grain PGAS-style communication between threads executing on different GPUs, we introduce Gravel. GPU-initiated messages are offloaded through a GPU-efficient concurrent queue to an aggregator (implemented with CPU threads), which combines messages targeting to the same destination. Gravel leverages diverged work-group-level semantics to amortize shared-memory synchronization across the GPU's data-parallel lanes.

Using Gravel, we can distribute six applications, each with frequent small messages, across a cluster of eight GPU-accelerated nodes. Compared to one node, these applications run 5.3x faster, on average. Furthermore, we show that Gravel is more programmable and usually performs better than prior GPU networking models.

Toward Standardized Near-Data Processing with Unrestricted Data Placement for GPUs

Authors: *Gwangsun Kim (ARM Ltd), Niladrish Chatterjee (Nvidia Corporation), Mike O'Connor (Nvidia Corporation), Kevin Hsieh (Carnegie Mellon University)*

3D-stacked memory devices with processing logic can help alleviate the memory bandwidth bottleneck in GPUs. However, in order for such Near-Data Processing (NDP) memory stacks to be used for different GPU architectures, it is desirable to standardize the NDP architecture. Our proposal enables this standardization by allowing data to be spread across multiple memory stacks as is the norm in high-performance systems without an MMU on the NDP stack. The keys to this architecture are the ability to move data between memory stacks as required for computation, and a partitioned execution mechanism that offloads memory-intensive application segments onto the NDP stack and decouples address translation from DRAM accesses. By enhancing this system with a smart offload selection mechanism that is cognizant of the compute capability of the NDP and cache locality on the host processor, system performance and energy are improved by up to 66.8% and 37.6%, respectively.

Room: 405-406-407
10:30 am - 12:00 pm

Reliability, Fault Tolerance, and Resilience

Experimental and Analytical Study of Xeon Phi Reliability

Authors: *Daniel Oliveira (Federal University of Rio Grande do Sul), Laercio Pilla (Federal University of Santa Catarina), Nathan DeBardeleben (Los Alamos National Laboratory), Sean Blanchard (Los Alamos National Laboratory), Heather Quinn (Los Alamos National Laboratory), Israel Koren (University of Massachusetts), Philippe Navaux (Federal University of Rio Grande do Sul), Paolo Rech (Federal University of Rio Grande do Sul)*

We present an in-depth analysis of transient faults effects on HPC applications in Intel Xeon Phi processors based on radiation experiments and high-level fault injection. Besides measuring the realistic error rates of Xeon Phi, we quantify Silent Data Corruption (SDCs) by correlating the distribution of corrupted elements in the output to the application's characteristics. We evaluate the benefits of imprecise computing for reducing the programs' error rate. For example, for HotSpot a 0.5% tolerance in the output value reduces the error rate by 85%.

We inject different fault models to analyze the sensitivity of given applications. We show that portions of applications can be graded by different criticalities. For example, faults occurring in the middle of LUD execution, or in the Sort and Tree portions of CLAMR, are more critical than the remaining portions. Mitigation techniques can then be relaxed or hardened based on the criticality of the particular portions.

REFINE: Realistic Fault Injection via Compiler-Based Instrumentation for Accuracy, Portability and Speed

Authors: *Giorgis Georgakoudis (Queen's University Belfast), Ignacio Laguna (Lawrence Livermore National Laboratory), Dimitrios S. Nikolopoulos (Queen's University Belfast), Martin Schulz (Lawrence Livermore National Laboratory, Technical University Munich)*

Compiler-based fault injection (FI) has become a popular technique for resilience studies to understand the impact of soft errors in supercomputing systems. Compiler-based FI frameworks inject faults at a high intermediate-representation level. However, they are less accurate than machine code, binary-level FI because they lack access to all dynamic instructions, thus they fail to mimic certain fault manifestations. In this paper, we study the limitations of current practices in compiler-based FI and how they impact the interpretation of results in resilience studies.

We propose REFINE, a novel framework that addresses these limitations, performing FI in a compiler backend. Our approach provides the portability and efficiency of compiler-based FI while keeping accuracy comparable to binary-level FI methods. We demonstrate our approach in 14 HPC benchmarks and show that, due to our unique design, its runtime overhead is significantly smaller than state-of-the-art compiler-based FI frameworks, reducing the time for large FI experiments.

Correcting Soft Errors Online in Fast Fourier Transform

Authors: Xin Liang (University of California, Riverside), Jieyang Chen (University of California, Riverside), Dingwen Tao (University of California, Riverside), Sihuan Li (University of California, Riverside), Panruo Wu (University of California, Riverside), Hongbo Li (University of California, Riverside), Kaiming Ouyang (University of California, Riverside), Yuanlai Liu (University of California, Riverside), Fengguang Song (Indiana University-Purdue University Indianapolis), Zizhong Chen (University of California, Riverside)

While many algorithm-based fault tolerance (ABFT) schemes have been proposed to detect soft errors offline in the fast Fourier transform (FFT) after computation finishes, none of the existing ABFT schemes detect soft errors online before the computation finishes. This paper presents an online ABFT scheme for FFT so that the corrupted computation can be terminated in a much more timely manner. We also extend our scheme to tolerate both arithmetic errors and memory errors, develop strategies to reduce its fault tolerance overhead and improve its numerical stability and fault coverage, and finally incorporate it into the widely used FFTW library - one of the today's fastest FFT software implementations. Experimental results demonstrate that: (1) the proposed online ABFT scheme introduces much lower overhead than the existing schemes; (2) it detects errors in a much more timely manner; and (3) it also has higher numerical stability and better fault coverage.

Room: 301-302-303
1:30 pm - 3:00 pm

Machine Learning and Quantum Computing

Performance Modeling under Resource Constraints Using Deep Transfer Learning

Authors: Aniruddha Marathe (Lawrence Livermore National Laboratory), Rushil Anirudh (Lawrence Livermore National Laboratory), Nikhil Jain (Lawrence Livermore National Laboratory), Abhinav Bhatele (Lawrence Livermore National Laboratory), Jayaraman Thiagarajan (Lawrence Livermore National Laboratory), Bhavya Kailkhura (Lawrence Livermore National Laboratory), Jae-Seung Yeom (Lawrence Livermore National Laboratory), Barry Rountree (Lawrence Livermore National Laboratory), Todd Gamblin (Lawrence Livermore National Laboratory)

Tuning application parameters for optimal performance is a combinatorially challenging problem. Hence, techniques for modeling the functional relationships between various input features in the parameter space and application performance are important. We show that simple statistical inference techniques are inadequate to capture these relationships, and that even with more complex ensembles of models, the minimum coverage of the parameter space required via experimental observations is still quite large. We propose a deep learning-based approach that can combine the knowledge from exhaustive observations collected at a smaller scale with limited observations collected at a larger target scale. The proposed approach is able to accurately predict performance in the regimes of interest to performance analysts, while outperforming many traditional techniques. In particular, our approach can identify the best performing configurations when trained using as little as 1% of observations at the target scale.

Obtaining Dynamic Scheduling Policies with Simulation and Machine Learning

Authors: Danilo Carastan-Santos (Federal University of ABC, Santo André, Brazil; University of Grenoble), Raphael Y. de Camargo (Federal University of ABC, Santo André, Brazil)

Dynamic scheduling of tasks in large-scale HPC platforms is normally accomplished using ad-hoc heuristics, based on task characteristics, combined with some backfilling strategy. Defining heuristics that work efficiently in different scenarios is a difficult task, specially when considering the large variety of task types and platform architectures.

In this work, we present a methodology based on simulation and machine learning to obtain dynamic scheduling policies. Using simulations and a workload generation model, we can determine the characteristics of tasks that lead to a reduction in the mean slowdown of tasks in an execution queue. Modeling these characteristics using a nonlinear function and applying this function to select the next task to execute in a queue improved the mean task slowdown in synthetic workloads. When applied to real workload traces from highly different machines, these functions still resulted in performance improvements, attesting the generalization capability of the obtained heuristics.

BP, BSP

0.5 Petabyte Simulation of a 45-Qubit Quantum Circuit

Authors: Thomas Häner (ETH Zurich), Damian S. Steiger (ETH Zurich)

Near-term quantum computers will soon reach sizes that are challenging to directly simulate, even when employing the most powerful supercomputers. Yet, the ability to simulate these early devices using classical computers is crucial for calibration, validation, and benchmarking. In order to make use of the full potential of systems featuring multi- and many-core processors, we use automatic code generation and optimization of compute kernels, which also enables performance portability. We apply a scheduling algorithm to quantum supremacy circuits in order to reduce the required communication and simulate a 45-qubit circuit on the Cori II supercomputer using 8,192 nodes and 0.5 petabytes of memory. To our knowledge, this constitutes the largest quantum circuit simulation to this date. Our highly-tuned kernels in combination with the reduced communication requirements allow an improvement in time-to-solution over state-of-the-art simulations by more than an order of magnitude at every scale.

Room: 402-403-404
2:00 pm - 3:00 pm

Performance Evaluation Tools

Representative Paths Analysis

Authors: *Nathan Tallent (Pacific Northwest National Laboratory), Darren Kerbyson (Pacific Northwest National Laboratory), Adolffy Hoisie (Pacific Northwest National Laboratory)*

Representative paths analysis generalizes and improves MPI critical path analysis. To improve diagnostic insight, we sample the distribution of program path costs and retain k representative paths. We describe scalable algorithms to collect representative paths and path profiles. To collect full paths efficiently, we introduce path pruning that reduces permanent space requirements from a trace (proportional to ranks) to path length (the minimum). To make space requirements independent of ranks and execution time --- a small constant in practice --- we profile program paths. Avoiding the limitations of prior path profiling approaches, we dynamically discover tasks and attribute costs in high resolution. We evaluate our algorithms on seven applications scaled up to 7000 MPI ranks. Full program paths use as little as 0.01% the permanent space of current methods; profiles require a nearly constant 100-1000 KB. Execution overhead is under 5% when synchronization intervals are sufficiently large (a few milliseconds).

ScrubJay: Deriving Knowledge from the Disarray of HPC Performance Data

Authors: *Alfredo Gimenez (Lawrence Livermore National Laboratory; University of California, Davis), Todd Gamblin (Lawrence Livermore National Laboratory), Abhinav Bhatele (Lawrence Livermore National Laboratory), Chad Wood (University of Oregon), Kathleen Shoga (Lawrence Livermore National Laboratory), Aniruddha Marathe (Lawrence Livermore National Laboratory), Peer-Timo Bremer (Lawrence Livermore National Laboratory), Bernd Hamann (University of California, Davis), Martin Schulz (Technical University Munich, Lawrence Livermore National Laboratory)*

Modern HPC centers comprise clusters, storage, networks, power and cooling infrastructure, and more. Analyzing the efficiency of these complex facilities is a daunting task. Increasingly, facilities deploy sensors and monitoring tools, but with millions of instrumented components, analyzing collected data manually is intractable. Data from an HPC center comprises different formats, granularities, and semantics, and handwritten scripts no longer suffice to transform the data into a digestible form.

We present ScrubJay, an intuitive, scalable framework for automatic analysis of disparate HPC data. ScrubJay decouples the task of specifying data relationships from the task of analyzing data. Domain experts can store reusable transformations that describe the projection of one domain onto another. ScrubJay also automates performance analysis. Analysts provide a query over logical domains of interest, and ScrubJay automatically derives needed steps to transform raw measurements. ScrubJay makes large-scale analysis tractable, reproducible, and provides insights into HPC facilities.

Room: 405-406-407
2:00 pm - 3:00 pm

Software for HPC Facilities

Charliecloud: Unprivileged Containers for User-Defined Software Stacks in HPC

Authors: *Reid Priedhorsky (Los Alamos National Laboratory), Tim Randles (Los Alamos National Laboratory)*

Supercomputing centers are seeing increasing demand for user-defined software stacks (UDSS) instead of or in addition to the stack provided by the center. These UDSS support user needs such as complex dependencies or build requirements, externally required configurations, portability, and consistency. The challenge for centers is to provide these services in a usable manner while minimizing the risks: security, support burden, missing functionality, and performance. We present Charliecloud, which uses the Linux user and mount namespaces to run industry-standard Docker containers with no privileged operations or daemons on center resources. Our simple approach avoids most security risks while maintaining access to the performance and functionality already on offer, doing so in just 800 lines of code. Charliecloud promises to bring an industry-standard UDSS user workflow to existing, minimally altered HPC resources.

Securing HPC: Development of a Low Cost, Open Source, Multi-Factor Authentication Infrastructure

Authors: *W. Cyrus Proctor (Texas Advanced Computing Center, University of Texas), Patrick Storm (Texas Advanced Computing Center, University of Texas), Matthew R. Hanlon (Texas Advanced Computing Center, University of Texas), Nathaniel Mendoza (Texas Advanced Computing Center, University of Texas)*

Multi-factor authentication (MFA) is rapidly becoming the de facto standard for access to all computing, whether via web, phone, or direct command-line access. HPC centers and other institutions supporting hundreds or thousands of users face challenging cost, licensing, user support, and infrastructure deployment decisions when considering a transition to MFA at scale.

This paper describes our experiences and lessons learned throughout the assessment, planning, and phased deployment of MFA across production systems supporting more than 10,000 accounts. It focuses on the ultimate curation, creation, and integration of a multitude of software components, some developed in-house and built to be compatible within existing HPC environments, and all

of which are freely available for open source distribution. We motivate the development of this customized infrastructure by highlighting some of the particular needs of our research community. What follows is an information resource for others when considering their own MFA deployments.

Room: 301-302-303

3:30 pm - 5:00 pm

Performance Tuning

Towards Fine-Grained Dynamic Tuning of HPC Applications on Modern Multi-Core Architectures

Authors: Mohammed Sourouri (Norwegian University of Science and Technology), Espen Birger Raknes (Aker BP ASA), Nico Reissmann (Norwegian University of Science and Technology), Johannes Langguth (Simula Research Laboratory), Daniel Hackenberg (Technical University Dresden), Robert Schöne (Technical University Dresden), Per Gunnar Kjeldsberg (Norwegian University of Science and Technology)

There is a consensus that exascale systems should operate within a power envelope of 20MW. Consequently, energy conservation is still considered as the most crucial constraint if such systems are to be realized.

So far, most research on this topic has focused on strategies such as power capping and dynamic power management. Although these approaches can reduce power consumption, we believe that they might not be sufficient to reach the exascale energy-efficiency goals. Hence, we aim to adopt techniques from embedded systems, where energy-efficiency has always been the fundamental objective.

A successful energy-saving technique used in embedded systems is to integrate fine-grained autotuning with dynamic voltage and frequency scaling. In this paper, we apply a similar technique to a real-world HPC application. Our experimental results on a HPC cluster indicate that such an approach can save up to 19% of energy compared to the baseline configuration, with negligible performance loss.

CAPEX: Unsupervised Storage Performance Tuning Using Neural Network-Based Deep Reinforcement Learning

Authors: Yan Li (University of California, Santa Cruz), Kenneth Chang (University of California, Santa Cruz), Oceane Bel (University of California, Santa Cruz), Ethan L. Miller (University of California, Santa Cruz), Darrell D. E. Long (University of California, Santa Cruz)

Parameter tuning is an important task of storage performance optimization. Current practice usually involves numerous tweak-benchmark cycles that are slow and costly. We present CAPEX, a model-less deep reinforcement learning-based unsupervised parameter tuning system driven by a deep neural network (DNN). It is designed to find optimal values for computer systems that have tunable parameters, from a simple client-server system to a large data center, where human tuning can be costly and often cannot achieve optimal performance. CAPEX takes periodic measurements of a target computer system's state, and trains a DNN which uses Q-learning to suggest changes to the system's current parameter values. CAPEX is minimally intrusive, and can be deployed into a production system to collect training data and suggest tuning actions during the system's daily operation. Evaluation of a prototype on a Lustre file system demonstrates an increase in I/O throughput up to 45% at saturation point.

BSP

sPIN: High-Performance Streaming Processing in the Network

Authors: Torsten Hoefler (ETH Zurich), Salvatore Di Girolamo (ETH Zurich), Konstantin Taranov (ETH Zurich), Ryan Grant (Sandia National Laboratories), Ronald Brightwell (Sandia National Laboratories)

Optimizing communication performance is imperative for large-scale computing because communication overheads limit the strong scalability of parallel applications. Today's network cards contain rather powerful processors optimized for data movement. However, these devices are limited to fixed functions, such as remote direct memory access. We develop sPIN, a portable programming model to offload simple packet processing functions to the network card. To demonstrate the potential of the model, we design a cycle-accurate simulation environment by combining the network simulator LogGOPSim and the CPU simulator gem5. We implement offloaded message matching, datatype processing, and collective communications and demonstrate full-application speedups of up to 3.7%. Furthermore, we show how sPIN can be used to accelerate redundant in-memory filesystems. Our work investigates a portable packet-processing network acceleration model similar to compute acceleration with CUDA or OpenCL. We show how such network acceleration enables an eco-system that can significantly speed up applications and system services.

BP

Room: 402-403-404

3:30 pm - 5:00 pm

Materials and Chemistry

Embracing a New Era of Highly Efficient and Productive Quantum Monte Carlo Simulations

Authors: Amrita Mathuriya (Intel Corporation), Ye Luo (Argonne National Laboratory), Raymond C. Clay III (Sandia National Laboratories), Anouar Benali (Argonne National Laboratory), Luke Shulenburger (Sandia National Laboratories), Jeongnim Kim

(Intel Corporation)

QMCPACK has enabled cutting-edge materials research on supercomputers for over a decade. It scales nearly ideally but has low single-node efficiency due to the physics-based abstractions using array-of-structures objects, causing inefficient vectorization. We present a systematic approach to transform QMCPACK to better exploit the new hardware features of modern CPUs in portable and maintainable ways. We develop miniapps for fast prototyping and optimizations. We implement new containers in structure-of-arrays data layout to facilitate vectorizations by the compilers. Further speedup and smaller memory-footprints are obtained by computing data on the fly with the vectorized routines and expanding single-precision use. All these are seamlessly incorporated in production QMCPACK. We demonstrate up-to 4.5x speedups on recent Intel® processors and IBM Blue Gene/Q for representative workloads. Energy consumption is reduced significantly commensurate to the speedup factor. Memory-footprints are reduced by up-to 3.8x, opening the possibility to solve much larger problems of future.

An Efficient MPI/OpenMP Parallelization of the Hartree-Fock Method for the Second Generation of Intel Xeon Phi Processor

Authors: Vladimir Mironov (*Lomonosov Moscow State University*), Alexander Moskovsky (*RSC Technologies*), Kristopher Keipert (*Iowa State University*), Michael D'mello (*Intel Corporation*), Mark Gordon (*Iowa State University*), Yuri Alexeev (*Argonne National Laboratory*)

Modern OpenMP threading techniques are used to convert the MPI-only Hartree-Fock code in the GAMESS program to a hybrid MPI/OpenMP algorithm. Two separate implementations that differ by the sharing or replication of key data structures among threads are considered, density and Fock matrices. All implementations are benchmarked on a supercomputer of 3,000 Intel Xeon Phi processors. With 64 cores per processor, scaling numbers are reported on up to 192,000 cores. The hybrid MPI/OpenMP implementation reduces the memory footprint by approximately 200 times compared to the legacy code. The MPI/OpenMP code was shown to run up to six times faster than the original for a range of molecular system sizes.

Efficient and Scalable Calculation of Complex Band Structure Using Sakurai-Sugiura Method

Authors: Shigeru Iwase (*University of Tsukuba*), Yasunori Futamura (*University of Tsukuba*), Akira Imakura (*University of Tsukuba*), Tetsuya Sakurai (*University of Tsukuba*), Tomoya Ono (*University of Tsukuba*)

Complex band structures (CBSs) are useful to characterize the static and dynamical electronic properties of materials. Despite the intensive developments, it is computationally demanding to calculate CBSs for systems consisting of over several hundred atoms. We propose an efficient and scalable first-principles calculation method to obtain CBSs. The basic idea is to express the Kohn-Sham equation of the real-space grid scheme as a quadratic eigenvalue problem and compute only the solutions which are necessary to construct the CBS by Sakurai-Sugiura method. The serial performance of the proposed method shows a significant advantage in both runtime and memory usage compared to the conventional method. Furthermore, owing to the hierarchical parallelism in Sakurai-Sugiura method, we can achieve an excellent scalability in the CBS calculation of a boron and nitrogen doped carbon nanotube consisting of more than 10,000 atoms using 2,048 nodes of Oakforest-PACS.

Room: 405-406-407
3:30 pm - 5:00 pm

State of the Practice: Operations

Failures in Large Scale Systems: Long-Term Measurement, Analysis, and Implications

Authors: Saurabh Gupta (*Intel Corporation*), Tirthak Patel (*Northeastern University*), Christian Engelmann (*Oak Ridge National Laboratory*), Devesh Tiwari (*Northeastern University*)

Resilience is one of the key challenges in maintaining high efficiency of future extreme scale supercomputers. Researchers and system practitioners rely on field-data studies to understand reliability characteristics and plan for future HPC systems. While the complexity of managing system reliability has increased, the number of studies covering comprehensive quantification and deep analysis of failures characteristics in large scale systems has not increased in the same proportion. To bridge this gap, in this work, we compare and contrast the reliability characteristics of multiple large-scale HPC production systems. Our study covers more than one billion compute node hours across five different systems over the period of 8 years. We confirm previous findings which continue to be valid, discover new findings, and discuss implications of new findings.

GUIDE: A Scalable Information Directory Service to Collect, Federate, and Analyze Logs for Operational Insights into a Leadership HPC Facility

Authors: Sudharshan S. Vazhkudai (*Oak Ridge National Laboratory*), Ross Miller (*Oak Ridge National Laboratory*), Devesh Tiwari (*Northeastern University*), Christopher Zimmer (*Oak Ridge National Laboratory*), Feiyi Wang (*Oak Ridge National Laboratory*), Sarp Oral (*Oak Ridge National Laboratory*), Raghul Gunasekaran (*Oak Ridge National Laboratory*), Deryl Steinert (*Oak Ridge National Laboratory*)

In this paper, we describe the GUIDE framework used to collect, federate, and analyze log data from the Oak Ridge Leadership Computing Facility (OLCF), and how we use that data to derive insights into facility operations. We collect system logs and extract monitoring data at every level of the various OLCF subsystems, and have developed a suite of pre-processing tools to make the

raw data consumable. The cleansed logs are then ingested and federated into a central, scalable data warehouse, Splunk, that offers storage, indexing, querying, and visualization capabilities. We have further developed and deployed a set of tools to analyze these multiple disparate log streams in concert and derive operational insights. We describe our experience from developing and deploying the GUIDE infrastructure, and deriving valuable insights on the various subsystems, based on two years of operations in the production OLCF environment.

Scientific User Behavior and Data-Sharing Trends in a Petascale File System

Authors: Seung-Hwan Lim (Oak Ridge National Laboratory), Hyogi Sim (Oak Ridge National Laboratory), Raghul Gunasekaran (Oak Ridge National Laboratory), Sudharshan S. Vazhkudai (Oak Ridge National Laboratory)

Oak Ridge Leadership Computing Facility (OLCF) runs some of the world's most powerful supercomputers, connected through a center-wide petascale file system. In this paper, we study the behavioral trends of 1,362 users and 387 projects from 39 scientific domains, by analyzing daily file system snapshots collected over 500 days.

The scope of our analysis encompasses: (i) a quantitative system-centric metrics of the file system; (ii) user behavior on the file system; and (iii) data-sharing trends between users and projects. To the best of our knowledge, our work is the first of its kind to provide comprehensive insights on user behavior through file system snapshot analysis of a large-scale shared file system. We envision that this study will provide valuable insights on the design, operation, and management of storage systems at scale, and also encourage other HPC centers to undertake similar such efforts.

Thursday, November 16th

Room: 402-403-404

10:30 am - 12:00 pm

Performance Analysis

Predicting the Performance Impact of Different Fat-Tree Configurations

Authors: Nikhil Jain (Lawrence Livermore National Laboratory), Abhinav Bhatele (Lawrence Livermore National Laboratory), Louis Howell (Lawrence Livermore National Laboratory), David Boehme (Lawrence Livermore National Laboratory), Ian Karlin (Lawrence Livermore National Laboratory), Edgar A. Leon (Lawrence Livermore National Laboratory), Misbah Mubarak (Argonne National Laboratory), Noah Wolfe (Rensselaer Polytechnic Institute), Todd Gamblin (Lawrence Livermore National Laboratory), Matthew Leining (Lawrence Livermore National Laboratory)

The fat-tree topology is one of the most commonly used network topologies in HPC systems. Vendors support several options that can be configured when deploying fat-tree networks for production systems, such as link bandwidth, number of rails, number of planes, and tapering. This paper showcases the use of simulations to compare the impact of these design options on representative production HPC applications, libraries, and multi-job workloads. We present advances in the TraceR-CODES simulation framework that enable this analysis and evaluate its prediction accuracy against experiments on a production fat-tree network. In order to understand the impact of network configurations on various anticipated scenarios, we study workloads with different communication patterns, computation-to-communication ratios, and scaling characteristics. Using multi-job workloads, we also study the impact of inter-job interference and different job placement schemes.

A Comparative Study of SDN and Adaptive Routing on Dragonfly Networks

Authors: Peyman Faizian (Florida State University), Md Atiqul Mollah (Florida State University), Zhou Tong (Florida State University), Xin Yuan (Florida State University), Michael Lang (Los Alamos National Laboratory)

The OpenFlow-style Software Defined Networking (SDN) technology has shown promising performance in data centers and campus networks; and the HPC community is significantly interested in adopting the SDN technology. However, while OpenFlow-style SDN allows dynamic per-flow resource management using a global network view, it does not support adaptive routing, which is widely used in HPC systems. This gives rise to the question whether SDN can achieve the performance that HPC systems expect with adaptive routing. In this work, we investigate possible methods to apply the SDN technology on the current generation HPC interconnects with the Dragonfly topology, and compare the performance of SDN with that of adaptive routing. Our results indicate that adaptive routing results in higher performance than SDN when both have similar resource allocation for a given traffic condition. However, SDN can use the global network view to compete with adaptive routing by allocating network resources more effectively.

Run-to-Run Variability on Xeon Phi Based Cray XC Systems

Authors: Sudheer Chunduri (Argonne National Laboratory), Kevin Harms (Argonne National Laboratory), Scott Parker (Argonne National Laboratory), Vitali Morozov (Argonne National Laboratory), Samuel Oshin (Intel Corporation), Naveen Cherukuri (Intel Corporation), Kalyan Kumaran (Argonne National Laboratory)

The increasing complexity of HPC systems has introduced new sources of variability, which can contribute to significant differences in run-to-run performance of applications. With components at various levels of the system contributing variability, application developers and system users are now faced with the difficult task of running and tuning their applications in an environment where

run-to-run performance measurements can vary by as much as a factor of two to three. In this study, we classify, quantify, and present ways to mitigate the sources of run-to-run variability on Cray XC systems with Intel Xeon Phi processors and a dragonfly interconnect. We further demonstrate that the code-tuning performance observed in a variability-mitigating environment correlates with the performance observed in production running conditions.

Room: 405-406-407
10:30 am - 12:00 pm

Communication Efficient Methods

Scaling Betweenness Centrality Using Communication-Efficient Sparse Matrix Multiplication

Authors: Edgar Solomonik (University of Illinois), Maciej Besta (ETH Zurich), Flavio Vella (Sapienza University of Rome), Torsten Hoefler (ETH Zurich)

Betweenness centrality (BC) is a crucial graph problem that measures the significance of a vertex by the number of shortest paths leading through it. We propose Maximal Frontier Betweenness Centrality (MFBC): a BC algorithm based on novel sparse matrix multiplication routines that performs asymptotically less communication than previous alternatives. We formulate, implement, and prove MFBC's correctness for weighted graphs by leveraging monoids instead of semirings, which enables a surprisingly succinct formulation. MFBC scales well for both extremely sparse and relatively dense graphs. It automatically searches a space of distributed data decompositions and sparse matrix multiplication algorithms for the most advantageous configuration. The MFBC implementation compares favorably to the CombBLAS library for social network graphs and is faster by up to 8x for denser random graphs. Our design methodology is readily extensible to other graph problems.

Distributed Southwell: An Iterative Method with Low Communication Costs

Authors: Jordi Wolfson-Pou (Georgia Institute of Technology), Edmond Chow (Georgia Institute of Technology)

We present a new algorithm, the Distributed Southwell method, as a competitor to Block Jacobi for preconditioning and multigrid smoothing. It is based on the Southwell iterative method, which is sequential, where only the equation with the largest residual is relaxed per iteration. The Parallel Southwell method extends this idea by relaxing equation i if it has the largest residual among all the equations coupled to variable i . Since communication is required for processes to exchange residuals, this method in distributed memory can be expensive. Distributed Southwell uses a novel scheme to reduce this communication of residuals while avoiding deadlock. Using test problems from the SuiteSparse Matrix Collection, we show that Distributed Southwell requires less communication to reach the same accuracy when compared to Parallel Southwell. Additionally, we show that the convergence of Distributed Southwell does not degrade like that of Block Jacobi when the number of processes is increased.

Tessellating Stencils

Authors: Liang Yuan (Chinese Academy of Sciences), Yunquan Zhang (Chinese Academy of Sciences), Peng Guo (Chinese Academy of Sciences), Shan Huang (Chinese Academy of Sciences)

Stencil computations represent a very common class of nested loops in scientific and engineering applications. The exhaustively studied tiling is one of the most powerful transformation techniques to explore the data locality and parallelism. Unlike previous work, which mostly blocks the iteration space of a stencil directly, this paper proposes a novel two-level tessellation scheme. A set of blocks are designed to tessellate the spatial space in various ways. The blocks can be processed in parallel without redundant computation. This corresponds to extending them along the time dimension and can form a tessellation of the iteration space. Experimental results show that our code performs up to 12% better than the existing highly concurrent schemes for the 3d27p stencil.

Room: 402-403-404
1:30 pm - 3:00 pm

Use and Management of Non-Volatile Memories

Transactional NVM Cache with High Performance and Crash Consistency

Authors: Qingsong Wei (Data Storage Institute), Chundong Wang (Data Storage Institute), Cheng Chen (Data Storage Institute, National University of Singapore), Yechao Yang (Data Storage Institute), Jun Yang (Data Storage Institute), Mingdi Xue (Data Storage Institute)

The byte-addressable non-volatile memory (NVM) is new promising storage medium. Compared to NAND flash memory, the next-generation NVM not only preserves the durability of stored data but has much shorter access latencies. An architect can utilize the fast and persistent NVM as an external disk cache. Regarding the system's crash consistency, a prevalent journaling file system needs to run atop an NVM disk cache. However, the performance is severely impaired by redundant efforts in achieving crash consistency in both file system and disk cache. Therefore, we propose a new mechanism called transactional NVM disk cache (Tinca). In brief, Tinca jointly guarantees consistency of file system and disk cache and removes the performance penalty of file system journaling with a lightweight transaction scheme. Evaluations confirm that Tinca significantly outperforms state-of-the-art

design by up to 2.5x in local and cluster tests without causing any inconsistency issue.

PapyrusKV: A High-Performance Parallel Key-Value Store for Distributed NVM Architectures

Authors: Jungwon Kim (Oak Ridge National Laboratory), Seyong Lee (Oak Ridge National Laboratory), Jeffrey Vetter (Oak Ridge National Laboratory)

This paper introduces PapyrusKV: a parallel embedded key-value store for a distributed HPC architectures that have nonvolatile memory (NVM). PapyrusKV stores keys and values in arbitrary byte arrays across multiple NVMs in a distributed system. PapyrusKV provides a set of standard key-value store operations such as put, get, and delete, and more advanced features specialized for HPC such as dynamic consistency configuration, database protection, zero-copy workflow, and asynchronous checkpoint/restart. Beyond filesystems, PapyrusKV provides HPC programmers a high-level interface to exploit massive pools of NVM in the system, and it organizes data to achieve high performance. Also, it allows HPC programmers to configure PapyrusKV to meet their application-specific requirements and/or needs. We evaluate PapyrusKV on three HPC systems, including OLCF's Summitdev, TACC's Stampede, and NERSC's Cori, using real NVM devices. Our results show that PapyrusKV can offer high performance, scalability, portability across various distributed NVM architectures, and usability for HPC.

Unimem: Runtime Data Management on Non-Volatile Memory-Based Heterogeneous Main Memory

Authors: Kai Wu (University of California, Merced), Yingchao Huang (University of California, Merced), Dong Li (University of California, Merced)

Non-volatile memory (NVM) provides a scalable and power-efficient solution to replace DRAM as main memory. However, because of relatively high latency and low bandwidth of NVM, NVM is often paired with DRAM to build a heterogeneous memory system (HMS). As a result, data objects of the application must be carefully placed to NVM and DRAM for best performance.

In this paper, we introduce a lightweight runtime solution that automatically and transparently manage data placement on HMS without the requirement of hardware modifications and disruptive change to applications. Leveraging online profiling and performance models, the runtime characterizes memory access patterns associated with data objects, and minimizes unnecessary data movement. Our runtime solution effectively bridges the performance gap between NVM and DRAM. We demonstrate that using NVM to replace the majority of DRAM can be a feasible solution for future HPC systems with the assistance of a software-based data management.

Room: 405-406-407

1:30 pm - 3:00 pm

Fast Multipole Methods and Linear Algebra

Geometry-Oblivious FMM for Compressing Dense SPD Matrices

Authors: Chenhan D. Yu (University of Texas), James Levitt (University of Texas), Severin Reiz (Technical University Munich), George Biros (University of Texas)

We present GOFMM (geometry-oblivious FMM), a novel method that creates a hierarchical low-rank approximation, or "compression," of an arbitrary dense symmetric positive definite (SPD) matrix. For many applications, GOFMM enables an approximate matrix-vector multiplication in $N \log N$ or even N time, where N is the matrix size. Compression requires $N \log N$ storage and work. In general, our scheme belongs to the family of 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 Pascal architecture for a variety of matrices.

Low Communication FMM-Accelerated FFT on GPUs

Authors: Cris Cecka (Nvidia Corporation)

Communication-avoiding algorithms have been the subject of growing interest in the last decade due to the growth of distributed memory systems and the disproportionate increase of computational throughput to communication bandwidth. For distributed 1D FFTs, communication costs quickly dominate execution time as all industry-standard implementations perform three all-to-all transpositions of the data.

In this work, we reformulate an existing algorithm that employs the Fast Multipole Method to reduce the communication requirements to approximately a single all-to-all transpose. We present a detailed and clear implementation strategy that relies on existing library primitives, demonstrate that this implementation achieves consistent speed-ups between 1.3x and 2.2x against cuFFTX on 2xP100 and 8xP100 GPUs, and develop an accurate compute model to analyze the performance dependencies.

Designing Vector-Friendly Compact BLAS and LAPACK Kernels

Authors: Kyungjoo Kim (Sandia National Laboratories), Timothy B. Costa (Intel Corporation), Mehmet Deveci (Sandia National Laboratories), Andrew M. Bradley (Sandia National Laboratories), Simon D. Hammond (Sandia National Laboratories), Murat E. Guney (Intel Corporation), Sarah Knepper (Intel Corporation), Shane Story (Intel Corporation), Sivasankaran Rajamanickam (Sandia National Laboratories)

Many applications rely on the use of blas/lapack routines on large groups of very small matrices. For example, many PDE-based simulations and machine learning applications require batched blas/lapack routines. While existing batched blas APIs provide meaningful speedup over alternatives like OpenMP loops around traditional blas/lapack, there exists potential for significant speedup by considering a non-canonical data-layout that allows for cross-matrix vectorization in batched blas/lapack routines.

We propose a new compact data-layout that interleaves matrices in blocks according to the architecture's SIMD vector-length and investigate its merits. Second, we discuss the proposed data-layout in two libraries, an open-source and a vendor implementation. In our experiments, the compact data layout provides up to 5x, 15x and 18x speedup against batched dgemm, dtrsm and dgetrf respectively with a blocksize 5 on Intel Knights Landing. Finally, we demonstrate the improved performance by using the compact data-layout in a line solver for coupled CFD codes.

Room: 402-403-404

3:30 pm - 5:00 pm

In-System Processing for Performance

Input-Aware Auto-Tuning of Compute-Bound HPC Kernels

Authors: Philippe Tillet (Harvard University), David Cox (Harvard University)

Efficient implementations of HPC applications for parallel architectures generally rely on external software packages (e.g., BLAS, LAPACK, CUDNN). While these libraries provide highly optimized routines for certain characteristics of inputs (e.g., square matrices), they generally don't retain optimal performance across the wide range of problems encountered in practice. In this paper, we present ISAAC, an input-aware auto-tuning framework for matrix multiplications and convolutions, capable of generating not only hardware, but also application-specific compute kernels, by combining highly parameterized PTX kernel templates with data-driven performance modeling. Numerical experiments on the NVIDIA Maxwell/Pascal architectures shows up to 3x performance gains over both cuBLAS and cuDNN after only a few hours of auto-tuning.

Leveraging Near Data Processing for High-Performance Checkpoint/Restart

Authors: Abhinav Agrawal (North Carolina State University), Gabriel H. Loh (Advanced Micro Devices Inc), James Tuck (North Carolina State University)

With the increasing size of HPC systems, the system mean time to interrupt will decrease. This requires checkpoints to be stored in a smaller time when using checkpoint/restart (C/R) for mitigation. Multilevel checkpointing improves C/R efficiency by saving most checkpoints to fast compute-node local storage. But it incurs a high cost for writing a few checkpoints to slow global-I/O. We show that leveraging NDP to offload writing of checkpoints to global-I/O improves C/R efficiency. We explore additional opportunities using NDP to further reduce C/R overhead and evaluate checkpoint compression using NDP as a starting point.

We evaluate the performance of our novel application of NDP for C/R and compare it to existing C/R optimizations. Our evaluation for a projected exascale system using multilevel checkpointing shows that with NDP, the host processor is able to increase its efficiency on an average from 51% to 78% (i.e., a >50% speedup in performance).

Melissa: Large Scale In Transit Global Sensitivity Analysis Avoiding Intermediate Files

Authors: Théophile Terraz (French Institute for Research in Computer Science and Automation (INRIA)), Alejandro Ribes (EDF France), Yvan Fournier (EDF France), Bertrand Iooss (EDF France), Bruno Raffin (French Institute for Research in Computer Science and Automation (INRIA))

Global sensitivity analysis is an important step for analyzing and validating numerical simulations. One classical approach consists in computing statistics from the outputs of multiple simulation runs. Results are stored to disk and statistics are computed postmortem. Scientists are constrained to run low resolution simulations with a limited number of probes to keep the amount of intermediate storage manageable. In this paper we propose a file avoiding, fault tolerant, and elastic framework that enables high resolution global sensitivity analysis at large scale. Our approach combines iterative statistics and in transit processing to compute Sobol' indices without any intermediate storage. Statistics are updated on-the-fly as soon as the in-transit parallel server receives results from one of the running simulations. For one experiment, we computed the Sobol' indices on 10M hexahedra and 100 timesteps, running 8000 parallel simulations executed in 1h27 on up to 28672 cores, avoiding 48TB of file storage.

Room: 405-406-407

3:30 pm - 5:00 pm

Optimizing MPI

Why Is MPI So Slow? Analyzing the Fundamental Limits in Implementing MPI-3.1

Authors: Ken Raffenetti (Argonne National Laboratory), Abdelhalim Amer (Argonne National Laboratory), Lena Oden (Argonne National Laboratory), Charles Archer (Intel Corporation), Wesley Bland (Intel Corporation), Hajime Fujita (Intel Corporation), Yanfei Guo (Argonne National Laboratory), Tomislav Janjusic (Mellanox Technologies), Dmitry Durnov (Intel Corporation), Michael Blocksome (Intel Corporation), Min Si (Argonne National Laboratory), Sangmin Seo (Argonne National Laboratory), Akhil Langer (Intel Corporation), Gengbin Zheng (Intel Corporation), Masamichi Takagi (RIKEN), Paul Coffman (Argonne National Laboratory), Jithin Jose (Intel Corporation), Sayantan Sur (Intel Corporation), Alexander Sannikov (Intel Corporation), Sergey Oblomov (Intel Corporation), Michael Chuvelev (Intel Corporation), Masayuki Hatanaka (RIKEN), Xin Zhao (Mellanox Technologies), Paul Fischer (University of Illinois), Thilina Rathnayake (University of Illinois), Matt Otten (Cornell University), Misun Min (Argonne National Laboratory), Pavan Balaji (Argonne National Laboratory)

This paper provides an in-depth analysis of the software overheads in the MPI performance-critical path and exposes mandatory performance overheads that are unavoidable based on the MPI-3.1 specification. We first present a highly optimized implementation of the MPI-3.1 standard where the communication stack---all the way from the application to the low-level network communication API---takes only a few tens of instructions. We then carefully study these instructions and analyze the root cause of these overheads based on specific requirements from the MPI standard that are unavoidable under the current MPI standard. We finally recommend potential changes to the MPI standard that can minimize these overheads. Our experimental results on a variety of network architectures and applications demonstrate significant benefits from our proposed changes.

ParaStack: Efficient Hang Detection for MPI Programs at Large Scale

Authors: Hongbo Li (University of California, Riverside), Zizhong Chen (University of California, Riverside), Rajiv Gupta (University of California, Riverside)

While program hangs on large parallel systems can be detected via the widely used timeout mechanism, it is difficult to set an optimal timeout threshold if users have limited knowledge of a program. Too small timeout will lead to high false alarm rates, and too large timeout will waste valuable computing resources. This paper presents a highly efficient hang detection tool, ParaStack, that does not rely on timeout. We have adapted ParaStack to work with Torque and Slurm parallel job schedulers and validated both its functionality and performance on the current world's tenth fastest supercomputer Stampede. Experimental results demonstrate that ParaStack can detect hangs accurately, in a timely manner, and at negligible runtime cost. Also ParaStack pinpoints the faulty processes with high accuracy when the hang is caused by errors in computation phase.

Scalable Reduction Collectives with Data Partitioning-Based Multi-Leader Design

Authors: Mohammadreza Bayatpour (Ohio State University), Sourav Chakraborty (Ohio State University), Hari Subramoni (Ohio State University), Xiaoyi Lu (Ohio State University), Dhabaleswar K. Panda (Ohio State University)

Existing designs for MPI Allreduce do not take advantage of the vast parallelism available in modern multi-/many-core processors like Intel Xeon/Xeon Phi or the increases in communication throughput and recent advances in high-end features seen with modern interconnects like InfiniBand and OmniPath. In this paper, we propose a high-performance and scalable Data Partitioning-based Multi-Leader (DPML) solution for MPI Allreduce that can take advantage of the parallelism offered by multi-/many-core architectures in conjunction with high throughput and high-end features offered by InfiniBand and Omni-Path to significantly enhance the performance of MPI Allreduce on modern HPC systems. We also model DPML-based designs to analyze the communication costs theoretically. Microbenchmark level evaluations show that the proposed DPML-based designs are able to deliver up to 3.5 times performance improvement for MPI Allreduce for multiple HPC systems at scale. At the application-level, up to 35% and 60% improvements is seen for HPCG and miniAMR respectively.

Tuesday, November 14th

Room: Mile High Prefunction
8:30 am - 5:00 pm

Scientific Visualization and Data Analytics Showcase Posters

Visualization of Decision-Making Support (DMS) Information for Responding to a Typhoon-Induced Disaster

Authors: Dongmin Jang (Korea Institute of Science and Technology Information), Jin-Hee Yuk (Korea Institute of Science and Technology Information), Junghyun Park (Korea Institute of Science and Technology Information), Jooneun An (Korea Institute of Science and Technology Information), Minsu Joh (Korea Institute of Science and Technology Information)

A high-resolution coupled atmosphere, ocean, and inundation (flood) modeling and simulation system was developed for scientific, accurate, fast, and efficient forecasting of typhoon-induced disasters. This is based on the KISTI decision-making support system (K-DMSS). Our prediction system consists of a typhoon, surge/wave, and flooding prediction and analysis systems (TPAS, SPAS, and FPAS). In this research, we simulated Typhoon 'CHABA' (1618), which was ranked third among the most intense tropical cyclones and was the most powerful typhoon in the Republic of Korea (South Korea) in 2016. The CHABA-induced storm surge and inundation were simulated using our prediction and analysis system. To understand intuitively the changes of physical phenomena and damage caused by the typhoon, numerical data sets produced by the prediction and analysis systems were visualized using VAPOR (Visualization and Analysis Platform for Ocean, atmosphere, and solar Researchers, and which is a part of the K-DMSS) as one of the visualization systems.

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Visualizations of a High-Resolution Global-Regional Nested, Ice-Sea-Wave Coupled Ocean Model System

Authors: Kangyou Zhong (Sun Yat-Sen University), Danya Xu (Sun Yat-Sen University), Changsheng Chen (University of Massachusetts, Dartmouth; Sun Yat-Sen University), Yutong Lu (Sun Yat-Sen University), Wenjie Dong (Sun Yat-Sen University), Jiang Li (Sun Yat-Sen University)

A multi-scale, global-regional nested ocean modeling system based on the unstructured grid Finite Volume Community Ocean Model (FVCOM) has been deployed on the Tianhe-2 supercomputer providing 24/7 marine forecasting since September 2016. The modeling system is part of the Sun Yat-Sen University Community Integrated Model (SYCIM) project for developing a new generation Earth System Model to explore the physical mechanisms of climate change. With a horizontal resolution up to ~17 m, this high-resolution modeling system can properly resolve the complex dynamical interactions of estuary, near shore coast, continental shelf, and deep ocean basins. The submitted animation shows the modeled global surface waves distribution pattern and propagation in the world ocean basins and in the South China Sea. Additionally, the variations of the global sea surface temperature, Arctic and Antarctic sea ice thickness and extension simulated by SYCIM is also presented. The animation can not only be used to visualize the big data for scientific research but also as a good example to demonstrate dynamical variations of the ocean to non-scientific communities. We hope this animation can help to arouse public attention on some issues such as global warming, polar sea ice melting, ocean environment, and marine hazards.

Milky Way Analogue Isolated Disk Galaxy

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications,

University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois), Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Sriv (University of Illinois)

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center; University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Room: Four Seasons Ballroom
8:30 am - 5:00 pm

Research Posters

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

8:30 am - 5:00 pm

ACM Student Research Competition

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

Room: Four Seasons Ballroom

5:15 pm - 7:00 pm

Poster Reception

A01: GEMM-Like Tensor-Tensor Contraction (GETT)

Authors: *Paul Springer (RWTH Aachen University), Paolo Bientinesi (RWTH Aachen University)*

Tensor contractions (TC) are a performance critical component in numerous scientific computations. Despite the close connection between matrix-matrix products (GEMM) and TCs, the performance of the latter is in general vastly inferior to that of an optimized GEMM. To close such a gap, we propose a novel approach: GEMM-like Tensor-Tensor multiplication (GETT). GETT mimics the design of a high-performance GEMM implementation; as such, it systematically reduces an arbitrary tensor contractions to a highly-optimized "macro-kernel". This macro-kernel operates on suitably "packed" sub-tensors that reside in specified levels of the cache hierarchy. GETT's decisive feature is its ability to pack subtensors via tensor transpositions, yielding efficient packing routines. In contrast to previous approaches to TCs, GETT attains the same I/O cost as an equally-sized GEMM, making GETT especially well-suited for bandwidth-bound TCs. GETT's excellent performance is highlighted across a wide range of random tensor contractions.

SF

A02: Accelerating the Higher Order Singular Value Decomposition Algorithm for Big Data with GPUs

Authors: *Yuhsiang M. Tsai (National Taiwan University)*

With the explosion of big data, finding ways of compressing large datasets with multi-way relationship - i.e., tensors - quickly and efficiently has become critical in HPC.

High-order singular value decomposition (HOSVD) method provides us with the means to attain both extremely high compression ratio and low error rate through low-rank approximation.

However, parallelizing HOSVD efficiently on GPUs remains a challenging problem, largely due to the lack of a fast SVD implementation that can stream data to the limited GPU memory through the PCIe bottleneck.

Our work studies, optimizes, and then contrasts four different methods for calculating singular vectors for performance, weak/strong scalability and accuracy in the context of HOSVD. We also discuss ways of load balancing the problem across multiple GPUs on a single node, and discuss the pros and cons of these different algorithms for GPU acceleration.

SF

A03: A High-Speed Algorithm for Genome-Wide Association Studies on Multi-GPU Systems

Authors: *Yen Chen Chen (National Taiwan University)*

We develop an algorithm as long as a CUDA code for GWAS (Genome-Wide Associate Studies). This algorithm can work efficiently on GPU and has high scalability. The core of the algorithm is an accurate and fast p-value integration reformation, which accelerates the most time-consuming part of the algorithm. With the algorithm, researchers can now deal with tens of billions of SNP to trait pair in only a few minutes. Even better, since this algorithm is highly scalable, you can increase the problem size as long as you have enough computing power.

SF

A04: Optimization of the AIREBO Many-Body Potential for KNL

Authors: *Markus Höhnerbach (RWTH Aachen University)*

Molecular dynamics simulations are an indispensable research tool for computational chemistry and material science. Empirical many-body potentials promise high-fidelity simulations that capture bonding and reaction behavior accurately, providing a level of detail in between more classical molecular dynamics and quantum methods.

The AIREBO potential is one such example that provides forces and energies for molecular dynamics (MD) simulations of carbon and carbohydrate structures. Allowing many-body potentials to profit from the recent architectural advances still poses a challenge due to deeply nested, short loops. We develop an optimized, vectorized AIREBO implementation for Intel's Xeon and Xeon Phi (co)processors and integrate it into the open-source LAMMPS molecular dynamics code. By both introducing improvements to the code and vectorization, we achieve a sustained real-word speedup of two on Broadwell, and a speedup of four on KNL. The

optimized code will be distributed with each LAMMPS download as part of the USER-INTEL package.
SF

A05: Parallel Prefix Algorithms for the Registration of Arbitrarily Long Electron Micrograph Series

Authors: Marcin Copik (RWTH Aachen University)

Recent advances in the technology of transmission electron microscopy have allowed for a more precise visualization of materials and physical processes, such as metal oxidation. Nevertheless, the quality of information is limited by the damage caused by an electron beam, movement of the specimen or other environmental factors. A novel registration method has been proposed to remove those limitations by acquiring a series of low dose microscopy frames and performing a computational registration on them to understand and visualize the sample. This process can be represented as a prefix sum with a complex and computationally intensive binary operator and a parallelization is necessary to enable processing long series of microscopy images. With our parallelization scheme, the time of registration of results from ten seconds of microscopy acquisition has been decreased from almost thirteen hours to less than seven minutes on 512 Intel IvyBridge cores.

A06: Accelerating Big Data Processing in the Cloud with Scalable Communication and I/O Schemes

Authors: Shashank Gugrani (Ohio State University)

With the advent of cloud computing, the field of Big Data has seen rapid growth. Most cloud providers provide hardware resources such as NVMe SSDs, large memory nodes, and SR-IOV. This opens up the possibility of large-scale high-performance data analytics and provides opportunities to use these resources to develop new designs. Cloud computing provides flexibility, scalability, and reliability, which are important requirements of Big Data frameworks. However, several important requirements are missing, such as performance, scalability, fault-tolerance, and consistency. The focus of this research work revolves around developing communication and I/O designs and concepts which can provide these requirements to Big Data frameworks. Specifically, we explore new ways to provide fault-tolerance and consistency in cloud storage systems, providing scalable and high-performance communication frameworks, and co-designing with Big Data stacks to leverage these features.

A07: Scalable Parallel Scripting in the Cloud

Authors: Benjamin H. Glick (Lewis & Clark College)

It's often complicated, time consuming, but frequently necessary to successfully port complex workflows to multiple high-performance environments. Parsl is a Python-based parallel scripting library that provides a simple model for describing and executing dataflow-based scripts over arbitrary execution resources such as clouds, campus clusters, and high-performance systems. Parsl's execution layer abstracts the differences between providers enabling provisioning and management of compute nodes for use with a pilot system. In this poster, we describe the development of a new execution provider designed to support Amazon Web Services (AWS) and Microsoft's Azure. This provider supports the transparent execution of implicitly parallel Python-based scripts using elastic cloud resources. We demonstrate that Parsl is capable of executing thousands of applications per second over this elastic execution fabric.

A08: Virtualized Big Data: Reproducing Simulation Output on Demand

Authors: Salvatore Di Girolamo (ETH Zurich)

Scientific simulations are being pushed to the extreme in terms of size and complexity of the addressed problems, producing astonishing amount of data. If the data is stored on disk, analysis applications can randomly access simulation output. Yet, storing the massive amounts simulation data is challenging. This is primarily due to the high storage costs and the fact that compute capabilities grow faster than storage capacities and bandwidths. In-situ analysis removes the storage costs but applications lose random access.

We propose to not store the full simulation output data but to produce it on demand. Our system intercepts I/O requests of both analysis tools and simulators, enabling data virtualization. This new paradigm allows us to explore the computation-storage tradeoff, by trading computation power for storage space. Overall, SDaVi offers a viable path towards exa-scale scientific simulations, by exploiting the growing computing power and relaxing the storage capacity requirements.

SF

A09: Ring: Unifying Replication and Erasure Coding to Rule Resilience in KV-Stores

Authors: Konstantin Taranov (ETH Zurich)

There is a wide range of storage schemes employed by KV-stores to ensure reliability of stored keys. However, previous implementations do not allow choosing the storage scheme dynamically, thereby forcing developers to commit to a single scheme. Such inefficient data management wastes cluster resources such as memory usage, network load, latency, availability and many others.

To solve this problem, we have designed a strongly consistent key-value store Ring that empowers its users to explicitly manage storage parameters like other resources, such as memory or processor time. The key feature of Ring is that all keys live in the same strongly consistent namespace and a user does not need to specify the resilience when looking up a key or value.

Our poster demonstrates how future applications that manage resilience of key-value pairs consciously can reduce the overall

operational cost and improve performance significantly.

A10: Revealing the Power of Neural Networks to Capture Accurate Job Resource Usage from Unparsed Job Scripts and Application Inputs

Authors: *Michael R. Wyatt (University of Delaware)*

Next generation HPC schedulers will rely heavily on accurate information about resource usage of submitted jobs. The information provided by users is often inaccurate and previous prediction models, which rely on parsed job script features, fail to accurately predict for all HPC jobs. We propose a new representation of job scripts and inclusion of application input decks for resource usage predictions with a neural network. Our contributions are a method for representing job scripts as image-like data, an automated method for predicting job resource usage from job script images and input deck features, and validation of our methods with real HPC data. We demonstrate that when job scripts for an application are very similar, our method performs better than other methods. We observe an average decrease in error of 2 node-hours compared to state of the art methods.

A11: Finding a Needle in a Field of Haystacks: Lightweight Metadata Search for Large-Scale Distributed Research Repositories

Authors: *Anna Blue Keleher (University of Maryland)*

Fast, scalable, and distributed search services are commonly available for single nodes, but lead to high infrastructure costs when scaled across tens of thousands of filesystems and repositories, as is the case with Globus. Endpoint-specific indexes may instead be stored on their respective nodes, but while this distributes storage costs between users, it also creates significant query overhead. Our solution provides a compromise by introducing two levels of indexes: a single centralized "second-level index" (SLI) that aggregates and summarizes terms from each endpoint; and many endpoint-level indexes that are referenced by the SLI and used only when needed. We show, via experiments on Globus-accessible filesystems, that the SLI reduces the amount of space needed on central servers by over 96% while also reducing the set of endpoints that need to execute user queries.

SF

A12: Applying Image Feature Extraction to Cluttered Scientific Repositories

Authors: *Emily Herron (Mercer University)*

Over time many scientific repositories and file systems become disorganized, containing poorly described and error-ridden data. As a result, it is often difficult for researchers to discover crucial data. In this poster, we present a collection of image processing modules that collectively extract metadata from a variety of image formats. We implement these modules in Skluma—a system designed to automatically extract metadata from structured and semi-structured scientific formats. Our modules apply several image metadata extraction techniques that include processing file system metadata, header information, color content statistics, extracted text, feature-based clusters, and predicting tags using a supervised learning model. Our goal is to collect a large number of metadata that may then be used to organize, understand, and analyze data stored in a repository.

SF

A13: Deep Learning with HPC Simulations for Extracting Hidden Signals: Detecting Gravitational Waves

Authors: *Daniel George (National Center for Supercomputing Applications, University of Illinois)*

We introduce Deep Filtering, new machine learning method for end-to-end time-series signal processing, which combines two deep one-dimensional convolutional neural networks for classification and regression to detect and characterize signals much weaker than the background noise. We trained this method with a novel curriculum learning scheme on data derived from HPC simulations and applied it for gravitational wave analysis specifically for mergers of black holes and demonstrated that it significantly outperforms conventional machine learning techniques, is far more efficient than matched-filtering, offering several orders-of-magnitude speed-up, allowing real-time processing of raw big data with minimal resources, and extends the range of detectable signals. This initiates a new paradigm for scientific research which employs massively-parallel numerical simulations to train artificial intelligence algorithms that exploit emerging hardware architectures such as deep-learning-optimized GPUs. Our approach offers a unique framework to enable coincident detection campaigns of gravitational wave sources and their electromagnetic counterparts.

SF

A14: Analysis of Synthetic Graph Generation Methods for Directed Network Graphs

Authors: *Spencer Callicott (Mississippi State University)*

Historically, scientific experiments have been conducted to generate scale-free network graphs based on structure. Metrics used to measure veracity ensure the integrity of a scale-free algorithm given a seed. However, studies do not explore the performance benefits or drawbacks of specific algorithms running on Apache Spark and GraphX. Recognizing the lack of performance benchmarks demands ensuring accuracy through experimenting. This study will utilize the Stochastic Kronecker Graph model to synthetically generate graphs given a seed graph.

A15: Quantifying Compiler Effects on Code Performance and Reproducibility Using FLiT

Authors: *Michael Bentley (University of Utah)*

A busy application developer likes to focus on doing science, but instead is often distracted by the sheer variety of available hardware platforms, their compilers, and associated optimization flags. Exclusive pursuit of speed may jeopardize the reproducibility

of scientific experiments. On the other hand, performance is central to HPC. Our previous work provided a unique testing framework called FLiT that helps developers exploit performance without jeopardizing reproducibility. To verify that FLiT is useful for real-world libraries and applications, it was applied to MFEM, a finite element library used in various HPC applications. I show that the compilation with the fastest average runtime for the converted MFEM examples is also bitwise reproducible. For these examples, clang had the fastest average runtimes and the best reproducibility. Our future work aims to enhance the open-source FLiT tool into a strong community resource and to follow up with found compiler oddities.

A16: Diagnosing Parallel I/O Bottlenecks in HPC Applications

Authors: *Peter Z. Harrington (University of California, Santa Cruz)*

HPC applications are generating increasingly large volumes of data (up to hundreds of TBs), which need to be stored in parallel to be scalable. Parallel I/O is a significant bottleneck in HPC applications, and is especially challenging in Adaptive Mesh Refinement (AMR) applications because the structure of output files changes dynamically during runtime. Data-intensive AMR applications run on the Cori supercomputer show variable and often poor I/O performance, but diagnosing the root cause remains challenging. Here we analyze logs from multiple levels of Cori's parallel I/O subsystems, and find bottlenecks during file metadata operations and during the writing of file contents that reduced I/O bandwidth by up to 40x. Such bottlenecks seemed to be system-dependent and not the application's fault. Increasing the granularity of file-system performance data will help provide conclusive causal relationships between file-system servers and metadata bottlenecks.

SF

A17: Toward Capturing Nondeterminism Motifs in HPC Applications

Authors: *Dylan Chapp (University of Delaware)*

High performance MPI applications employ nondeterministic asynchronous communication to achieve greater performance. However, this nondeterminism can significantly hamper debugging. Various software tools have been developed to control nondeterminism in HPC applications, but a high-level application-agnostic taxonomy for this nondeterminism is absent and limits these tools' effectiveness in practice. We propose to address this need by extracting common nondeterministic communication motifs from representative applications.

We present a first step toward capturing nondeterminism motifs by way of a workflow for detecting and summarizing sender nondeterminism in HPC applications.

A18: Understanding the Impact of Fat-Tree Network Locality on Application Performance

Authors: *Philip Taffet (Rice University)*

Network congestion can be a significant cause of performance loss and variability for many message passing programs. However, few studies have used a controlled environment with virtually no other extraneous sources of network traffic to observe the impact of application placement and multi-job interactions on overall performance. We study different placements and pairings for three DOE applications. We observe that for a job size typical for an LLNL commodity cluster, the impact of congestion and poor placement is typically less than 2%, which is less dramatic than on torus networks. In addition, in most cases, the cyclic MPI task mapping strategy increases performance and reduces placement sensitivity despite also increasing total network traffic. We also found that the performance difference between controlled placements and runs scheduled through the batch system was less than 3%.

SF

A19: Performance Analysis of a Parallelized Restricted Boltzmann Machine Artificial Neural Network Using OpenACC Framework and TAU Profiling System

Authors: *Abhishek Kumar (Brookhaven National Laboratory)*

Restricted Boltzmann Machines are stochastic neural networks that create probability distribution based off connection weight between nodes of the hidden and visible layer. The distribution makes the program optimal at classifying large amounts of data, which could be useful in work settings, such as a research lab. The parallelization of these neural networks would allow for the classification of data at a much faster rate than before. Using a high-performance computer it was determined that parallelizing the neural networks could decrease the runtime of the algorithm by over 35% when offloading the work to a GPU through OpenACC. Using Tuning and Analysis Utilities Profiling Systems, it was found that scheduling the program would only be effective if the data size was large enough and an increase in the number of thread blocks used for scheduling would allow for greater performance gains than the number of threads in each thread block.

A20: Correctness Verification and Boundary Conditions for Chapel Iterator-Based Loop Optimization

Authors: *Daniel A. Feshbach (Haverford College)*

We explore two issues of correctness concerning iteration space transformation techniques: data dependencies and boundary conditions. First, we present a data structure which automatically verifies correctness of data dependencies for stencil computations with transformed iteration spaces. This further confirms the viability of Chapel iterators for defining iteration space transformations, by demonstrating that simple tool support can verify data dependencies and assist debugging. Second, we explore the performance and simplicity of three strategies for implementing boundary conditions in transformed iteration spaces: if statements, loop peeling, and an array of coefficients. We find that the coefficient array technique performs the best, often at 70 to 80 percent speed of the

benchmark of ignoring the boundary condition. If statements are not far behind, while loop peeling performs much worse. The coefficient array and if statements are indifferent to the transformation technique applied, while loop peeling must be implemented within the transformation.

A21: Runtime Support for Concurrent Execution of Overdecomposed Heterogeneous Tasks

Authors: Jaemin Choi (University of Illinois)

With the rise of heterogeneous systems in high performance computing, how we utilize accelerators has become a critical factor in achieving the optimal performance. We explore several issues with using accelerators in Charm++, a parallel programming model that employs overdecomposition. We propose a runtime support scheme that enables concurrent execution of heterogeneous tasks and evaluate its performance. Using a synthetic benchmark that utilizes busy-waiting to simulate workload, we observe that the effectiveness of the runtime support varies with the application characteristics, with a maximum speedup of 4.79x. With a two-dimensional five-point stencil benchmark designed to represent a realistic workload, we obtain up to 2.75x speedup.

A22: Verifying Functional Equivalence Between C and Fortran Programs

Authors: Wenhao Wu (University of Delaware)

Software verification is a mature research area with many techniques. These verification approaches can be applied to programs written in different programming languages; nevertheless, most verification tools are only designed for programs written in C or Java. As a result, verification tools are inadequate for other languages, such as Fortran. A high level of software safety is mandatory in most of its application scenarios, which makes verification tools for Fortran programs necessary and significant.

In this poster, the author illustrates the motivation and objectives of the project with examples. Also, this poster shows an extension (as a Fortran program verifier) of an existing verification platform -- CIVL. Additionally, the results of a set of extensive experiments conducted by the author is shown in this poster to indicate that the performance is satisfactory.

A23: Evaluation of Data-Intensive Applications on Intel Knights Landing Cluster

Authors: Tao Gao (University of Delaware)

Analyzing and understanding large datasets on high performance computing platforms is becoming more and more important in various scientific domains. MapReduce is the dominant programming model for processing these datasets. Platforms for data processing are empowered by many-core nodes with cutting-edge processing units. Intel Knights Landing (KNL) is the new arrival in the field. However, this new architecture has not been fully evaluated for data-intensive applications. In this poster, we present the assess of KNL on the performance of three key data-intensive applications based on a high-performance MapReduce programming framework on the latest KNL-cluster, Stampede2. We focus on the impact of different KNL memory models, we compare Stampede2 with other clusters such as Tianhe-2 and Mira, and we measure the scalability of large datasets. We observe how KNL-based clusters are a promising architecture for data-intensive applications. We also identify key aspects to enable more efficient usage of KNL-based clusters.

A24: Comparison of Machine Learning Algorithms and Their Ensembles for Botnet Detection

Authors: Songhui Ryu (Purdue University)

A Botnet is a network of compromised devices that is controlled by malicious 'botmaster' in order to perform various tasks, such as executing DoS attack, sending SPAM and obtaining personal data etc. As botmasters generate network traffic while communicating with their bots, analyzing network traffic to detect Botnet traffic can be a promising feature of Intrusion Detection System(IDS). Although IDS has been applying various machine learning (ML) techniques, comparison of ML algorithms including their ensembles on Botnet detection has not been figured out yet. In this study, not only the three most popular classification ML algorithms – Naïve Bayes, Decision tree, and Neural network are evaluated, but also the ensemble methods known to strengthen ML algorithms are tested to see if they indeed provide enhanced predictions on Botnet detection. This evaluation is conducted with CTU-13 public dataset, measuring running time of each ML and its f measure and MCC score.

A25: Investigating Performance of Serialization Methods for Networked Data Transfer in HPC Applications

Authors: Max Yang (Georgia Institute of Technology)

Cluster-to-user data transfers present challenges with cross-platform endianness (byte-order) compatibility and handling a variety of numeric types, and may occur over suboptimal network links. Two serialization libraries, Protocol Buffers and Conduit, were selected for their ability to handle endianness and their cross-language support, and their performance in both size and speed was measured. It was found that the throughput of Protocol Buffers was significantly more than that of Conduit while exhibiting less protocol overhead. Adding a compression stage after serialization dramatically reduced the size of messages on certain types of data, but had some impact on throughput.

A26: Co-Designing MPI Runtimes and Deep Learning Frameworks for Scalable Distributed Training on GPU Clusters

Authors: Ammar Ahmad Awan (Ohio State University)

Deep Learning frameworks like Caffe, TensorFlow, and CNTK have brought forward new requirements and challenges for communication runtimes like MVAPICH2-GDR. These include support for low-latency and high-bandwidth communication of very-large GPU-resident buffers. This support is essential to enable scalable distributed training of Deep Neural Networks on GPU clusters. However, current MPI runtimes have limited support for large-message GPU-based collectives. To address this, we propose the S-Caffe framework; a co-design of distributed training in Caffe and large-message collectives in MVAPICH2-GDR. We highlight two designs for MPI_Bcast, one that exploits NVIDIA NCCL and the other that exploits ring-based algorithms. Further, we present designs for MPI_Reduce that provide up-to 2.5X improvement. We also present layer-wise gradient aggregation designs in S-Caffe that exploit overlap of computation and communication as well as the proposed reduce design. S-Caffe provides a scale-out to 160 GPUs for GoogLeNet training and delivers performance comparable to CNTK for AlexNet training.

A27: High-Performance and Scalable Broadcast Schemes for Deep Learning on GPU Clusters

Authors: *Ching-Hsiang Chu (Ohio State University)*

Broadcast operations are a widely used operation in many streaming and deep learning applications to disseminate large amounts of data on emerging heterogeneous High-Performance Computing (HPC) systems. Further, traditional broadcast schemes are not well optimized for upcoming large-scale Graphics Processing Unit (GPU)-based systems. However, utilizing cutting-edge features of modern HPC technologies such like InfiniBand (IB) and NVIDIA GPUs to enable scalable heterogeneous broadcast operations remains an open challenge.

Toward delivering the best performance for streaming and deep learning workloads, we propose high-performance and scalable broadcast schemes that exploit IB hardware multicast (IB-MCAST) and NVIDIA GPUDirect technology. We present experimental results and find that they indicate improved scalability and up to 68% reduction of latency compared to the state-of-the-art solutions in the benchmark-level evaluation. Furthermore, the proposed design yields up to 24% performance improvement for the popular deep learning framework, Microsoft cognitive toolkit (CNTK), with no application changes.

A28: Exploring Use Cases for Non-Volatile Memories in Support of HPC Resilience

Authors: *Onkar Patil (North Carolina State University)*

Improving resilience and creating resilient architectures is one of the major goals of exascale computing. With the advent of Non-volatile memory technologies, memory architectures with persistent memory regions will be a significant part of future architectures. There is potential to use them in more than one way to benefit different applications. We look to take advantage of this technology to enable more fine-grained and novel methodology that will improve resilience and efficiency of exascale applications. We have developed three modes of memory usage for persistent memory to enable efficient checkpointing in HPC applications. We have developed a simple API that is evaluated with the DGEMM benchmark on a 16-node cluster with independent SSDs on every node. Our aim is to build on this work and enable static and dynamic runtime systems that will inherently make the HPC applications more fault-tolerant and resistant to errors.

P01: Cache-Blocking Tiling of Large Stencil Codes at Runtime

Authors: *Istvan Z. Reguly (Pazmany Peter Catholic University), Gihan R. Mudalige (University of Warwick), Mike B. Giles (University of Oxford)*

Stencil codes on structured meshes are well-known to be bound by memory bandwidth. Previous research has shown that compiler techniques that reorder loop schedules to improve temporal locality across loop nests, such as tiling, work particularly well. However in large codes the scope of such analysis is limited by the large number of code paths, compilation units, and run-time parameters. We present how, through run-time analysis of data dependencies across stencil loops enables the OPS domain specific language to tile across a large number of different loops. This lets us tackle much larger applications than previously studied: we demonstrate 1.7-3.5x performance improvement on CloverLeaf 2D, CloverLeaf 3D, TeaLeaf and OpenSBLI, tiling across up to 650 subsequent loopnests accessing up to 30 different state variables per gridpoint with up to 46 different stencils. We also demonstrate excellent strong and weak scalability of our approach on up to 4608 Broadwell cores.

P02: Strassen's Algorithm for Tensor Contraction

Authors: *Jianyu Huang (University of Texas), Devin A. Matthews (University of Texas), Robert A. van de Geijn (University of Texas)*

Tensor contraction(TC) is an important computational kernel widely used in numerous applications. It is a multi-dimensional generalization of matrix multiplication(GEMM). While Strassen's algorithm for GEMM is well studied in theory and practice, extending it to accelerate TC has not been previously pursued. Thus, we believe this to be the first work to demonstrate how one can in practice speed up tensor contraction with Strassen's algorithm. By adopting a Block-Scatter-Matrix format, a novel matrix-centric tensor layout, we can conceptually view TC as GEMM for general stride storage, with an implicit tensor-to-matrix transformation. This insight enables us to tailor a recent state-of-the-art implementation of Strassen's algorithm to TC, avoiding explicit transpositions(permutations) and extra workspace, and reducing the overhead of memory movement that is incurred. Performance benefits are demonstrated with a performance model as well as in practice on modern single core, multicore, and distributed memory parallel architectures, achieving up to 1.3x speedup.

P03: BEM4I: A Massively Parallel Boundary Element Solver

Authors: *Michal Merta (Technical University of Ostrava), Jan Zapletal (Technical University of Ostrava), Michal Kravcenko*

(Technical University of Ostrava), Lukas Maly (Technical University of Ostrava)

In this work we present a library of parallel solvers based on the boundary element method (BEM). We provide a brief description of BEM and its parallelization, focus on SIMD vectorization and shared- and distributed-memory parallelization by OpenMP and MPI, respectively. Two approaches for distributed parallelization of BEM are discussed - the first one based on a novel parallel adaptive cross approximation (ACA) method, the second one on the boundary element tearing and interconnecting (BETI) domain decomposition method. To demonstrate the efficiency of the library we provide results of numerical experiments on the Xeon and Xeon Phi based clusters.

BP

P04: Unstructured-Grid CFD Algorithms on Many-Core Architectures

Authors: Aaron Walden (NASA Langley Research Center), Eric J. Nielsen (NASA Langley Research Center), Mohammad Zubair (Old Dominion University), John C. Linford (ParaTools), John G. Wohlbiel (Engility Corporation), Justin P. Luitjens (Nvidia Corporation), Jason Orender (Old Dominion University), Izaak Beekman (ParaTools), Samuel Khuvis (ParaTools), Sameer S. Shende (ParaTools)

In the field of computational fluid dynamics (CFD), the Navier-Stokes equations are often solved using an unstructured-grid approach to accommodate geometric complexity. Furthermore, turbulent flows encountered in aerospace applications generally require highly anisotropic meshes, driving the need for implicit solution methodologies to efficiently solve the discrete equations. These approaches require frequent construction and solution of large, tightly-coupled systems of block-sparse linear equations.

We explore the transition of two representative CFD kernels from a coarse-grained MPI-based model originally developed for multi-core systems to a shared-memory model suitable for many-core platforms. Results for the Intel Xeon Phi Knights Landing, NVIDIA Pascal P100, and NVIDIA Volta V100 architectures are compared with the aforementioned MPI-based implementation for the multi-core Intel Xeon Broadwell (BWL) processor. We observe substantial speedups over BWL as well as higher performance per dollar MSRP and performance per watt for the many-core architectures.

P05: ooc_cuDNN : A Deep Learning Library Supporting CNNs over GPU Memory Capacity

Authors: Yuki Ito (Tokyo Institute of Technology), Ryo Matsumiya (Tokyo Institute of Technology), Toshio Endo (Tokyo Institute of Technology)

GPUs are widely used to accelerate deep learning with convolutional neural network (CNN). However, since GPU memory capacity is limited, it is difficult to implement efficient programs that compute large CNN on GPU. This poster describes the design and implementation of out-of-core cuDNN (ooc_cuDNN) library, which supports to compute CNN exceeding GPU memory capacity using capacity of CPU memory. ooc_cuDNN is an extension of cuDNN, which is high performance and popular deep learning library. ooc_cuDNN divides CNN computation based on its performance model for better performance. In addition, ooc_cuDNN provides fused functions combined some computation to reduce extra communication. With ooc_cuDNN, we successfully computed CNN requiring more than 60 GB memory on a single GPU with 16 GB memory. Compared with an in-core case using cuDNN, performance degradation was 13 %.

P06: Large Scale FFT-Based Stress-Strain Simulations with Irregular Domain Decomposition

Authors: Anuva Kulkarni (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University), Jelena Kovacevic (Carnegie Mellon University)

Large-scale stress-strain simulations involving parallel Fast Fourier Transforms (FFTs) suffer from high memory requirements and high communication overhead. We propose an irregular domain decomposition method to reduce the memory requirement of an FFT-based stress-strain simulation algorithm for composite materials, the Moulinec-Suquet Composite (MSC) - Basic Scheme. This algorithm uses Green's functions to solve a partial differential equation. FFTs are used for convolution of large 3-D tensor fields with the Green's function.

In this preliminary work, we propose a modified algorithm, the MSC-Alternate Scheme, to show that processing the composite with smaller, local FFTs on irregular domains (grains in the material's microstructure) can reduce memory usage without adversely impacting accuracy of the result. Additionally, data models can reduce communication by compressing the data in the domains before the communication step. Our poster presents our proof-of-concept results and charts out the path towards a GPU implementation.

P07: PORTAGE - A Flexible Conservative Remapping Framework for Modern HPC Architectures

Authors: Rao V. Garimella (Los Alamos National Laboratory), Peter D. Crossman (Los Alamos National Laboratory), Gary A. Dilts (Los Alamos National Laboratory), Rajeev S. Erramilli (Los Alamos National Laboratory), Charles R. Ferenbaugh (Los Alamos National Laboratory), Angela M. Herring (Los Alamos National Laboratory), Eugene Kikinon (Los Alamos National Laboratory), Chris M. Malone (Los Alamos National Laboratory), Navamita Ray (Los Alamos National Laboratory), Mike L. Rogers (Los Alamos National Laboratory)

Portage is a massively parallel remapping framework to transfer fields between general polyhedral meshes while conserving integral quantities of interest. The framework also has the capability to remap data between two point clouds. Portage is templated on the component classes required in conservative remapping - search, intersection and interpolation as well as on the mesh and field managers. Applications supply Portage with custom components while the framework takes care of distributed parallelism

using MPI and threaded parallelism using NVIDIA Thrust to scale to many thousands of cores. Moreover, the imposition of a functional design on the components used by Portage makes it very amenable to achieve task parallelism with runtime systems such as Legion. Portage has been tested in 2D/3D for remapping between general polygonal and polyhedral meshes and between point clouds. We present scaling results for distributed (MPI) and on-node parallelism (OpenMP) on LANL's HPC machines.

P08: Performance Optimization of Matrix-free Finite-Element Algorithms within deal.II

Authors: Martin Kronbichler (Technical University Munich), Karl Ljungkvist (Uppsala University), Momme Allalen (Leibniz Supercomputing Centre), Martin Ohlerich (Leibniz Supercomputing Centre), Igor Pasichnyk (IBM), Wolfgang A. Wall (Technical University Munich)

We present a performance comparison of highly tuned matrix-free finite element kernels from the deal.II finite element library on three contemporary computer architectures, an NVIDIA P100 GPU, an Intel Knights Landing Xeon Phi, and two multi-core Intel CPUs. The algorithms are based on fast integration on hexahedra using sum factorization techniques. On Cartesian meshes with a relatively high arithmetic intensity, the four architectures provide a surprisingly similar computational throughput. On curved meshes, the kernel is heavily memory bandwidth limited which reveals distinct differences between the architectures: the P100 is twice as fast as KNL, and almost four times as fast as the Haswell and Broadwell CPUs, effectively leveraging the higher memory bandwidth and the favorable shared memory programming model on the GPU.

P09: Adaptive Multistep Predictor for Accelerating Dynamic Implicit Finite-Element Simulations

Authors: Kohei Fujita (University of Tokyo, RIKEN), Tsuyoshi Ichimura (University of Tokyo, RIKEN), Masashi Horikoshi (Intel Corporation), Muneo Hori (University of Tokyo, RIKEN), Lalith Maddegadara (University of Tokyo, RIKEN)

We develop an adaptive multistep predictor for accelerating memory bandwidth-bound dynamic implicit finite-element simulations. We predict the solutions for future time steps adaptively using highly-efficient matrix-vector product kernels with multiple right-hand sides to reduce the number of iterations required in the solver. By applying the method to a conjugate gradient solver with 3×3 block Jacobi preconditioning, we were able to achieve a 42% speedup on a Skylake-SP Xeon Gold cluster for a typical earthquake ground motion problem. As the method enables the number of iterations, and thus the communication frequency, to be reduced, the developed solver was able to attain high size-up scalability: 80.6% up to 32,768 compute nodes on the K computer. The developed predictor can also be applied to other iterative solvers and is thus expected to be useful for wide range of dynamic implicit finite-element simulations.

P10: HiCOO: A Hierarchical Sparse Tensor Format for Tensor Decompositions

Authors: Jiajia Li (Georgia Institute of Technology), Jimeng Sun (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

This paper proposes a new Hierarchical COOrdinate (HiCOO) format for sparse tensors, which compresses its indices to units of sparse tensor blocks. HiCOO format does not favor one tensor mode over the others, thus can be used as a replacement of the traditional COOrdinate (COO) format. In this paper, we use HiCOO format for the Matriced Tensor Times Khatri-Rao Product (MTTKRP) operation, the most expensive computational core in the popular CANDECOMP/PARAFAC decomposition, then accelerate it on multicore CPU architecture using two parallel strategies for irregular shaped tensors. Parallel MTTKRP using HiCOO format achieves up to $3.5\times$ ($2.0\times$ on average) speedup over COO format and up to $4.3\times$ ($2.2\times$ on average) speedup over CSF format.

P11: Energy-Efficient Transprecision Techniques for Iterative Refinement

Authors: JunKyu Lee (Queen's University Belfast), Hans Vandierendonck (Queen's University Belfast), Dimitrios S. Nikolopoulos (Queen's University Belfast)

This paper presents transprecision techniques for iterative refinement, which utilize various precision arithmetic dynamically according to numeric properties of the algorithm and computational latencies depending on precisions. The transprecision techniques were plugged into a mixed precision iterative refinement on an Intel Xeon E5-2650 2GHz core with MKL 2017 and XBLAS 1.0. The transprecision techniques brought further 2.0-3.4X speedups and 3.0-4.1X energy reductions to a mixed precision iterative refinement when double precision solution accuracy was required for forward error and a matrix size was ranged from 4K to 32K.

P12: Multi-Size Optional Offline Caching Algorithms

Authors: Andrew Y. Choliy (Rutgers University), Max D. Whitmore (Brandeis University), Gruia Calinescu (Illinois Institute of Technology)

The optional offline caching (paging) problem, where all future file requests are known, is a variant of the heavily studied online caching problem. This offline problem has applications in web caching and distributed storage systems. Given a set of unique files with varying sizes, a series of requests for these files, fast cache memory of limited size, and slow main memory, an efficient replacement policy is necessary to decide when it is best to evict some file(s) from the cache in favor of another. It is known that this problem is NP-complete, and few approximation algorithms have been proposed. We propose three new heuristics, as well as a 4-approximation algorithm. We then evaluate each algorithm by the metrics of runtime complexity and proximity to the optimal solutions of many synthetic data sets.

P13: Large-Scale GW Calculations on Pre-Exascale HPC Systems

Authors: Mauro Del Ben (Lawrence Berkeley National Laboratory), Felipe H. da Jornada (University of California, Berkeley), Andrew Canning (Lawrence Berkeley National Laboratory), Nathan Wichmann (Cray Inc), Karthik Raman (Intel Corporation), Ruchira Sasanka (Intel Corporation), Chao Yang (Lawrence Berkeley National Laboratory), Steven G. Louie (Lawrence Berkeley National Laboratory; University of California, Berkeley)

The accurate determination of excitation spectra of materials, such as the electronic band gap, is critical for the design of novel devices, including photovoltaics, transistors, batteries, and LEDs. Many-body perturbation-theory methods, and the ab-initio GW approach in particular, have emerged over the last decades as the gold standard for computing these quantities. However, the ab-initio GW formalism is often limited to systems of at most 100 atoms due to its computational complexity. We present here large scale GW calculations of crystalline defect problems, relevant for the performance of semiconductors, with up to 1000 atoms, on the Cori system at NERSC. We show that the GW method is particularly well suited for exascale/pre-exascale systems. Our implementation, which uses a combination of new algorithms and optimizations targeted at many-core CPU architectures, scales well to the entire Cori system, and obtains a significant fraction of peak performance.

P14: Robust SA-AMG Solver by Extraction of Near-Kernel Vectors

Authors: Naoya Nomura (University of Tokyo), Kengo Nakajima (University of Tokyo), Akihiro Fujii (Kogakuin University)

The smoothed aggregation algebraic multigrid (SA-AMG) method is among the fastest solvers for large-scale linear equations, $Ax=b$. The SA-AMG method achieves good convergence and scalability by damping various wavelength components efficiently. To achieve this damping, this method creates multi-level matrices which are hierarchically smaller in dimension than the original matrix. Moreover, the convergence can be further improved by setting near-kernel vectors p , which satisfy $Ap=0$ and $p \neq 0$. Generally, the same number of near-kernel vectors are used at each level. In the present work, we propose a method that extracts and adds near-kernel vectors at each level. We evaluate the performance of the solver that extracts the near-kernel vectors and adds them at each level. We use the three-dimensional elastic problem and employ up to 512 processes on the FX10 supercomputer system. By using this method, the performance is improved compared with previous work.

P15: Toward Decoupling the Selection of Compression Algorithms from Quality Constraints

Authors: Julian Kunkel (German Climate Computing Center), Anastasiia Novikova (University of Hamburg), Eugen Betke (German Climate Computing Center)

With the Scientific Compression Library (SCIL), we are developing a meta-compressor that allows users to set various quantities that define the acceptable error and the expected performance behavior. The library then chooses the appropriate chain of algorithms to yield the users requirements. This approach is a crucial step towards a scientifically safe use of much-needed lossy data compression, because it disentangles the tasks of determining scientific ground characteristics of tolerable noise, from the task of determining an optimal compression strategy given target noise levels and constraints. Without changing applications, it allows these codes to utilize future algorithms once they are integrated into the library.

P16: Scaling Analysis of a Hierarchical Parallelization of Large Inverse Multiple-Scattering Solutions

Authors: Mert Hidayetoglu (University of Illinois), Carl Pearson (University of Illinois), Izzat El-Hajj (University of Illinois), Weng Cho Chew (University of Illinois), Levent Gurel (University of Illinois), Wen-Mei Hwu (University of Illinois)

We propose a hierarchical parallelization strategy to improve the scalability of inverse multiple-scattering solutions. The inverse solver parallelizes the independent forward solutions corresponding to different illuminations. For further scaling out on large numbers of computing nodes, each forward solver parallelizes the dense and large matrix-vector multiplications accelerated by the multilevel fast multipole algorithm. An inverse problem involving a large Shepp-Logan phantom is solved on up to 1,024 CPU nodes of the Blue Waters supercomputer in order to demonstrate the strong-scaling efficiency of the proposed parallelization scheme. The results show that parallelizing illuminations has almost perfect scaling efficiency of 95% because of the independent nature of forward-scattering solutions, however, parallelization of MLFMA has 73% efficiency due to MPI communications in MLFMA multiplications. Nevertheless, the proposed strategy improves granularity and allows spreading DBIM solutions on large numbers of nodes.

P17: Fully Non-Blocking Communication-Computation Overlap Using Assistant Cores toward Exascale Computing

Authors: Motoki Nakata (National Institute for Fusion Science), Masanori Nunami (National Institute for Fusion Science), Shinsuke Satake (National Institute for Fusion Science), Yoshihiro Kasai (Fujitsu Ltd), Shinya Maeyama (Nagoya University), Tomo-Hiko Watanabe (Nagoya University), Yasuhiro Idomura (Japan Atomic Energy Agency)

A fully non-blocking optimized Communication-Computation overlap technique using assistant cores (AC), which are independent from the calculation cores, is proposed for the application to the five-dimensional plasma turbulence simulation code with spectral (FFT) and finite-difference schemes, toward exascale supercomputing. The effects of optimization are examined in Fujitsu FX100 (2.62PFlop/s) with 32 ordinary cores and 2 Assistant cores/node, where AC enables us to employ the fully non-blocking MPI communications overlapped by the thread-parallelized calculations with OpenMP Static scheduling with much less overheads. It is clarified that the combination of the non-blocking communications by AC and the static scheduling leads to not only reduction in OpenMP overhead, but also improved load/store and cash performance, where about 22.5% improved numerical performance is

confirmed in comparison to the conventional overlap by the master thread communications with dynamic scheduling.

P18: A Parallel Python Implementation of BLAST+ (PPIB) for Characterization of Complex Microbial Consortia

Authors: Amina Jackson (Naval Research Laboratory), William Connor Horne (Naval Research Laboratory), Daniel Beall (Naval Research Laboratory), Kenneth Jiang (Naval Research Laboratory), William Judson Hervey (Naval Research Laboratory)

Technological advancements in analytical instrumentation have enabled large-scale data acquisitions among the ‘-omics’ sciences of genomics, transcriptomics, and proteomics. An essential application among ‘-omics’ disciplines is the Basic Local Alignment Search Tool (BLAST) for functional inference of biomolecules. Though implementations of BLAST+ have been modified to address data volume growth, such improvements have neither been consistently maintained for high performance computing (HPC), nor have they been applied to complex microbiomes. Further, such implementations do not scale well to microbiomes of Naval interest on HPC systems in our hands.

Here, we compare 2 HPC implementations: BLAST+ and a Parallel Python Implementation of BLAST+ (PPIB) for protein functional inference. PPIB enabled functional inference of 2 complex microbiomes, which may be attributed to a combination of MPI and Python multiprocessing to query up to 3,600 proteins simultaneously. In contrast, BLAST+ did not complete functional assignments relative to PPIB at a comparable walltime.

P19: MPI-GIS: An MPI System for Big Spatial Data

Authors: Satish Puri (Marquette University)

In recent times, geospatial datasets are growing in terms of size, complexity and heterogeneity. High performance systems are needed to analyze such data to produce actionable insights in an efficient manner. For polygonal a.k.a vector datasets, operations such as I/O, data partitioning, and communication becomes challenging in a cluster environment.

In this work, we present MPI-GIS equipped with MPI-Vector-IO, a parallel I/O library that we have designed using MPI-IO specifically for irregular polygonal (vector) data formats such as Well Known Text, XML, etc. Our system can perform spatial in-memory indexing and join efficiently for an order of magnitude larger datasets compared to our previous work. It makes MPI aware of spatial data and spatial primitives and provides support for spatial data types embedded within collective computation and communication using MPI message-passing library. It takes less than 2 minutes to scan through 2.7 billion geometries in 96GB file using 160 processes.

P20: Facilitating the Scalability of ParSplice for Exascale Testbeds

Authors: Vinay B. Ramakrishnaiah (University of Wyoming), Jonas L. Landsgesell (University of Stuttgart), Ying Zhou (Loughborough University), Iris Linck (University of Colorado, Denver), Mouad Ramil (National School of Bridges and Roads - ParisTech), Joshua Bevan (University of Illinois), Danny Perez (Los Alamos National Laboratory), Louis J. Vernon (Los Alamos National Laboratory), Thomas D. Swinburne (Los Alamos National Laboratory), Robert S. Pavel (Los Alamos National Laboratory), Christoph Junghans (Los Alamos National Laboratory)

Parallel trajectory splicing (ParSplice) is an attempt to solve the enduring challenge of simulating the evolution of materials over long time scales for complex atomistic systems. A novel version of ParSplice is introduced with features that could be useful in its scaling to exascale architectures. A two-pronged approach is used. First, latent parallelism is exploited by extending support to heterogeneous architectures, including GPUs and KNLs. Second, the efficiency of the Kinetic Monte Carlo predictor is improved, allowing enhanced parallel speculative execution. The key idea in these predictor modifications is to include statistics from higher temperature simulations. The issue of inherent uncertainty in the prediction model was addressed in order to improve the performance, as the current predictor only takes into account the previous observations to formulate the problem. The predictor was also improved by using a hybrid approach of message-passing + multi-threading. (LA-UR-17-26181)

P21: The First Real-Scale DEM Simulation of a Sandbox Experiment Using 2.4 Billion Particles

Authors: Mikito Furuichi (Japan Agency for Marine-Earth Science and Technology), Daisuke Nishiura (Japan Agency for Marine-Earth Science and Technology), Mitsuteru Asai (Kyushu University), Takane Hori (Japan Agency for Marine-Earth Science and Technology)

A novel implementation of the Discrete Element Method (DEM) for a large parallel computer system is presented to simulate a sandbox experiment with realistic particle sizes. To save memory in the pairwise tangential forces and halve the arithmetic costs, interactions are calculated using the action-reaction law. An iterative load-balancer the flexible 2D orthogonal domain decomposition is applied to overcome the load-imbalance problem caused by the Lagrangian nature of DEM. An overlapping communication technique combined with cell-ordering with space-filling curves is also applied to hide the overhead cost because of the MPI communication tasks. We verify our complex parallel implementation with the action-reaction law via a reproducibility test. The parallel scaling test shows good strong, and weak scalabilities up to 2.4 billion particles on the Earth Simulator and the K computer. The world's first real-scaled numerical sandbox simulation successfully captures the characteristics of real observations.

P22: Numerical Simulation of Snow Accretion by Airflow Simulator and Particle Simulator

Authors: Kohei Murotani (Railway Technical Research Institute), Koji Nakade (Railway Technical Research Institute), Yasushi Kamata (Railway Technical Research Institute)

In this research, to take countermeasures for the snow accretion damage, we developed a simulator of realizing the snow accretion process in the following steps. Firstly, air flow analysis is performed by "Airflow simulator" developed by RTRI (Railway Technical Research Institute). Secondly, trajectory of flying snow is calculated by Basset-Boussinesq-Oseen equation using distributed velocity of air flow. Thirdly, snow accretion analysis is performed by "Particle simulator" developed by RTRI. The shape modified by snow accretion is reflected onto the boundary conditions of the air flow analysis. In this year, snow accretion analysis for simple cubic shapes is performed in order to aim at system development and validation.

P23: AI with Super-Computed Data for Monte Carlo Earthquake Hazard Classification

Authors: Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Muneo Hori (University of Tokyo, RIKEN), Madgededara Lalith (University of Tokyo, RIKEN), Naonori Ueda (RIKEN)

Many problems associated with earthquakes are yet to be solved using heroic computing, which is defined as computing at the largest scale possible using the best supercomputers and algorithms. Thus, a continuous effort has been pursued in HPC to solve these problems. However, even when heroic computing is applied, its practical use is difficult without considering the uncertainties in models. In this study, we constructed an AI methodology that uses super-computed data generated using heroic computing. We applied this AI to an earthquake hazard classification including uncertainty analyses in order to demonstrate its utility. This study can be regarded as an innovative step towards realizing high quality computing for Earthquakes by exploiting the potential of HPC through AI.

BP

P24: A Deployment of HPC Algorithm into Pre/Post-Processing for Industrial CFD on K-Computer

Authors: Keiji Onishi (RIKEN), Niclas Jansson (KTH Royal Institute of Technology), Rahul Bale (RIKEN), Wei-Hsiang Wang (RIKEN), Chung-Gang Li (Kobe University, RIKEN), Makoto Tsubokura (Kobe University, RIKEN)

Pre- and post-processing is still a major problem in industrial computational fluid dynamics (CFD). With the rapid development of computers, physical solvers are getting faster, while pre- remains slow because it's mainly a serial process. A methodology using MPI+OpenMP hybrid parallelization has been proposed to eliminate the manual work required during pre-processing for correcting the surface imperfections of CAD data. Compared to the rapidly increasing amount of data in recent years, the speed-up of visualization is insufficient. We address this limitation of post- by adapting the in-situ visualization to parallelize the post-processing using libsim (Visit) library. The performance of pre-/post- processing is investigated in this work, and we show that the pre-processing time has been reduced from several days in the conventional framework to order of minutes. The post-processing time has been reduced seconds order per frame, and approximately 30% increase of computational time was observed in vehicle aerodynamics cases.

P25: Large-Scale Adaptive Mesh Simulations Through Non-Volatile Byte-Addressable Memory

Authors: Bao Nguyen (Washington State University, Vancouver), Hua Tan (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

Octree-based mesh adaptation has enabled simulations of complex physical phenomena. Existing meshing algorithms were proposed with the assumption that computer memory is volatile. Consequently, for failure recovery, the in-core algorithms need to save memory states as snapshots with slow file I/Os. The out-of-core algorithms store octants on disks for persistence. However, neither of them was designed to leverage unique characteristics of non-volatile byte-addressable memory (NVBM). We propose a novel data structure Persistent Merged octree (PM-octree) for both meshing and in-memory storage of persistent octrees using NVBM. It is a multi-version data structure and can recover from failures using its earlier persistent version stored in NVBM. In addition, we design a feature-directed sampling approach to help dynamically transform the PM-octree layout for reducing NVBM-induced memory write latency.

P26: Optimizing Gravity and Nuclear Physics in FLASH for Exascale

Authors: Hannah Klion (University of California, Berkeley; Oak Ridge National Laboratory), Bronson Messer (Oak Ridge National Laboratory, University of Tennessee), J. Austin Harris (Oak Ridge National Laboratory), Thomas Papatheodore (Oak Ridge National Laboratory)

In a Type Ia supernova, runaway fusion ignites in a white dwarf, causing it to explode. The heavy element yields of these events remain uncertain, and high-performance multiphysics simulations with tools like FLASH are critical for our understanding. Current simulations track approximately a dozen nuclear isotopes, as opposed to the thousands required to completely capture the event's nuclear physics.

Simulating nuclear physics and self-gravity accurately and efficiently is critical for modeling a Type Ia supernova, since supernovae are competitions between energy-releasing nuclear reactions and gravity. Currently, the FLASH nuclear reaction network and self-gravity solver requires substantial inter-node communication. We use non-blocking MPI collectives to overlap communication in the self-gravity calculation with the computation-heavy nuclear burning calculation. We find that speedups from this technique are possible, but are MPI implementation-dependent. We highlight some of the challenges associated with this type of optimization.

P27: Parallelization of the Particle-In-Cell Monte Carlo Collision (PIC-MCC) Algorithm for Plasma Simulation on Intel MIC

Xeon Phi Architecture

Authors: Keval Shah (*Dhirubhai Ambani Institute of Information and Communication Technology*), Anusha Phadnis (*Dhirubhai Ambani Institute of Information and Communication Technology*), Miral Shah (*Dhirubhai Ambani Institute of Information and Communication Technology*), Bhaskar Chaudhury (*Dhirubhai Ambani Institute of Information and Communication Technology*)

The implementation of 2D-3v (2D in space and 3D in velocity space) PIC-MCC (Particle-In-Cell Monte Carlo Collision) method involves the computational solution of Vlasov-Poisson equations. This provides the spatial and temporal evolution of the charged-particle velocity distribution functions in plasma under the effect of self-consistent electromagnetic fields and collisions. Stringent numerical constraints associated with the PIC code makes it computationally prohibitive on CPU.

In our work, parallelization and optimization techniques have been extended to this simulation, along with a novel approach that involves developing a 'self-aware' code that triggers sorting in order to maintain cache-coherence while reducing the total sorting time during iterations.

We present the effect of important numerical parameters on speed-up. Finally, we compare the scalability and performance of the parallelization and optimization strategies on Intel® Xeon™ E5-2630, Xeon Phi™ 5110p and Xeon Phi™ 7250 relative to a serial implementation on Intel® i5.

P28: High-Fidelity Blade-Resolved Wind Plant Modeling

Authors: Andrew C. Kirby (*University of Wyoming*), Zhi Yang (*University of Wyoming*), Michael J. Brazell (*University of Wyoming*), Behzad R. Ahrabi (*University of Wyoming*), Jay Sitaraman (*Parallel Geometric Algorithms LLC*), Dimitri J. Mavriplis (*University of Wyoming*)

Blade-resolved numerical simulations of wind energy applications using full blade and tower models are presented. The computational methodology combines solution technologies in a multi-mesh, multi-solver paradigm through a dynamic overset framework. The coupling of a finite-volume solver and a high-order, hp-adaptive finite-element solver is utilized. Additional technologies including in-situ visualization and atmospheric micro-scale modeling are incorporated into the analysis environment. Validation of the computational framework is performed on the NREL 5MW wind turbine, the unsteady aerodynamics experimental NREL Phase VI turbine, and the Siemens SWT-2.3-93 wind turbine. The power and thrust results of all single turbine simulations agree well with low-fidelity model simulation results and field experiments when available. Scalability of the computational framework is demonstrated using 6, 12, 24, 48, and 96 wind turbine wind plant set-ups including the 48 turbine wind plant known as Lillgrund. Demonstration of the coupling of atmospheric micro-scale and CFD solvers is presented.

P29: A Deep Learning Tool for Fast Simulation

Authors: Sofia Vallecorsa (*CERN*), Gul rukh Khattak (*CERN*), Shruti Sharan (*CERN*)

We present the first application of Volumetric Generative Adversarial Network (VGAN) to High Energy Physics simulation. We generate three dimensional images of particles depositing energy in calorimeters. This is the first time such an approach is taken in HEP where most of data is three dimensional in nature but it is customary to convert it into two dimensional slices. The volumetric approach leads to a larger number of parameters, but two dimensional slicing loses the volumetric dependencies inherent in the dataset. The present work proves the success of handling those dependencies through VGANs. Energy showers are faithfully reproduced in all dimensions and show a reasonable agreement with standard techniques. We also demonstrate the ability to condition training on several parameters such as particle type and energy. This work aims at proving Deep Learning techniques represent a valid fast alternative to standard MonteCarlo approaches and is part of the GEANTV project.

P30: MPI/OpenMP Parallelization of the Hartree-Fock Method for the Second Generation Intel Xeon Phi

Authors: Kristopher Keipert (*Iowa State University*), Vladimir Mironov (*Lomonosov Moscow State University*), Yuri Alexeev (*Argonne National Laboratory*), Michael D'mello (*Intel Corporation*), Alexander Moskovsky (*RSC Technologies*), Mark S. Gordon (*Iowa State University*)

Replication of critical data structures in the MPI-only GAMESS Hartree-Fock algorithm limits the full utilization of the manycore Intel Xeon Phi processor. In this work, modern OpenMP threading techniques are used to implement hybrid MPI/OpenMP algorithms. Two separate implementations that differ by the sharing and replication details of key data structures among threads are considered. The hybrid MPI/OpenMP implementations reduce the memory footprint by approximately 200 times compared to the legacy code. The MPI/OpenMP code was shown to run up to six times faster than the original for a range of molecular system sizes. The implementation details and strategies will be presented for both hybrid algorithms. Benchmark scaling results utilizing up to 3000 Intel Xeon Phi processors will also be discussed.

BP

P31: Understanding the Performance of Small Convolution Operations for CNN on Intel Architecture

Authors: Alexander Heinecke (*Intel Corporation*), Evangelos Georganas (*Intel Corporation*), Kunal Banerjee (*Intel Corporation*), Dhiraj Kalmakar (*Intel Corporation*), Narayanan Sundaram (*Intel Corporation*), Anand Venkat (*Intel Corporation*), Greg Henry (*Intel Corporation*), Hans Pabst (*Intel Corporation*)

Convolution layers are prevalent in many classes of deep neural networks, including Convolutional Neural Networks (CNNs) which provide state-of-the-art results for tasks like image recognition, natural language processing, and speech recognition. The computationally expensive nature of a convolution operation has led to the proliferation of implementations including matrix-matrix

multiplication formulation, FFT-formulation, Winograd transformation, and direct convolution primarily targeting GPUs. In this paper, we optimize a direct convolution and Winograd implementation for x86 architectures, in particular for Xeon Phi systems, via a dynamic compilation approach. We then show how these JIT optimizations can be integrated in a high-level domain-specific language setting. We shed light on what is possible and what is not possible based on different data-formats and blocking techniques. Our JIT-based Ninja implementation shows close to theoretical peak results on modern x86 architectures, depending on setting and the CPU architecture at hand.

P32: Exploring the Performance of Electron Correlation Method Implementations on Kove XPDs

Authors: Colleen Bertoni (Argonne National Laboratory), Brian Toonen (Argonne National Laboratory), William Allcock (Argonne National Laboratory), Spencer R. Pruitt (Worcester Polytechnic Institute), Mark S. Gordon (Iowa State University)

In electron correlation methods in quantum chemistry, there are often high memory requirements which can reach terabytes for medium-sized molecular systems. For second-order perturbation theory (MP2), the two-electron integral arrays are the main memory bottleneck. Previously the two-electron integrals were recomputed, stored in distributed memory, or stored on disk. A way of storing the arrays which would remove the dependence on compute node memory and large latency associated with using disk is by using an external memory appliance, like Kove's XPD.

In this work, we modified a distributed memory implementation of MP2 to use XPDs instead of distributed memory. We evaluated the performance of our implementation against the distributed memory version for several molecular systems by considering scaling behavior with respect to compute processes and connections to XPDs. In the poster, we present an outline of the MP2 implementation using XPDs and the scaling results.

P33: Massively Parallel Evolutionary Computation for Empowering Electoral Reform: Quantifying Gerrymandering via Multi-objective Optimization and Statistical Analysis

Authors: Wendy K. Cho (National Center for Supercomputing Applications, University of Illinois), Yan Liu (National Center for Supercomputing Applications, University of Illinois)

Important insights into redistricting can be gained by formulating and analyzing the problem within a large-scale optimization framework. Redistricting is an application of the set-partitioning problem that is NP-hard. We design a hybrid metaheuristic as the base search algorithm. With our grant on the Blue Waters supercomputer, we extend our algorithm to the high-performance-computing realm by using MPI to implement an asynchronous processor communication framework. We experimentally demonstrate the effectiveness of our algorithm to utilize multiple processors and to scale to 131,072 processors. The massive computing power allows us to extract new substantive insights that closely mesh with the framework that the Supreme Court has elucidated for electoral reform.

P34: GPU Acceleration for the Impurity Solver in GW+DMFT Packages

Authors: Kwangmin Yu (Brookhaven National Laboratory), Patrick Semon (Brookhaven National Laboratory), Nicholas D'Imperio (Brookhaven National Laboratory)

The combination of dynamical mean field theory (DMFT) and GW (or density functional theory) has become a powerful tool to study and predict properties of real materials with strongly correlated electrons, such as high temperature superconductors. At the core of this combined theory lies the solution of a quantum impurity model, and continuous-time quantum Monte Carlo (CT-QMC) has proven an indispensable algorithm in this respect. However, depending on the material, this algorithm is computationally very expensive, and enhancements are crucial for bringing new materials within reach of GW+DMFT. Based on a CPU implementation, GPU acceleration is added, and two times speedup is achieved. New techniques are invented and implemented to deal with various GPU acceleration environment.

P36: A Novel Feature-Preserving Spatial Mapping for Deep Learning Classification of Ras Structures

Authors: Thomas Corcoran (Lawrence Berkeley National Laboratory), Rafael Zamora-Resendiz (Lawrence Berkeley National Laboratory), Xinlian Liu (Lawrence Berkeley National Laboratory), Silvia Crivelli (Lawrence Berkeley National Laboratory)

A protein's 3D structure determines its functionality, and is therefore a topic of great importance. This work leverages the power of Convolutional Neural Networks (CNNs) to classify proteins and extract features directly from their 3D structures. So far, researchers have been unable to fully exploit 3D structural information with 2D CNNs, partly because it is difficult to encode 3D data into the 2D format that can be ingested by such networks. We designed and implemented a novel method that maps 3D models to 2D data grids as a preprocessing step for 2D CNN use. Our experiments focused on the Ras protein family, which has been linked to various forms of cancer. Our trained CNNs are able to distinguish between two branches within the Ras family, HRas and KRas, which are similar in sequence and structure. Analysis of saliency maps suggests classification is accomplished by detection of structurally and biologically-meaningful sites.

P37: PaSTRI: A Novel Data Compression Algorithm for Two-Electron Integrals in Quantum Chemistry

Authors: Ali Murat Gok (Argonne National Laboratory, Northwestern University), Dingwen Tao (University of California, Riverside), Sheng Di (Argonne National Laboratory), Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Franck Cappello (Argonne National Laboratory)

Integral computations for two-electron repulsion energies are very frequently used applications in quantum chemistry. Computational complexity, energy consumption and the size of the output data generated by these computations scales with $O(N^4)$, where N is the number of atoms simulated. Typically, the same integrals are calculated multiple times. Storing these values and reusing them requires impractical amounts of storage space; whereas recalculating them requires a lot of computations. We propose PaSTRI (Pattern Scaling for Two-electron Repulsion Integrals), a fast novel compression algorithm which makes it possible to calculate these integrals only once, store them, and reuse them at much smaller computational cost than recalculation. PaSTRI is "lossy" compared to floating point numbers, but still maintains the precision level required by the integral computations. PaSTRI is an extension to SZ compressor package as a part of ECP-EZ. PaSTRI achieves 17.5:1 compression ratio whereas vanilla SZ achieves 8.0:1 and ZFP achieves 7.1:1.

P38: Benchmarking Parallelized File Aggregation Tools for Large Scale Data Management

Authors: *Tiffany Li (National Center for Supercomputing Applications, University of Illinois), Craig Steffen (National Center for Supercomputing Applications, University of Illinois), Ryan Chui (National Center for Supercomputing Applications, University of Illinois), Roland Haas (National Center for Supercomputing Applications, University of Illinois), Liudmila S. Mainzer (National Center for Supercomputing Applications, University of Illinois)*

Large-scale genomic data analyses have given rise to bottlenecks in data management due to the production of many small files. Existing file-archiving utilities, such as tar, are unable to efficiently package large datasets with upward of multiple terabytes and hundreds of thousands of files. To create parallelized and multi-threaded alternatives, ParFu (parallel archiving file utility), MPItar, and ptgz (parallel tar gzip) were developed by the Blue Waters team and the NCSA Genomics team as efficient data management tools, with the ability to perform parallel archiving (and eventually extracting). Scalability was tested for each tool as a function of the number of ranks executed and stripe count on a Lustre filesystem. We used two datasets typically seen in genomic analyses to measure the effects of different file-size distributions. These tests suggest the best user parameters and subsequent costs for usage as efficient replacements of data-packaging tools.

P39: Extremely Large, Wide-Area Power-Line Models

Authors: *Ross N. Adelman (US Army Research Laboratory)*

The electric and magnetic fields around power lines carry an immense amount of information about the power grid, and can be used to improve stability, balance loads, and reduce outages. To study this, extremely large models of transmission lines over a 49.5-sq-km tract of land near Washington, DC have been built. The terrain is modeled accurately using 1-m-resolution LIDAR data. The models are solved using the boundary element method, and the solvers are parallelized across Army Research Laboratory's Centennial supercomputer using a modified version of the domain decomposition method. The code on each node is accelerated using the fast multipole method and, when available, GPUs. Additionally, larger test models were used to characterize the scaling properties of the code. The largest test model had 10,020,913,152 elements, and was solved across 1024 nodes in 3.0 hours.

P40: Running Large-Scale Ultrasound Simulations on Piz Daint with 512 Pascal GPUs

Authors: *Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)*

Ultrasound simulation is a critical component of model-based treatment planning for ultrasound therapy. However, the domains are typically thousands of wavelengths in size, leading to large-scale numerical models with 10s of billions of unknowns. This paper presents a novel local Fourier basis domain decomposition for full-wave ultrasound propagation simulations with a custom bell function which ensures that the numerical error stays below 0.1% while enabling almost ideal strong scaling. Realistic benchmarks with 512 Nvidia P100 GPUs in the best EU supercomputer Piz Daint achieved efficiency between 90 and 100% with a speed-up over 100 and computational cost reduction by a factor of 12 compared to 1024 Haswell cores.

P41: OpenCL-Based High-Performance 3D Stencil Computation on FPGAs

Authors: *Hamid Reza Zohouri (Tokyo Institute of Technology), Artur Podobas (Tokyo Institute of Technology), Naoya Maruyama (RIKEN), Satoshi Matsuoka (Tokyo Institute of Technology)*

With the recent advancements in OpenCL-based High-Level Synthesis, FPGAs are now more attractive choices for accelerating High Performance Computing workloads. Despite their power efficiency advantage, FPGAs usually fall short in terms of sheer performance against GPUs due to having multiple times lower memory bandwidth and compute performance. In this work, we show that due to the architectural advantage of FPGAs for stencil computation, apart from power efficiency, these devices can also offer comparable performance to high-end GPUs. We achieve this goal using a parameterized OpenCL-based implementation that employs both spatial and temporal blocking, and multiple advanced FPGA-specific optimizations to maximize performance. We show that it is possible to achieve up to 60 GBps and 230 GBps of effective throughput for 3D stencil computation on Intel Stratix V and Arria 10 FPGAs, respectively, which is comparable to a highly-optimized implementation on high-end GPUs.

P42: TRIP: An Ultra-Low Latency, TeraOps/s Reconfigurable Inference Processor for Multi-Layer Perceptrons

Authors: *Ahmed Sanaullah (Boston University), Chen Yang (Boston University), Yuri Alexeev (Argonne National Laboratory), Kazutomo Yoshii (Argonne National Laboratory), Martin C. Herbordt (Boston University)*

Multi-Layer Perceptron (MLP) is one of the most commonly deployed Deep Neural Networks, representing 61% of the workload in

Google data-centers. MLP Inference, a memory bound problem, typically has hard response time deadlines and prefers latency over throughput. In our work, we designed a TeraOps/s Reconfigurable Inference Processor for MLPs (TRIP) on FPGAs that alleviates the memory bottleneck by storing all application specific weights on-chip. It can be deployed in multiple configurations, including host-independent operation. We have shown that TRIP achieves 60x better performance than the current state-of-the-art Google Tensor Processing Unit (TPU) for MLP Inference. It was demonstrated on the cancer patient datasets used in the Candle Exascale Computing Project (ECP).

P43: Deep Packet/Flow Analysis Using GPUs

Authors: Qian Gong (Fermi National Laboratory), Wenji Wu (Fermi National Laboratory), Phil DeMar (Fermi National Laboratory)

Deep packet inspection (DPI) faces severe challenges in high-speed networks as it requires high I/O throughputs and intensive computations. The parallel architecture of GPUs fits exceptionally well for per-packet traffic processing. However, TCP data stream need to be reconstructed in a per-flow level to deliver a consistent content analysis. Since the flow-centric operations are naturally anti-parallel and often require large memory space for buffering out-of-sequence packets, they can be problematic for GPUs. Here, we present a highly efficient DPI framework, which includes a purely GPU-implemented TCP flow tracking and stream reassembly. Instead of buffering till the TCP packets become in sequence, we process the packets in batch with pattern matching states between consecutive batches connected by a Aho-Corasick with a prefix-/suffix- tree method. Evaluation shows that our code can reassemble tens of millions of packets per second and conduct a signature-based DPI at 55 Gbit/s using an NVIDIA K40 GPU.

P44: Increasing Throughput of Multiprogram HPC Workloads: Evaluating a SMT Co-Scheduling Approach

Authors: Elias Lundmark (University West Sweden), Chris Persson (University West Sweden), Andreas de Blanche (University West Sweden, Tetra Pak), Thomas Lundqvist (University West Sweden)

Simultaneous Multithreading (SMT) is a technique that allows for more efficient processor utilization by scheduling multiple threads on a single physical core, thus increasing the instruction level parallelism as it mixes instructions from several threads. Previous research has shown an average throughput increase of around 20% with an SMT level of two, e.g. two threads per core. However, a bad combination of threads can actually result in decreased performance. To be conservative, many HPC-systems have SMT disabled, thus, limiting the number of scheduling slots in the system to one per core. However, for SMT to not hurt performance, we need to determine which threads should share a core. In this poster, we use 30 random SPEC CPU job mixed on a twelve-core Broadwell based node, to study the impact of enabling SMT using two different co-scheduling strategies. The results show that SMT can increase performance especially when using no-same-program co-scheduling.

P45: Campaign Storage: Erasure Coding with GPUs

Authors: Walker Haddock (University of Alabama, Birmingham), Matthew L. Curry (Sandia National Laboratories), Purushotham Bangalore (University of Alabama, Birmingham), Anthony Skjellum (University of Tennessee, Chattanooga)

Cloud computing has developed high capacity, reliable and economical storage systems based on object technology. Los Alamos National Labs has designed the storage systems for Trinity to include a cloud type object storage system as a layer between the Parallel File System (PFS). This pre-archive system has been dubbed "Campaign Storage" with the purpose of storing data products to be quickly accessible during the life of a research project. Data stored on the Campaign Storage can be pulled into the PFS or moved to archive after the data has been curated. Campaign Storage reduces the capacity requirements for PFS storage and reduces the data transfer bandwidth requirements for the archive storage.

We make these contributions to the pre-archive storage layer: * GPU assisted erasure coding * Demonstrating erasure on File Transfer Agents * Reducing erasure recovery costs with "lazy recovery" * Enabling larger erasure coded disk pools

P46: Understanding How OpenCL Parameters Impact on Off-Chip Memory Performance of FPGA Platforms

Authors: Yingyi Luo (Northwestern University), Zheming Jin (Argonne National Laboratory), Kazutomo Yoshii (Argonne National Laboratory), Seda Ogrenci-Memik (Northwestern University)

Reconfigurability has strong potential to achieve higher performance and energy efficiency in the post-Moore era. Field-programmable gate arrays (FPGAs), the most practical reconfigurable architecture today, are becoming more relevant to scientific computing thanks to hardened floating-point circuits and emerging FPGA high-level synthesis (HLS) technology. Most notably, FPGA vendors started supporting OpenCL for FPGA platforms, and some OpenCL-based codes have been ported to FPGAs. However, OpenCL offers no guarantee for performance portability; optimal OpenCL parameters such as global size and local size are different between platforms, which could lead to unfair comparisons. In this study, our objective is twofold: 1) to understand how OpenCL parameters impact off-chip memory access performance of the current generation of OpenCL-FPGA platforms and 2) to find effective OpenCL parameters empirically from microbenchmark results.

P47: Understanding Congestion on Omni-Path Fabrics

Authors: Lauren Gillespie (Southwestern University), Christopher Leap (University of New Mexico), Dan Cassidy (Los Alamos National Laboratory)

High-performance computing systems require high-speed interconnects, such as InfiniBand (IB), to efficiently transmit data. Intel's Omni-Path Architecture (OPA) is a new interconnect similar to IB that is implemented on some of Los Alamos National Laboratory's

recent clusters. Both interconnects suffer from degraded performance under heavy network traffic loads, resulting in packet discards. However, unlike IB, OPA specifically calls out these drops in the form of the performance counter, congestion discards. Owing to the relative immaturity of the OPA fabric technology, the correlation between performance degradation and congestion discards has not been fully evaluated to date. This research aims to increase the level of understanding of the effects congestion has on cluster performance by presenting a sufficiently high data injection load to the OPA fabric such that performance degradation is induced and the cause of this performance degradation can be evaluated. LA-UR-17-26341

P48: Prototyping of Offloaded Persistent Broadcast on Tofu2 Interconnect

Authors: Yoshiyuki Morie (RIKEN), Masayuki Hatanaka (RIKEN), Masamichi Takagi (RIKEN), Atsushi Hori (RIKEN), Yutaka Ishikawa (RIKEN)

With the increasing scale of parallel computers, it has become more important to reduce communication time. Overlapping computation and communication is one effective method for hiding communication delay. Although standard non-blocking collective communication is an overlap method, it requires generating a communication command sequence for each collective communication. In contrast, persistent non-blocking collective communication can generate the sequence at initialization and reuse it at the start of collective communication. Moreover, if the sequence can be offloaded to a network device, more efficient execution is possible without using CPU cycles.

In this poster, a persistent non-blocking broadcast is implemented using the offloading functionality of the Tofu2 interconnect on the Fujitsu FX100 supercomputer, the successor to the K computer. We report the performance improvement by offloading persistent non-blocking collective communication in a real machine.

P49: Toward Exascale HPC Systems: Exploiting Advances in High Bandwidth Memory (HBM2) through Scalable All-to-All Optical Interconnect Architectures

Authors: Pouya Fotouhi (University of California, Davis), Roberto Proietti (University of California, Davis), Paolo Grani (University of California, Davis), Mohammad Amin Nazirzadeh (University of California, Davis), S. J. Ben Yoo (University of California, Davis)

As we reach the limits of miniaturization in fabrication processes, the interpretation of Moore's law has changed from doubling the frequency every eighteen months to doubling the core count every three to four years (from 2 cores in 2004 to 16 cores in 2015). To reach exascale-level computation, the communication and data transfers between processors and memory is expected to increase drastically; the on-chip interconnect plays a key role in the overall system latency and energy-efficiency. Therefore, novel solutions providing one order of magnitude higher bandwidth and lower energy consumption than what is possible with current electrical interconnects are needed. This poster discusses an optical interconnected compute node that makes use of embedded photonic interconnects together with emerging high bandwidth memory technologies (such as HBM and HBM2). Two different multi-processors architectures with different requirements in terms of number of lasers, high-speed SERDES, and memory bandwidth per processors are presented.

P50: Energy-Efficient and Scalable Bio-Inspired Nanophotonic Computing

Authors: Mohammadamin Nazirzadeh (University of California, Davis), Pouya Fotouhi (University of California, Davis), Mohammadsadeh Shamsabardeh (University of California, Davis), Roberto Proietti (University of California, Davis), S. J. Ben Yoo (University of California, Davis)

This paper discusses bio-inspired neuromorphic computing utilizing nanophotonic, nanoelectronic, and NEMS technologies integrated into reconfigurable 2D-3D integrated circuits as hierarchical neural networks. The goal is to achieve $\geq 1000\times$ improvements in energy-per-operation compare to the state-of-the-art implementations of neural networks on Von-Neumann based computers. We combine nanophotonic and nanoelectronic technologies to build energy-efficient (~ 10 fJ/b) artificial spiking neurons with required functionality (spiking, integration, thresholding, reset). Photonic interconnects exploiting 2x2 NEMS-MZIs enables distance independent propagation of signal with weighted addition among the neurons as well as possibility of on-line learning capability. Using low-leakage nanophotonic and nanoelectronic devices, and NEMS, the static power consumption of the system can be decreased down to nearly zero. Realizing 2D-3D photonic integrated circuit technologies, the proposed system can overcome the scalability limitations of current neuromorphic computing architectures.

BP

P51: TuPiX-Flow: Workflow-Based Large-Scale Scientific Data Analysis System

Authors: Sunggeun Han (Korea Institute of Science and Technology Information), Jung-Ho Um (Korea Institute of Science and Technology Information), Hyunwoo Kim (Korea Institute of Science and Technology Information), Kyongseok Park (Korea Institute of Science and Technology Information)

With more research being actively conducted on big data, there has been a growing interest towards data-intensive data science in various fields. Against this backdrop, extensive efforts have been made to apply High Performance Computing (HPC) technology such as distributed computing or parallel computing. However, many researchers are unable to fully utilize such technology due to their lack of knowledge, programming skills, and analytical skills for large-scale data. TuPiX-Flow, which provides a workflow-based user interface, enables researchers to easily analyze large-scale data using workflow diagrams for end-to-end analytical processes even without programming knowledge or implementation technology. In addition, large-scale data can be efficiently analyzed in software compatible with HPC, including distributed computing and parallel computing. As a case study of large-scale data analysis using TuPiX-Flow, a model for the detection of red tides surrounding the Korean peninsula was analyzed based on ocean color satellite data.

P52: A Simulation-Based Analysis on the Configuration of Burst Buffer

Authors: *Tianqi Xu (Tokyo Institute of Technology), Kento Sato (Lawrence Livermore National Laboratory), Satoshi Matsuoka (Tokyo Institute of Technology)*

Burst buffers have been widely deployed in many supercomputer systems to absorb bursty I/O and accelerate I/O performance. Previous work has shown that with burst buffer systems, I/O operations from computer nodes can be greatly accelerated. Different configurations of burst buffers can have huge impact on performance of applications. However, the proper configuration of burst buffers for given systems and workloads still remains an open problem. In this paper, we address this challenge by simulating the behavior of a burst buffer under different buffer sizes with trace logs from a set of HPC applications in a distributed environment. We tuned our simulator with a production level burst buffer system. From the results of the simulation, we found that for most of HPC applications, using a buffer size that is less than half of the total access space of the application can still achieve high performance. BP

P53: TensorViz: Visualizing the Training of Convolutional Neural Network Using ParaView

Authors: *Xinyu Chen (University of New Mexico), Qiang Guan (Los Alamos National Laboratory), Xin Liang (University of California, Riverside), Li-Ta Lo (Los Alamos National Laboratory), Simon Su (US Army Research Laboratory), Trilce Estrada (University of New Mexico), James Ahrens (Los Alamos National Laboratory)*

Deep Convolutional Networks have been very successful in visual recognition tasks recently. Previous works visualize learned features at different layers to help people understand how CNNs learn visual recognition tasks. However, they do not help to accelerate the training process. We use ParaView to provides both qualitative and quantitative visualization that help understand the learning procedure, tune the learning parameters, and direct merging and pruning of neural networks.

P54: Investigating Hardware Offloading for Reed-Solomon Encoding

Authors: *John W. Dermer (Los Alamos National Laboratory), Gustavo De Leon (Los Alamos National Laboratory; University of California, Berkeley), Tyler S. Rau (Los Alamos National Laboratory)*

Reed-Solomon (RS) encoding is a storage scheme which offers better scalability, but requires heavier computation, compared to other models. This presents a problem as it requires users to purchase brawnier CPUs than would be otherwise necessitated. However, Mellanox's ConnectX-4 Infiniband cards have the capability to perform RS encoding on the HCA hardware; removing the need for powerful CPUs to calculate it. We investigated the performance, measured in throughput, between these cards and Intel's ISA-Library, with regard to various block sizes and concurrency. We conclude that the MLX cards encoded faster and more consistently than ISA-L. Furthermore, the ConnectX-5 cards support the Galois Field (GF) 2^8 , this grants compatibility with data encoded by any system using GF(2^8) or less, including ISA-L. These cards enable users to substantially increase encoding and decoding throughput by using more cards; additionally enabling the use of less powerful CPUs to achieve similar performance. LA-UR-17-26333

P55: Incorporating Proactive Data Rescue into ZFS Disk Recovery for Enhanced Storage Reliability

Authors: *Zhi Qiao (University of North Texas), Song Fu (University of North Texas), Hsing-bung Chen (Los Alamos National Laboratory), Michael Lang (Los Alamos National Laboratory)*

As tremendous amount of data are generated every day, storage systems store exabytes of data on hundreds of thousands of disk drives. At such a scale, disk failures become the norm. Data recovery takes longer time due to increased disk capacity. ZFS is a widely used filesystem, providing data recovery from corruption. Many factors may affect ZFS's recovery performance in a production environment. Additionally, disk failure prediction techniques enables ZFS to proactively rescue data prior to disk failures. In this poster, we extensively evaluate the recovery performance with a variety of ZFS configurations. We also compare the performance of different data rescue strategies, including post-failure disk recovery, proactive disk cloning, and proactive data recovery. Our proposed analytic model uses the collected zpool utilization data and system configuration to derive the optimal data rescue strategy that best suits the storage array in the current state.

P56: ZoneTier: A Zone-Based Storage Tiering and Caching Co-Design to Integrate SSDs with Host-Aware SMR Drives

Authors: *Xuchao Xie (National University of Defense Technology), Liquan Xiao (National University of Defense Technology), David H.C. Du (University of Minnesota)*

Integrating solid state drives (SSDs) and host-aware shingled magnetic recording (HA-SMR) drives can potentially build a cost-effective high-performance storage system. However, existing SSD tiering and caching designs in such a hybrid system are not fully matched with the intrinsic properties of HA-SMR drives due to their handling of non-sequential writes (NSWs) from both workloads and data migration. We propose ZoneTier, a zone-based storage tiering and caching co-design, to effectively control all the NSWs by leveraging the host-aware property of HA-SMR drives. ZoneTier exploits the real-time data layouts of HA-SMR zones to optimize zone placements, reshape NSWs generated from zone demotions to HA-SMR drive preferred sequential writes, and transforms the inevitable NSWs to HA-SMR zones to cleaning-friendly write traffics. Our experiments show that ZoneTier can utilize SSDs with high efficiency, minimize relocation overhead, shorten performance recovery time of HA-SMR drives, and finally accomplish better system performance than existing hybrid storage designs.

P57: Adaptive Tier Selection for NetCDF and HDF5

Authors: Jakob Luettgau (German Climate Computing Center), Eugen Betke (German Climate Computing Center), Olga Perevalova (University of Hamburg), Julian Kunkel (German Climate Computing Center), Michael Kuhn (University of Hamburg)

Scientific applications on supercomputers tend to be I/O-intensive. To achieve portability and performance, data description libraries such as HDF5 and NetCDF are commonly used. Unfortunately, the libraries often default to suboptimal access patterns for reading/writing data to multi-tier distributed storage. This work explores the feasibility of adaptively selecting tiers depending on an application's I/O behavior.

P58: Wharf: Sharing Docker Images across Hosts from a Distributed Filesystem

Authors: Chao Zheng (University of Notre Dame), Lukas Rupprecht (IBM), Vasily Tarasov (IBM), Mohamed Mohamed (IBM), Dimitrios Skourtis (IBM), Amit S. Warke (IBM), Dean Hildebrand (IBM), Douglas Thain (University of Notre Dame)

Due to their portability and less overhead compared to traditional virtual machines, containers are becoming an attractive solution for running HPC workloads. Docker is a popular toolset which enables convenient provisioning and management of containers and their corresponding images. However, Docker does not natively support running on shared storage, a crucial requirement in large-scale HPC clusters which are often diskless or access data via a shared burst buffer layer. This lack of distributed storage support can lead to overhead when running containerized HPC applications. In this work, we explore how Docker images can be served efficiently from a shared distributed storage layer. We implement a distributed layer on top of Docker that allows multiple Docker daemons to access container images from a shared filesystem such as IBM Spectrum Scale or NFS. Our design is independent of the underlying storage layer and minimizes the synchronization overhead between different daemons.

P59: Secure Enclaves: An Isolation-Centric Approach for Creating Secure High-Performance Computing Environments

Authors: Ferrol Aderholdt (Oak Ridge National Laboratory), Susan Hicks (Oak Ridge National Laboratory), Thomas Naughton (Oak Ridge National Laboratory), Lawrence Sorrillo (Oak Ridge National Laboratory), Blake Caldwell (University of Colorado, Boulder), James Pogge (Tennessee Technological University), Stephen L. Scott (Tennessee Technological University)

High performance computing environments are used for a wide variety of workloads. These systems may process data at various security levels but in so doing are often enclaved at the highest security posture, which may limit usability or performance. The traditional approach used to provide isolation is effective at the creation of secure enclaves, but poses significant challenges with respect to the use of shared infrastructure in HPC environments. We evaluate the use of system-level (i.e., hypervisor-based) and operating system level (i.e., containers) virtualization as well as software defined networking (SDN) as possible mechanisms for secure, isolation-centric enclaves (secure enclaves). We describe our approach to secure HPC enclaves and provide benchmark results for three focus areas (compute, network and data storage) where isolation mechanisms are most significant.

P60: Managing dbGaP Data with Stratus, a Research Cloud for Protected Data

Authors: Evan F. Bollig (University of Minnesota), Graham Allan (University of Minnesota), Benjamin J. Lynch (University of Minnesota), Yectli Huerta (University of Minnesota), Mathew Mix (University of Minnesota), Brent Swartz (University of Minnesota), Edward A. Munsell (University of Minnesota), Joshua Leibfried (University of Minnesota), Naomi Hospodarsky (University of Minnesota)

Modern research computing needs at academic institutions are evolving. While traditional HPC continues to satisfy most workflows, a new generation of researcher has emerged looking for sophisticated, self-service control of compute infrastructure in a cloud-like environment. Often, these demands are not for their own interest, but nonetheless present due to constraints imposed by data governance and protection policies that cannot be satisfied by traditional HPC.

To cater to these modern users, the Minnesota Supercomputing Institute deployed a cloud service for research computing called Stratus. Stratus is designed expressly to satisfy the requirements set forth by the NIH Genomic Data Sharing (GDS) Policy for dbGaP data. It is powered by the Newton version of OpenStack, and backed by Ceph storage. The service offers three features not available on traditional HPC systems: a) on-demand availability of compute resources; b) long-running jobs (i.e., > 30 days); and c) container-based computing with Docker applications.

P61: Cloud Resource Selection Based on PLS Method for Deploying Optimal Infrastructures for Genomic Analytics Application

Authors: Katsunori Miura (Kitami Institute of Technology), Courtney Powell (Hokkaido University), Masaharu Munetomo (Hokkaido University)

This poster proposes a method for determining infrastructures composed of cloud resources that concurrently meet satisfiability and optimality system requirements, such as computational performance, maximum price payable, and deployment location of a genomic analytics application. The input to the proposed method is a mathematical formula that captures the system requirements given by the user, which is defined in accordance with first-order predicate logic, whereas the output is a set of unit clauses representing infrastructures for deploying the genomic analytics application. In the proposed method, an equivalent transformation algorithm is used to generate valid solutions with respect to system requirements, and a genetic algorithm is used to evaluate the optimality of the solutions.

P62: How To Do Machine Learning on Big Clusters

Authors: Thomas Ashby (IMEC), Tom Vander Aa (IMEC), Stanislav Bohm (Technical University of Ostrava), Vojtech Cima (Technical University of Ostrava), Jan Martinovic (Technical University of Ostrava), Vladimir Chupakhin (Janssen Global Services LLC)

Scientific pipelines, such those in chemogenomics machine learning applications, often compose of multiple interdependent data processing tasks. We are developing HyperLoom - a platform for defining and executing workflow pipelines in large-scale distributed environments. HyperLoom users can easily define dependencies between computational tasks and create a pipeline which can then be executed on HPC systems. The high-performance core of HyperLoom dynamically orchestrates the tasks over available resources respecting task requirements. The entire system was designed to have a minimal overhead and to efficiently deal with varying computational times of the tasks. HyperLoom allows to execute pipelines that contain basic built-in tasks, user-defined Python tasks, tasks wrapping third-party applications or a combination of those.

BP

P63: FleCSPH: a Parallel and Distributed Smoothed Particle Hydrodynamics Framework Based on FleCSI

Authors: Julien Loiseau (University of Reims Champagne-Ardenne), Hyun Lim (Brigham Young University), Ben Bergen (Los Alamos National Laboratory), Nicholas Moss (Los Alamos National Laboratory)

In this poster session, we introduce our new parallel and distributed Smoothed Particle Hydrodynamics implementation, FleCSPH, which is based on the open-source framework FleCSI. This framework provides us the basic data structures and runtime required in our work. We intend to provide a parallel and distributed version of Binary, Quad, and Oct trees data structure, respectively for 1, 2 and 3 dimensions dedicated for SPH problems.

Also, we present various test problems in several dimensions that indicate the flexibility and the scalability of our toolkit. For application and tests we simulate classical physics test cases like Sod Shock Tube, Sedov Blast Wave or Fluid Flows. The aim of this work is to simulate binary compact object mergers such as white dwarfs and neutron stars that address many interesting astrophysical phenomena in the universe.

P64: romeoLAB : HPC Training Platform on HPC facility

Authors: Jean-Matthieu Etancelin (University of Reims Champagne-Ardenne), Arnaud Renard (University of Reims Champagne-Ardenne)

In this pre-exascale era, we are observing a dramatic increase of the necessity of computer science courses dedicated to parallel programming with advanced technologies on hybrid architectures. The full hybrid cluster Romeo has long been used for that purpose in order to train master students and cluster users. The main issue for trainees is the cost of accessing and exploiting a production facility in a pedagogic context. The use of some specific techniques and software (SSH, workload manager, remote file system, ...) is mandatory without being part of courses prerequisites nor pedagogic objectives. The romeoLAB platform we developed at ROMEO HPC Center is an online interactive pedagogic platform for HPC technologies courses. Its main purpose is to simplify the process of resource usage in order to focus on the taught subjects. This paper presents the context, the romeoLAB architecture and its motivations, usages and pedagogical contents.

P65: CAPES: Unsupervised System Performance Tuning Using Neural Network-Based Deep Reinforcement Learning

Authors: Yan Li (University of California, Santa Cruz), Kenneth Chang (University of California, Santa Cruz), Oceane Bel (University of California, Santa Cruz), Ethan Miller (University of California, Santa Cruz), Darrell Long (University of California, Santa Cruz)

Parameter tuning is an important task of storage performance optimization. Current practice usually involves numerous tweak-benchmark cycles that are slow and costly. To address this issue, we developed CAPES, a model-less deep reinforcement learning-based unsupervised parameter tuning system driven by a deep neural network (DNN). It is designed to find optimal values for computer systems that have tunable parameters when human tuning can be costly and often cannot achieve optimal performance. CAPES takes periodic measurements of a target computer system's state, and trains a DNN which uses Q-learning to suggest changes to the system's current parameter values. CAPES is minimally intrusive, and can be deployed into a production system to collect training data and suggest tuning actions during the system's daily operation. Evaluation of a prototype on a Lustre file system demonstrates an increase in I/O throughput up to 45% at saturation point.

P66: Analyzing Multi-Layer I/O Behavior of HPC Applications

Authors: Ronny Tschüter (Technical University Dresden), Christian Herold (Technical University Dresden), Bert Wesarg (Technical University Dresden), Matthias Weber (Technical University Dresden)

In this work, we present an approach to analyze multi-layer I/O behavior of HPC applications. For a comprehensive analysis it is not sufficient to track I/O operations on a single layer because applications may combine multiple I/O paradigms (e.g., MPI I/O, NetCDF, HDF5). Furthermore, I/O libraries may use another I/O paradigm internally. With our approach, I/O activities can be captured at any layer. This work introduces methodologies to intercept calls to I/O libraries and presents an event design to record applications' I/O activities. We implement our approach in the Score-P measurement system and prove its usability with a study of the Met Office/NERC Cloud Model (MONC) code.

P67: Measuring I/O Behavior on Upcoming Systems with NVRAM

Authors: *Christian Herold (Technical University Dresden), Ronny Tschüter (Technical University Dresden)*

Upcoming HPC systems will use NVRAM to address existing I/O bottlenecks. The I/O performance thereby is one of the keys for the exascale challenges. NVRAM introduces a new level in the memory hierarchy and can be utilized by different technologies, e.g. memory mapped files or block transfer operations. Using NVRAM without considering the complete hierarchy may lead to an inefficient usage and bad I/O performance. Therefore, in this work, we evaluate techniques for measuring the I/O behavior of applications that utilize NVRAM in various use cases. We extended the application tracing of our tool Score-P and introduced several metrics as well as events for different use cases of NVRAM.

P68: Continuous Clock Synchronization for Accurate Performance Measurements

Authors: *Johannes Ziegenbalg (Technical University Dresden), Matthias Weber (Technical University Dresden)*

Clock synchronization is a key prerequisite for accurate performance analysis. This is particularly true for most HPC systems, due to the lack of a system-wide timer. The effects of different clock properties lead to false measurements and wrong conclusions. Often these errors remain undetected by the user. Different environmental effects cause continuous changes to clock properties over time. This behavior is ignored by the usual post-mortem approach for clock synchronization. In order to improve time measurement accuracy we implemented a method for continuous clock synchronization. In this work, we share our experiences and draw conclusions for tool development.

P69: Portable Methods for Measuring Cache Hierarchy Performance

Authors: *Tom Deakin (University of Bristol), James Price (University of Bristol), Simon McIntosh-Smith (University of Bristol)*

There has been a recent influx of different processor architecture designs into the market, with many of them targeting HPC applications. When estimating application performance, developers are used to considering the most common figures of merit, such as peak FLOP/s, memory bandwidth, core counts, and so on. In this study, we present a detailed comparison of on-chip memory bandwidths, including single core and aggregate across a node, for a set of next-generation CPUs.

We do this in such a way as to be portable across different architectures and instruction sets. Our study indicates that, while two processors might look superficially similar when only considering the common figures of merit, those two processors might have radically different on-chip memory bandwidth, a fact which may be crucial when understanding observed application performance. Our results and methods will be made available on GitHub to aid the community in evaluating cache bandwidths.

P70: FFT, FMM, and Multigrid on the Road to Exascale: Performance Challenges and Opportunities

Authors: *Huda Ibeid (University of Illinois), Luke Olson (University of Illinois), William Gropp (University of Illinois)*

FFT, FMM, and multigrid methods are widely used fast and highly scalable solvers for elliptic PDEs. However, emerging systems are introducing challenges in comparison to current petascale computers. The International Exascale Software Project Roadmap identifies several constraints on the design of exascale software. Challenges include massive concurrency, energy efficiency, resilience management, exploiting the high performance of heterogeneous systems, and utilizing the deeper and more complex memory hierarchy expected at exascale. In this study, we perform model-based comparison of the FFT, FMM, and multigrid methods in the context of these constraints and use the performance models to offer predictions about the methods performance on possible exascale system configurations, based on current technology trends.

P71: Is ARM Software Ecosystem Ready for HPC?

Authors: *Fabio Banchelli (Barcelona Supercomputing Center), Daniel Ruiz (Barcelona Supercomputing Center), Ying Hao Xu Lin (Barcelona Supercomputing Center), Filippo Mantovani (Barcelona Supercomputing Center)*

In recent years, the HPC community has increasingly grown its interest towards the ARM architecture with research projects targeting primarily the deployment of ARM-based clusters. Attention is usually given to hardware platforms, however the availability of a mature software ecosystem and the possibility of running large and complex HPC applications plays a key role in the consolidation process of a new technology.

For this reason in this poster we present a preliminary evaluation of the ARM system software ecosystem, limited here to the ARM HPC Compiler and the ARM Performance Libraries, together with a porting and testing of three fairly complex HPC code suites: QuantumESPRESSO, WRF, and FEniCS.

These codes have been proposed as HPC challenges during the last two editions of the Student Cluster Competition at ISC where all the authors have been involved operating an ARM-based cluster and awarded with the Fan Favorite award.

P72: New Developments for PAPI 5.6+

Authors: *Anthony Danalis (University of Tennessee), Heike Jagode (University of Tennessee), Vince Weaver (University of Maine), Yan Liu (University of Maine), Jack Dongarra (University of Tennessee)*

The HPC community has relied on PAPI to track low-level hardware operations for over 15 years. In that time, the needs of software

developers have changed immensely, and the PAPI team aims to meet these demands through a better understanding of deep and heterogeneous memory hierarchies and finer-grain power-management support.

This poster demonstrates how PAPI enables power-tuning to reduce overall energy consumption without, in many cases, a loss in performance. Furthermore, we discuss efforts to develop microbenchmarks intended to assist application developers who are interested in performance analysis by automatically categorizing and disambiguating performance counters. Finally, the poster illustrates efforts to update PAPI's internal sanity checks, designed to inspect that PAPI's predefined events are in fact measuring the values they claim to measure, and modernize the implementation of critical API functions, e.g., `PAPI_read()`, and the sampling interface so that more information can be captured and reported with lower overhead.

BP

P73: HPC Production Job Quality Assessment

Authors: Omar Aaziz (*New Mexico State University*), Jonathan Cook (*New Mexico State University*)

Users of HPC systems would benefit from more feedback about the quality of their application runs, such as knowing whether or not the performance of a particular run was good, or whether the resources requested were enough, or too much. Such feedback requires more information to be kept regarding production application runs, and requires some analytics to assess any new runs. In this research, we assess the practicality of using job data, system data, and hardware performance counters in a near-zero overhead manner to assess job performance, in particular whether or not the job runtime was in line with expectations from historical application performance. We show over four proxy applications and two real application that our assessment is within 10% of actual performance.

P74: A Methodology for Bridging the Native and Simulated Executions of Parallel Applications

Authors: Ali Mohammed (*University of Basel*), Ahmed Eleliemy (*University of Basel*), Florina M. Ciorba (*University of Basel*)

Simulation is considered as the third pillar of science, following experimentation and theory. Bridging the native and simulated executions of parallel applications is needed for attaining trustworthiness in simulation results. Yet, bridging the native and simulated executions of parallel applications is challenging. This work proposes a methodology for bridging the native and simulated executions of message passing parallel applications on high performance computing (HPC) systems in two steps: Expression of the software characteristics, and representation and verification of the hardware characteristics in the simulation. This work exploits the capabilities of the SimGrid [3] simulation toolkit's interfaces to reduce the effort of bridging the native and simulated executions of a parallel application on an HPC system. For an application from computer vision, the simulation of its parallel execution using straightforward parallelization on an HPC cluster approaches the native performance with a minimum relative percentage difference of 5.6%.

P75: Model-Agnostic Influence Analysis for Performance Data

Authors: Rahul Sridhar (*University of California, Irvine; Lawrence Livermore National Laboratory*), Rushil Anirudh (*Lawrence Livermore National Laboratory*), Jayaraman J. Thiagarajan (*Lawrence Livermore National Laboratory*), Nikhil Jain (*Lawrence Livermore National Laboratory*), Todd Gamblin (*Lawrence Livermore National Laboratory*)

Execution time of an application is affected by several performance parameters, e.g. number of threads, decomposition, etc. Hence, an important problem in high performance computing is to study the influence of these parameters on the performance of an application. Additionally, quantifying the influence of individual parameter configurations (data samples) on performance also aids in identifying sub-domains of interest in high-dimensional parameter spaces. Conventionally, such analysis is performed using a surrogate model, which introduces its own bias that is often non-trivial to undo, leading to inaccurate results. In this work, we propose an entirely data-driven, model-agnostic influence analysis approach based on recent advances in analyzing functions on graphs. We show that the problem of identifying influential parameters (features) and configurations (samples) can be effectively addressed within this framework.

P76: A Compiler Agnostic and Architecture Aware Predictive Modeling Framework for Kernels

Authors: William Killian (*University of Delaware*), Ian Karlin (*Lawrence Livermore National Laboratory*), David Beckingsale (*Lawrence Livermore National Laboratory*), John Cavazos (*University of Delaware*)

Multi-architecture machines make program characterization for modeling a regression outcome difficult. Determining where to offload compute-dense portions requires accurate prediction models for multiple architectures. To productively achieve portable performance across these diverse architectures, users are adopting portable programming models such as OpenMP and RAJA.

When adopted, portable models make traditional high-level source code analysis inadequate for program characterization. In this poster, we introduce a common microarchitecture instruction format (ComIL) for program characterization. ComIL is capable of representing programs in an architecture-aware compiler-agnostic manner. We evaluate feature extraction with ComIL by constructing multiple regression-objective models for performance (execution time) and correctness (maximum absolute error). These models perform better than the current state of the art -- achieving a mean error rate of only 4.7% when predicting execution time. We plan to extend this work to handle multiple architectures concurrently and evaluate with more representative physics kernels.

P77: AutoTuneTMP: Auto Tuning in C++ With Runtime Template Metaprogramming

Authors: David Pfander (University of Stuttgart), Dirk Pflüger (University of Stuttgart)

Maximizing the performance on modern hardware platforms has become more and more difficult, due to different levels of parallelism and complicated memory hierarchies. Auto tuning can help developers to address these challenges by writing code that automatically adjusts to the underlying hardware platform. AutoTuneTMP is a new C++-based auto tuning framework that uses just-in-time compilation to enable runtime-instantiable C++ templates. We use C++ template metaprogramming to provide data structures and algorithms that can be used to develop tunable compute kernels. These compute kernels can be tuned with different optimization strategies. We demonstrate for a first prototype the applicability and usefulness of our approach by tuning 6 parameters of a DGEMM implementation, achieving 68% peak performance on an Intel Skylake processor.

P78: Performance Evaluation of Graph500 Considering CPU-DRAM Power Shifting

Authors: Yuta Kakibuka (Kyushu University), Yuichiro Yasui (Kyushu University), Takatsugu Ono (Kyushu University), Katsuki Fujisawa (Kyushu University), Koji Inoue (Kyushu University)

There are power constraints on computer systems which comes from technical, costly or social demands. Power wall is one of the most serious issues for post petascale high-performance computing. A promising solution to tackle this problem is to effectively manage power resources based on the characteristics of workloads. In power constrained computing, the key is to translate the limited power budget into sustained performance effectively. To achieve this goal, assigning the appropriate amount of power budget to each hardware component, or power shifting, is a critical challenge.

In this work, we focus on large-scale graph processing. Graph analysis algorithms are increasing its importance with growing demands of big data analysis. However, the impact of power constraint on the performance of graph processing application is not declared. Our work is the performance evaluation of Graph500 under power constraints to CPU and DRAM.

P79: Porting the Opacity Client Library to a CPU-GPU Cluster Using OpenMP 4.5

Authors: Jason S. Kimko (College of William and Mary), Michael M. Pozulp (Lawrence Livermore National Laboratory), Riyaz Haque (Lawrence Livermore National Laboratory), Leopold Grinberg (IBM)

The poster accompanying this summary exhibits our experience porting the Opacity client library to IBM's "Minsky" nodes using OpenMP 4.5. We constructed a GPU-friendly container class that mimics existing library functionality. We benchmarked our implementation on Lawrence Livermore National Laboratory's (LLNL) RZManta, a Minsky cluster. In our benchmarks on a single POWER8 CPU and Tesla P100 GPU, we observed up to a 4x speedup including CPU-GPU data transfers and up to a 30x speedup excluding data transfers. Optimizing to reduce register pressure and increase occupancy may improve speedups. Our results demonstrate a successful and beneficial library port to the CPU-GPU architecture.

P80: Adaptive Loop Scheduling with Charm++ to Improve Performance of Scientific Applications

Authors: Vivek Kale (University of Southern California), Harshitha Menon (Lawrence Livermore National Laboratory), Karthik Senthil (University of Illinois)

Supercomputers today employ a large number of cores on each node. The Charm++ parallel programming system provides an intelligent runtime which has been highly effective at providing dynamic load balancing across nodes of a supercomputer. Modern multi-core nodes present new challenges and opportunities for Charm++. The large degree of over-decomposition required may lead to high overhead. We modified the Charm++ Runtime System (RTS) to assign Charm++ objects to nodes, thus reducing over-decomposition, and spreading work across cores via parallel loops. We modify a library of the Charm++ software suite that supports loop parallelism by adding to it a loop scheduling strategy that maximizes load balance across cores while minimizing data movement. We tune parameters of the RTS and the loop scheduling strategy to improve performance of benchmark codes run on a variety of architectures. Our technique improves performance of a Particle-in-Cell code run on the Blue Waters supercomputer by 17.2%.

BP

P81: Offloading Python Kernels to Micro-Core Architectures

Authors: Nick Brown (University of Edinburgh)

Micro-core architectures combine many low memory, low power computing cores together in a single package. These can be used as a co-processor or standalone but due to limited on-chip memory and esoteric nature of the hardware, writing efficient parallel codes for them is challenging. We previously developed ePython, a low memory (24Kb) implementation of Python supporting the rapid development of parallel Python codes and education for these architectures. In this poster we present our work on an offload abstraction to support the use of micro-cores as an accelerator. Programmers decorate specific functions in their Python code, running under any Python interpreter on the host, with our underlying technology then responsible for the low-level data movement, scheduling and execution of kernels on the micro-cores. Aimed at education and fast prototyping, a machine learning code for detecting lung cancer, where computational kernels are offloaded to micro-cores, is used to illustrate the approach.

P82: Performance Evaluation of the NVIDIA Tesla P100: Our Directive-Based Partitioning and Pipelining vs. NVIDIA's Unified Memory

Authors: Xuwen Cui (Virginia Tech), Thomas R. W. Scogland (Lawrence Livermore National Laboratory), Bronis R. de Supinski (Lawrence Livermore National Laboratory), Wu-chun Feng (Virginia Tech)

We need simpler mechanisms to leverage the performance of accelerators, such as GPUs, in supercomputers. Programming models like OpenMP offer simple-to-use but powerful directive-based offload mechanisms. By default, these models naively copy data to or from the device without overlapping computation. Achieving performance can require extensive hand-tuning to apply optimizations such as pipelining. Users must manually partition data whenever it exceeds device memory. Our directive-based partitioning and pipelining extension for accelerators overlaps data transfers and kernel computation without explicit user data-splitting. We compare a prototype implementation of our extension to NVIDIA's Unified Memory on the Pascal P100 GPU and find that our extension outperforms Unified Memory on average by 68% for data sets that fit into GPU memory and 550% for those that do not.

P83: Contracts for Message-Passing Programs

Authors: *Ziqing Luo (University of Delaware), Stephen F. Siegel (University of Delaware)*

Verification of message-passing parallel programs is challenging because of the state-explosion problem. A procedure-level contract system for parallel programs can help overcome this challenge by verifying contracts individually, in a modular way. A contract system is used to specify intended behaviors of programs. Contracts can be checked at run-time or verified formally. There is a mature theory of contracts for sequential programs, but little work has been done on parallel programs, and even less for message-passing parallel programs. We developed a theory of contracts for message-passing programs and realize this theory in a contract language for programs that use the Message Passing Interface (MPI). We are also developing a verification tool that uses symbolic execution and model checking techniques to prove that MPI programs satisfy their contracts.

P84: PRESAGE: Selective Low Overhead Error Amplification for Easy Detection

Authors: *Vishal Chandra Sharma (University of Utah), Arnab Das (University of Utah), Ian Briggs (University of Utah), Ganesh Gopalakrishnan (University of Utah), Sriram Krishnamoorthy (Pacific Northwest National Laboratory)*

Soft-errors remain a vexing challenge. Today's error detectors either come with high false positives or high omissions. Our work focuses on not losing an error sown, but amplifying it so that cheap detection is enabled.

Consider structured address generation in loops where data from a base address plus offset is used. An erroneous offset no longer crashes today's applications thanks to large address spaces; instead, it silently corrupts data (unintended fetch). We relativize address generation using LLVM, thus each new address is not base plus offset but previous relative address plus offset. If one address is corrupted, all future address are corrupted in a chain. This permits efficient loop exit-point detection.

Relativization has low overhead, actually lowered by some ISAs down to zero. These advantages survive crucial compiler memory access optimizations. We demonstrate 100% SDC detection for a class of benchmarks with respect to structured address protection.

P85: GPU Mekong: Simplified Multi-GPU Programming Using Automated Partitioning

Authors: *Alexander Matz (University of Heidelberg)*

GPU accelerators are pervasively used in the HPC community, because they provide excellent computational performance at a reasonable power efficiency. While programming single-GPU applications is comparatively productive, programming multiple GPUs using data-parallel languages is tedious and error prone as the user has to manually orchestrate data movements and kernel launches.

The Mekong research project is driven by the motivation to improve productivity of multi-GPU systems by compiler based partitioning of single-device data-parallel programs. Key to scalable performance improvement is the resolution of data dependencies between kernels and the orchestration of these kernels. Mekong relies on polyhedral compilation to identify memory access patterns in order to compile a single-GPU application into a multi-GPU application.

In this work, the Mekong project is introduced and its components explained. While the tool is still under development, preliminary results are available and are shortly discussed demonstrating the potential of this approach.

P86: HyGraph: High Performance Graph Processing on Hybrid CPU+GPUs platforms

Authors: *Stijn Heldens (University of Twente), Ana Lucia Varbanescu (University of Amsterdam), Alexandru Iosup (Vrije University Amsterdam)*

Graph analytics is becoming increasingly important in many domains, such as in biology, social sciences, and data mining. Many large-scale graph-processing systems have been proposed, either targeting distributed clusters or GPU-based accelerated platforms. However, little research exists on designing systems for hybrid CPU-GPU platforms, i.e., exploiting both the CPU and the GPU efficiently.

In this work, we present HyGraph, a novel graph-processing system for hybrid platforms which delivers performance by using both the CPU and GPUs concurrently. Its core feature is dynamic scheduling of tasks onto both the CPU and the GPUs, thus providing load balancing, contrary to the state-of-the-art approach based on static partitioning. Additionally, communication overhead is minimized by overlapping computation and communication.

Our results demonstrate that HyGraph outperforms CPU-only and GPU-only solutions, delivering close-to-optimal performance. Moreover, it supports large-scale graphs which do not fit into GPU memory and is competitive against state-of-the-art systems.

P87: EoCoE Performance Benchmarking Methodology for Renewable Energy Applications

Authors: Paul Gibbon (*Forschungszentrum Juelich*), Matthieu Haefele (*French Alternative Energies and Atomic Energy Commission*), Sebastian Luehrs (*Forschungszentrum Juelich, Juelich Supercomputing Center*)

An optimisation strategy developed by the Energy Oriented Centre of Excellence (EoCoE) is presented for computational models used in a variety of renewable energy domains. It is found that typical applications in this comparatively new sector exhibit a huge range of HPC maturity, from simple parallelization needs to near-exascale readiness. An essential part of this process has therefore been the introduction of a flexible, quantitative performance assessment of applications using the benchmarking tool JUBE to automatically extract up to 28 different metrics taken with several state-of-the-art performance tools. An initial hands-on workshop to establish this baseline status is consolidated by follow-up actions by joint code-teams comprising members of both developer groups and HPC centres involved with the EoCoE consortium. Examples of early successes achieved with this policy are given, together with an outlook on challenges faced for energy applications with next-generation, pre-exascale architectures.

P88: PetaVision Neural Simulation Toolbox on Intel KNLs

Authors: Boram Yoon (*Los Alamos National Laboratory*), Pete Schultz (*Los Alamos National Laboratory, New Mexico Consortium*), Garrett Kenyon (*Los Alamos National Laboratory, New Mexico Consortium*)

We implemented a large-scale neuromorphic algorithm called the Sparse Prediction Machine (SPM), on the Los Alamos Trinity supercomputer. Specifically, we used PetaVision, an open source high-performance neural simulation toolbox, to implement a 4-layer SPM applied to ImageNet video. Various optimization techniques were applied to efficiently utilize up to 8192 KNL nodes. The goal of the SPM is to predict future states of a system from a sequence of previous states, or in the case of video, to predict a subsequent frame from previous frames. In our training, the SPM was able to predict the 8th frame from the preceding 7 frames, including successful separation of foreground and background motion.

P89: Desh: Deep Learning for HPC System Health Resilience

Authors: Anwesha Das (*North Carolina State University*), Abhinav Vishnu (*Pacific Northwest National Laboratory*), Charles Siegel (*Pacific Northwest National Laboratory*), Frank Mueller (*North Carolina State University*)

HPC systems are well known to endure service downtime due to increasing failures. With enhancements in HPC architectures and design, enabling resilience is extremely challenging due to component scaling and absence of well defined failure indicators. HPC system logs are notorious to be complex and unstructured. Efficient fault prediction to enable proactive recovery mechanisms is the need of the hour to make such systems more robust and reliable. This work addresses such faults in computing systems using a recurrent neural network based technique called LSTM (long short-term memory).

We present our framework Desh : Deep Learning for HPC System Health, which entails a procedure to diagnose and predict failures with acceptable lead times. Desh indicates prospects of indicating failure indicators with enhanced training and classification for generic applicability to other systems. This deep learning based framework gives interesting insights for further work on HPC system reliability.

P90: Global Survey of Energy and Power-Aware Job Scheduling and Resource Management in Supercomputing Centers

Authors: Siddhartha Jana (*Intel Corporation*), Gregory A. Koenig (*Energy Efficient HPC Working Group*), Matthias Maiterth (*Intel Corporation*), Kevin T. Pedretti (*Sandia National Laboratories*), Andrea Borghesi (*University of Bologna*), Andrea Bartolini (*ETH Zurich*), Bilel Hadri (*King Abdullah University of Science and Technology*), Natalie J. Bates (*Energy Efficient HPC Working Group*)

The two major driving forces that are leading centers to investigate dynamic power and energy management strategies are (1) limitations to the availability of resources from the electricity service provider, and (2) the desire to spend limited budgets on computational cycles rather than infrastructure requirements such as electricity. In addition, supercomputer systems have increasingly rapid, unpredictable, and large power fluctuations. In addition, electricity service providers may request supercomputing centers to change their timing and/or magnitude of demand to help address electricity supply constraints. To adapt to this new landscape, centers may employ energy and power-aware job scheduling and resource management (EPA-JSRM) strategies to dynamically, and in real-time, control their electricity demand. This poster summarizes the lessons learned from one of the first global surveys of supercomputing centers that are actively investigating such strategies.

P91: Assessing the Availability of Source Code in Computational Physics

Authors: Matthew Krafczyk (*National Center for Supercomputing Applications, University of Illinois*), Victoria Stodden (*University of Illinois*), Yantong Zheng (*University of Illinois*), David Wong (*University of Illinois*)

Replicability of scientific work based on computation is a subject which has been receiving increased scrutiny recently. One approach to replicating a computational finding is to run the original source code. Availability of source code however is not routine; Only 3/33 computationally based article authors released source code from JASA in 2006, and a 2015 study showed that only 44% of computer science article authors released their source code. The field of Computational Physics has yet to be examined in this

way, nor has the effect of author knowledge of such a study been measured.

We present our findings regarding the availability of source code in recent articles of the Journal of Computational Physics as well as how author knowledge of the study affects their willingness to make source code available. This work extends current reproducibility efforts being explored by the ACM, SIGHPC, and the SC conference community.

P92: Characterization and Comparison of Application Resilience for Serial and Parallel Executions

Authors: Kai Wu (*University of California, Merced*), Qiang Guan (*Los Alamos National Laboratory, Ultrascale Systems Research Center*), Nathan DeBardeleben (*Los Alamos National Laboratory, Ultrascale Systems Research Center*), Dong Li (*University of California, Merced*)

Soft error of exascale application is a challenge problem in modern HPC. In order to quantify an application's resilience and vulnerability, the application-level fault injection method is widely adopted by HPC users. However, it is not easy since users need to inject a large number of faults to ensure statistical significance, especially for parallel version program. Normally, parallel execution is more complex and requires more hardware resources than its serial execution. Therefore, it is essential that we can predict error rate of parallel application based on its corresponding serial version. In this poster, we characterize fault pattern in serial and parallel executions. We find first there are same fault sources in serial and parallel execution. Second, parallel execution also has some unique fault sources compared with serial executions. Those unique fault sources are important for us to understand the difference of fault pattern between serial and parallel executions.

P93: Spacehog: Evaluating the Costs of Dedicating Resources to In Situ Analysis

Authors: Rebecca Kreitinger (*University of New Mexico*), Scott Levy (*Sandia National Laboratories*), Kurt B. Ferreira (*Sandia National Laboratories*), Patrick Widener (*Sandia National Laboratories*)

Using in situ analytics requires that computational resources be shared between the simulation and the analysis. With space-sharing, there is a possibility for contention over these shared resources such as memory, memory bandwidth, network bandwidth, or filesystem bandwidth. In our analysis, we explore the sensitivity of different applications with a set of microbenchmarks that are representative of analytics that may be used with scientific simulation. These tasks are modeled using a library called libspacehog. The experimentation consisted of examining three different dimensions of how simulation workloads might be space-shared with analysis codes. The results indicate that contention does need to be considered when applying in situ analytic techniques and can be of greater concern than simply the number of analysis processes or overall process density. This research provides an explanation on how the application's performance is affected by space-sharing to further understand in situ analytic techniques.

P94: Fully Hierarchical Scheduling: Paving the Way to Exascale Workloads

Authors: Stephen Herbein (*University of Delaware*), Tapasya Patki (*Lawrence Livermore National Laboratory*), Dong H. Ahn (*Lawrence Livermore National Laboratory*), Don Lipari (*Lawrence Livermore National Laboratory*), Tamara Dahlgren (*Lawrence Livermore National Laboratory*), David Domyancic (*Lawrence Livermore National Laboratory*), Michela Taufer (*University of Delaware*)

Exascale workloads, such as uncertainty quantification (UQ), represent an order of magnitude increase in both scheduling scale and complexity. Batch schedulers with their decades-old, centralized scheduling model will fail to address the needs of these new workloads. To address these upcoming challenges, we claim that HPC schedulers must transition from the centralized to the fully hierarchical scheduling model. In this work, we assess the impact of the fully hierarchical model on both a synthetic stress test and a real-world UQ workload. We observe over a 100x increase in scheduler scalability on the synthetic stress test and a 37% decrease in the runtime of real-world UQ workloads under the fully hierarchical model. Our empirical results demonstrate that the fully hierarchical scheduling model can overcome the limitations of existing schedulers to meet the needs of UQ and other exascale workloads.

BP

P95: GEOPM: A Scalable Open Runtime Framework for Power Management

Authors: Siddhartha Jana (*Intel Corporation*), Asma H. Al-rawi (*Intel Corporation*), Steve S. Sylvester (*Intel Corporation*), Christopher M. Cantalupo (*Intel Corporation*), Brad Geltz (*Intel Corporation*), Brandon Baker (*Intel Corporation*), Jonathan M. Eastep (*Intel Corporation*)

The power scaling challenges associated with exascale systems is a well-known issue. In this work, we introduce the Global Extensible Open Power Manager (GEOPM): a tree-hierarchical, open source runtime framework we are contributing to the HPC community to foster increased collaboration and accelerated progress toward software-hardware co-designed energy management solutions that address exascale power challenges and improve performance and energy efficiency in current systems. Through its plugin extensible architecture, GEOPM enables rapid prototyping of new energy management strategies. Different plugins can be tailored to the specific performance or energy efficiency priorities of each HPC center. To demonstrate the potential of the framework, this work develops an example plugin for GEOPM. This power rebalancing plugin targets power-capped systems and improves efficiency by minimizing job time-to-solution within a power budget. Our results demonstrate up to 30% improvements in the time-to-solution of CORAL system procurement benchmarks on a Xeon Phi cluster.

P96: Correcting Detectable Uncorrectable Errors in Memory

Authors: Grzegorz Pawelczak (*University of Bristol*), Simon McIntosh-Smith (*University of Bristol*)

With the expected decrease in Mean Time Between Failures, Fault Tolerance has been identified as one of the major challenges for exascale computing. One source of faults are soft errors caused by cosmic rays, which can cause bit corruptions to the data held in memory. Current solutions for protection against these errors include Error Correcting Codes, which can detect and/or correct these errors. When an error that can be detected but not corrected occurs, a Detectable Uncorrectable Error (DUE) results, and unless checkpoint-restart is used, the system will usually fail. In our work we present a probabilistic method of correcting DUEs which occur in the part of the memory where the program instructions are stored. We devise a correction technique for DUEs for the ARM A64 instruction set which combines extended Hamming code with Cyclic Redundancy Check code to provide near 100% Successful Correction Rate of DUEs.

P97: Profile Guided Kernel Optimization for Individual Container Execution on Bare-Metal Container

Authors: Kuniyasu Suzuki (National Institute of Advanced Industrial Science and Technology), Hidetaka Koie (National Institute of Advanced Industrial Science and Technology), Ryousei Takano (National Institute of Advanced Industrial Science and Technology)

Container technologies become popular on supercomputers as well as in data centers. They use a container image as a package of an application, which makes easy to customize the computing environment. Unfortunately, they are not allowed to change the kernel. It means that an application cannot get the benefit of kernel optimization. Especially, Profile Guided Kernel Optimization (PGKO) is not applied.

Bare-Metal Container (BMC) tries to solve this problem. BMC utilizes remote machine management technologies (IPMI, Intel AMT, and WakeupOnLAN) to run a container image on a remote machine with a suitable Linux kernel. It enables to use PGKO easily, because the trial execution to get a profile and the optimized execution are executed automatically. Furthermore, BMC easily changes the target machine, and the user can compare the effects. We measured the performance of PGKO on big data workloads (Apache and Redis) on Xeon and i7 and found the difference.

P98: Energy Efficiency in HPC with Machine Learning and Control Theory

Authors: Connor Imes (University of Chicago), Steven Hofmeyr (Lawrence Berkeley National Laboratory), Henry Hoffmann (University of Chicago)

Performance and power management in HPC has historically favored a race-to-idle approach in order to complete applications as quickly as possible, but this is not energy-efficient on modern systems. As we move toward exascale and hardware over-provisioning, power management is becoming more critical than ever for HPC system administrators, opening the door for more balanced approaches to performance and power management. We propose two projects to address balancing application performance and system power consumption in HPC during application runtime, using closed loop feedback designs based on the Self-Aware Computing Model to observe, decide, and act.

P99: The Intersection of Big Data and HPC: Using Asynchronous Many Task Runtime Systems for HPC and Big Data

Authors: Joshua Daniel Suetterlein (Pacific Northwest National Laboratory), Joshua Landwehr (Trovares Inc), Andres Marquez (Pacific Northwest National Laboratory), Joseph Manzano (Pacific Northwest National Laboratory), Kevin Barker (Pacific Northwest National Laboratory), Guang Gao (University of Delaware)

Although the primary objectives of the HPC and Big data fields seem disparate, HPC is beginning to suffer from a growing size of its workloads and the limitation of its techniques to handle large amount of data. This places interesting research challenges for both HPC and Big Data on how to marriage both fields together. This poster presents a case study which uses Asynchronous Many Task Runtimes (AMTs) as an exploratory vehicle to highlight possible solutions to these challenges. AMTs presents the unique opportunity for better load balancing, reconfigurable schedulers and data layouts that can take advantage of introspection frameworks, and the ability to exploit a massive amount of concurrency. We use the Performance Open Community Runtime (P-OCR) as a vehicle to port MapReduce operators to the HPC realm. We conduct experiments with both strong and weak scaling experimental format using WordCount and TeraSort as our kernels.

Wednesday, November 15th

Room: Mile High Prefunction
8:30 am - 5:00 pm

Scientific Visualization & Data Analytics Showcase Posters

Visualization of Decision-Making Support (DMS) Information for Responding to a Typhoon-Induced Disaster

Authors: Dongmin Jang (Korea Institute of Science and Technology Information), Jin-Hee Yuk (Korea Institute of Science and Technology Information), Junghyun Park (Korea Institute of Science and Technology Information), Jooneun An (Korea Institute of Science and Technology Information), Minsu Joh (Korea Institute of Science and Technology Information)

A high-resolution coupled atmosphere, ocean, and inundation (flood) modeling and simulation system was developed for scientific, accurate, fast, and efficient forecasting of typhoon-induced disasters. This is based on the KISTI decision-making support system

(K-DMSS). Our prediction system consists of a typhoon, surge/wave, and flooding prediction and analysis systems (TPAS, SPAS, and FPAS). In this research, we simulated Typhoon 'CHABA' (1618), which was ranked third among the most intense tropical cyclones and was the most powerful typhoon in the Republic of Korea (South Korea) in 2016. The CHABA-induced storm surge and inundation were simulated using our prediction and analysis system. To understand intuitively the changes of physical phenomena and damage caused by the typhoon, numerical data sets produced by the prediction and analysis systems were visualized using VAPOR (Visualization and Analysis Platform for Ocean, atmosphere, and solar Researchers, and which is a part of the K-DMSS) as one of the visualization systems.

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: *Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)*

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Visualizations of a High-Resolution Global-Regional Nested, Ice-Sea-Wave Coupled Ocean Model System

Authors: *Kangyou Zhong (Sun Yat-Sen University), Danya Xu (Sun Yat-Sen University), Changsheng Chen (University of Massachusetts, Dartmouth; Sun Yat-Sen University), Yutong Lu (Sun Yat-Sen University), Wenjie Dong (Sun Yat-Sen University), Jiang Li (Sun Yat-Sen University)*

A multi-scale, global-regional nested ocean modeling system based on the unstructured grid Finite Volume Community Ocean Model (FVCOM) has been deployed on the Tianhe-2 supercomputer providing 24/7 marine forecasting since September 2016. The modeling system is part of the Sun Yat-Sen University Community Integrated Model (SYCIM) project for developing a new generation Earth System Model to explore the physical mechanisms of climate change. With a horizontal resolution up to ~17 m, this high-resolution modeling system can properly resolve the complex dynamical interactions of estuary, near shore coast, continental shelf, and deep ocean basins. The submitted animation shows the modeled global surface waves distribution pattern and propagation in the world ocean basins and in the South China Sea. Additionally, the variations of the global sea surface temperature, Arctic and Antarctic sea ice thickness and extension simulated by SYCIM is also presented. The animation can not only be used to visualize the big data for scientific research but also as a good example to demonstrate dynamical variations of the ocean to non-scientific communities. We hope this animation can help to arouse public attention on some issues such as global warming, polar sea ice melting, ocean environment, and marine hazards.

Milky Way Analogue Isolated Disk Galaxy

Authors: *Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois), Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)*

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: *David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Srivier (University of Illinois)*

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on

the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center, University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Room: Four Seasons Ballroom

8:30 am - 5:00 pm

Research Posters

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

8:30 am - 5:00 pm

ACM Student Research Competition

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

Room: 701

3:00 pm - 4:45 pm

ACM Student Research Competition: Presentations by Semi-Finalists

A13: Deep Learning with HPC Simulations for Extracting Hidden Signals: Detecting Gravitational Waves

Authors: Daniel George (National Center for Supercomputing Applications, University of Illinois)

We introduce Deep Filtering, new machine learning method for end-to-end time-series signal processing, which combines two deep one-dimensional convolutional neural networks for classification and regression to detect and characterize signals much weaker than the background noise. We trained this method with a novel curriculum learning scheme on data derived from HPC simulations and applied it for gravitational wave analysis specifically for mergers of black holes and demonstrated that it significantly outperforms conventional machine learning techniques, is far more efficient than matched-filtering, offering several orders-of-magnitude speed-up, allowing real-time processing of raw big data with minimal resources, and extends the range of detectable signals. This initiates a new paradigm for scientific research which employs massively-parallel numerical simulations to train artificial intelligence algorithms that exploit emerging hardware architectures such as deep-learning-optimized GPUs. Our approach offers a unique framework to enable coincident detection campaigns of gravitational wave sources and their electromagnetic counterparts.

SF

A01: GEMM-Like Tensor-Tensor Contraction (GETT)

Authors: Paul Springer (RWTH Aachen University), Paolo Bientinesi (RWTH Aachen University)

Tensor contractions (TC) are a performance critical component in numerous scientific computations. Despite the close connection between matrix-matrix products (GEMM) and TCs, the performance of the latter is in general vastly inferior to that of an optimized GEMM. To close such a gap, we propose a novel approach: GEMM-like Tensor-Tensor multiplication (GETT). GETT mimics the design of a high-performance GEMM implementation; as such, it systematically reduces an arbitrary tensor contractions to a highly-optimized "macro-kernel". This macro-kernel operates on suitably "packed" sub-tensors that reside in specified levels of the cache hierarchy. GETT's decisive feature is its ability to pack subtensors via tensor transpositions, yielding efficient packing routines. In contrast to previous approaches to TCs, GETT attains the same I/O cost as an equally-sized GEMM, making GETT especially well-suited for bandwidth-bound TCs. GETT's excellent performance is highlighted across a wide range of random tensor contractions.

SF

A04: Optimization of the AIREBO Many-Body Potential for KNL

Authors: Markus Höhnerbach (RWTH Aachen University)

Molecular dynamics simulations are an indispensable research tool for computational chemistry and material science. Empirical many-body potentials promise high-fidelity simulations that capture bonding and reaction behavior accurately, providing a level of detail in between more classical molecular dynamics and quantum methods.

The AIREBO potential is one such example that provides forces and energies for molecular dynamics (MD) simulations of carbon and carbohydrate structures. Allowing many-body potentials to profit from the recent architectural advances still poses a challenge due to deeply nested, short loops. We develop an optimized, vectorized AIREBO implementation for Intel's Xeon and Xeon Phi (co)processors and integrate it into the open-source LAMMPS molecular dynamics code. By both introducing improvements to the code and vectorization, we achieve a sustained real-word speedup of two on Broadwell, and a speedup of four on KNL. The optimized code will be distributed with each LAMMPS download as part of the USER-INTEL package.

SF

A08: Virtualized Big Data: Reproducing Simulation Output on Demand

Authors: Salvatore Di Girolamo (ETH Zurich)

Scientific simulations are being pushed to the extreme in terms of size and complexity of the addressed problems, producing astonishing amount of data. If the data is stored on disk, analysis applications can randomly access simulation output. Yet, storing the massive amounts simulation data is challenging. This is primarily due to the high storage costs and the fact that compute capabilities grow faster than storage capacities and bandwidths. In-situ analysis removes the storage costs but applications lose random access.

We propose to not store the full simulation output data but to produce it on demand. Our system intercepts I/O requests of both analysis tools and simulators, enabling data virtualization. This new paradigm allows us to explore the computation-storage tradeoff, by trading computation power for storage space. Overall, SDAVi offers a viable path towards exa-scale scientific simulations, by exploiting the growing computing power and relaxing the storage capacity requirements.

SF

A02: Accelerating the Higher Order Singular Value Decomposition Algorithm for Big Data with GPUs

Authors: Yuhsiang M. Tsai (National Taiwan University)

With the explosion of big data, finding ways of compressing large datasets with multi-way relationship - i.e., tensors - quickly and efficiently has become critical in HPC.

High-order singular value decomposition (HOSVD) method provides us with the means to attain both extremely high compression ratio and low error rate through low-rank approximation.

However, parallelizing HOSVD efficiently on GPUs remains a challenging problem, largely due to the lack of a fast SVD implementation that can stream data to the limited GPU memory through the PCIe bottleneck.

Our work studies, optimizes, and then contrasts four different methods for calculating singular vectors for performance, weak/strong scalability and accuracy in the context of HOSVD. We also discuss ways of load balancing the problem across multiple GPUs on a single node, and discuss the pros and cons of these different algorithms for GPU acceleration.

SF

Break

A16: Diagnosing Parallel I/O Bottlenecks in HPC Applications

Authors: *Peter Z. Harrington (University of California, Santa Cruz)*

HPC applications are generating increasingly large volumes of data (up to hundreds of TBs), which need to be stored in parallel to be scalable. Parallel I/O is a significant bottleneck in HPC applications, and is especially challenging in Adaptive Mesh Refinement (AMR) applications because the structure of output files changes dynamically during runtime. Data-intensive AMR applications run on the Cori supercomputer show variable and often poor I/O performance, but diagnosing the root cause remains challenging. Here we analyze logs from multiple levels of Cori's parallel I/O subsystems, and find bottlenecks during file metadata operations and during the writing of file contents that reduced I/O bandwidth by up to 40x. Such bottlenecks seemed to be system-dependent and not the application's fault. Increasing the granularity of file-system performance data will help provide conclusive causal relationships between file-system servers and metadata bottlenecks.

SF

A11: Finding a Needle in a Field of Haystacks: Lightweight Metadata Search for Large-Scale Distributed Research Repositories

Authors: *Anna Blue Keleher (University of Maryland)*

Fast, scalable, and distributed search services are commonly available for single nodes, but lead to high infrastructure costs when scaled across tens of thousands of filesystems and repositories, as is the case with Globus. Endpoint-specific indexes may instead be stored on their respective nodes, but while this distributes storage costs between users, it also creates significant query overhead. Our solution provides a compromise by introducing two levels of indexes: a single centralized "second-level index" (SLI) that aggregates and summarizes terms from each endpoint; and many endpoint-level indexes that are referenced by the SLI and used only when needed. We show, via experiments on Globus-accessible filesystems, that the SLI reduces the amount of space needed on central servers by over 96% while also reducing the set of endpoints that need to execute user queries.

SF

A12: Applying Image Feature Extraction to Cluttered Scientific Repositories

Authors: *Emily Herron (Mercer University)*

Over time many scientific repositories and file systems become disorganized, containing poorly described and error-ridden data. As a result, it is often difficult for researchers to discover crucial data. In this poster, we present a collection of image processing modules that collectively extract metadata from a variety of image formats. We implement these modules in Skluma—a system designed to automatically extract metadata from structured and semi-structured scientific formats. Our modules apply several image metadata extraction techniques that include processing file system metadata, header information, color content statistics, extracted text, feature-based clusters, and predicting tags using a supervised learning model. Our goal is to collect a large number of metadata that may then be used to organize, understand, and analyze data stored in a repository.

SF

A18: Understanding the Impact of Fat-Tree Network Locality on Application Performance

Authors: *Philip Taffet (Rice University)*

Network congestion can be a significant cause of performance loss and variability for many message passing programs. However, few studies have used a controlled environment with virtually no other extraneous sources of network traffic to observe the impact of application placement and multi-job interactions on overall performance. We study different placements and pairings for three DOE applications. We observe that for a job size typical for an LLNL commodity cluster, the impact of congestion and poor placement is typically less than 2%, which is less dramatic than on torus networks. In addition, in most cases, the cyclic MPI task mapping strategy increases performance and reduces placement sensitivity despite also increasing total network traffic. We also found that the performance difference between controlled placements and runs scheduled through the batch system was less than 3%.

SF

A03: A High-Speed Algorithm for Genome-Wide Association Studies on Multi-GPU Systems

Authors: *Yen Chen Chen (National Taiwan University)*

We develop an algorithm as long as a CUDA code for GWAS (Genome-Wide Associate Studies). This algorithm can work efficiently on GPU and has high scalability. The core of the algorithm is an accurate and fast p-value integration reformation, which accelerates the most time-consuming part of the algorithm. With the algorithm, researchers can now deal with tens of billions of SNP to trait pair in only a few minutes. Even better, since this algorithm is highly scalable, you can increase the problem size as long as you have enough computing power.

SF

Thursday, November 16th

Room: Mile High Prefunction
8:30 am - 5:00 pm

Scientific Visualization & Data Analytics Showcase Posters

Visualization of Decision-Making Support (DMS) Information for Responding to a Typhoon-Induced Disaster

Authors: Dongmin Jang (Korea Institute of Science and Technology Information), Jin-Hee Yuk (Korea Institute of Science and Technology Information), Junghyun Park (Korea Institute of Science and Technology Information), Jooneun An (Korea Institute of Science and Technology Information), Minsu Joh (Korea Institute of Science and Technology Information)

A high-resolution coupled atmosphere, ocean, and inundation (flood) modeling and simulation system was developed for scientific, accurate, fast, and efficient forecasting of typhoon-induced disasters. This is based on the KISTI decision-making support system (K-DMSS). Our prediction system consists of a typhoon, surge/wave, and flooding prediction and analysis systems (TPAS, SPAS, and FPAS). In this research, we simulated Typhoon 'CHABA' (1618), which was ranked third among the most intense tropical cyclones and was the most powerful typhoon in the Republic of Korea (South Korea) in 2016. The CHABA-induced storm surge and inundation were simulated using our prediction and analysis system. To understand intuitively the changes of physical phenomena and damage caused by the typhoon, numerical data sets produced by the prediction and analysis systems were visualized using VAPOR (Visualization and Analysis Platform for Ocean, atmosphere, and solar Researchers, and which is a part of the K-DMSS) as one of the visualization systems.

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Visualizations of a High-Resolution Global-Regional Nested, Ice-Sea-Wave Coupled Ocean Model System

Authors: Kangyou Zhong (Sun Yat-Sen University), Danya Xu (Sun Yat-Sen University), Changsheng Chen (University of Massachusetts, Dartmouth; Sun Yat-Sen University), Yutong Lu (Sun Yat-Sen University), Wenjie Dong (Sun Yat-Sen University), Jiang Li (Sun Yat-Sen University)

A multi-scale, global-regional nested ocean modeling system based on the unstructured grid Finite Volume Community Ocean Model (FVCOM) has been deployed on the Tianhe-2 supercomputer providing 24/7 marine forecasting since September 2016. The modeling system is part of the Sun Yat-Sen University Community Integrated Model (SYCIM) project for developing a new generation Earth System Model to explore the physical mechanisms of climate change. With a horizontal resolution up to ~17 m, this high-resolution modeling system can properly resolve the complex dynamical interactions of estuary, near shore coast, continental shelf, and deep ocean basins. The submitted animation shows the modeled global surface waves distribution pattern and propagation in the world ocean basins and in the South China Sea. Additionally, the variations of the global sea surface temperature, Arctic and Antarctic sea ice thickness and extension simulated by SYCIM is also presented. The animation can not only be used to visualize the big data for scientific research but also as a good example to demonstrate dynamical variations of the ocean to non-scientific communities. We hope this animation can help to arouse public attention on some issues such as global warming, polar sea ice melting, ocean environment, and marine hazards.

Milky Way Analogue Isolated Disk Galaxy

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois), Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Sriver (University of Illinois)

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center; University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Room: Four Seasons Ballroom

8:30 am - 5:00 pm

Research Posters

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

8:30 am - 5:00 pm

ACM Student Research Competition

SC17 Research Posters

SC17 Research Posters will be on display on Tuesday, Wednesday, Thursday from 8:30am to 5pm in the Four Seasons Ballroom.

Reception [\(back to top\)](#)

Sunday, November 12th

Room: McNichols Civic Center
6:00 pm - 9:00 pm

Exhibitors' Reception

Exhibitors' Reception

SC17 will host an Exhibitor Party at the McNichols Civic Center (144 West Colfax Avenue) for registered exhibitors. The party is SC17's way of thanking exhibitors for their participation and support of the conference. The event will include entertainment along with food and drinks.

The McNichols Civic Center Building is located in the heart of downtown Denver at the northwest corner of Civic Center Park. In 1909, the cornerstone of the building was laid, setting the foundation for the then Carnegie Library that would become a center of learning in the park. That tradition continues, as the building was re-opened in 2012 as a contemporary hub for arts, culture and events for the people of Denver. This stunning Greek Revival building with its classic Corinthian columns and iconic colonnade across its front, offers new experiences in a classic space.

The McNichols Civic Center Building is within walking distance (@ .6 mile) from the Colorado Convention Center. Limited busing will be provided to/from the Convention Center. An Exhibitor badge is required to attend this event. Guest tickets may be purchased in advance at the Registration desk (no tickets will be available at the party).

Monday, November 13th

Room: Exhibit Halls A, B, E and F
7:00 pm - 9:00 pm

Gala Opening Reception

Gala Opening Reception

SC17 will host its annual Grand Opening Gala in the Exhibit Hall. This will be your first opportunity to see the latest high performance computing, networking, storage, analysis, and research products, services, and innovations. This event is open to all Technical Program, Exhibitors and Students@SC registrants.

Tuesday, November 14th

Room: Four Seasons Ballroom
5:15 pm - 7:00 pm

Poster Reception

A01: GEMM-Like Tensor-Tensor Contraction (GETT)

Authors: Paul Springer (RWTH Aachen University), Paolo Bientinesi (RWTH Aachen University)

Tensor contractions (TC) are a performance critical component in numerous scientific computations. Despite the close connection between matrix-matrix products (GEMM) and TCs, the performance of the latter is in general vastly inferior to that of an optimized GEMM. To close such a gap, we propose a novel approach: GEMM-like Tensor-Tensor multiplication (GETT). GETT mimics the design of a high-performance GEMM implementation; as such, it systematically reduces an arbitrary tensor contractions to a highly-optimized "macro-kernel". This macro-kernel operates on suitably "packed" sub-tensors that reside in specified levels of the cache hierarchy. GETT's decisive feature is its ability to pack sub-tensors via tensor transpositions, yielding efficient packing routines. In contrast to previous approaches to TCs, GETT attains the same I/O cost as an equally-sized GEMM, making GETT especially well-suited for bandwidth-bound TCs. GETT's excellent performance is highlighted across a wide range of random tensor contractions.

SF

A02: Accelerating the Higher Order Singular Value Decomposition Algorithm for Big Data with GPUs

Authors: Yuhsiang M. Tsai (National Taiwan University)

With the explosion of big data, finding ways of compressing large datasets with multi-way relationship - i.e., tensors - quickly and efficiently has become critical in HPC.

High-order singular value decomposition (HOSVD) method provides us with the means to attain both extremely high compression ratio and low error rate through low-rank approximation.

However, parallelizing HOSVD efficiently on GPUs remains a challenging problem, largely due to the lack of a fast SVD implementation that can stream data to the limited GPU memory through the PCIe bottleneck.

Our work studies, optimizes, and then contrasts four different methods for calculating singular vectors for performance, weak/strong scalability and accuracy in the context of HOSVD. We also discuss ways of load balancing the problem across multiple GPUs on a single node, and discuss the pros and cons of these different algorithms for GPU acceleration.

SF

A03: A High-Speed Algorithm for Genome-Wide Association Studies on Multi-GPU Systems

Authors: Yen Chen Chen (National Taiwan University)

We develop an algorithm as long as a CUDA code for GWAS (Genome-Wide Associate Studies). This algorithm can work efficiently on GPU and has high scalability. The core of the algorithm is an accurate and fast p-value integration reformation, which accelerates the most time-consuming part of the algorithm. With the algorithm, researchers can now deal with tens of billions of SNP to trait pair in only a few minutes. Even better, since this algorithm is highly scalable, you can increase the problem size as long as you have enough computing power.

SF

A04: Optimization of the AIREBO Many-Body Potential for KNL

Authors: Markus Höhnnerbach (RWTH Aachen University)

Molecular dynamics simulations are an indispensable research tool for computational chemistry and material science. Empirical many-body potentials promise high-fidelity simulations that capture bonding and reaction behavior accurately, providing a level of detail in between more classical molecular dynamics and quantum methods.

The AIREBO potential is one such example that provides forces and energies for molecular dynamics (MD) simulations of carbon and carbohydrate structures. Allowing many-body potentials to profit from the recent architectural advances still poses a challenge due to deeply nested, short loops. We develop an optimized, vectorized AIREBO implementation for Intel's Xeon and Xeon Phi (co)processors and integrate it into the open-source LAMMPS molecular dynamics code. By both introducing improvements to the code and vectorization, we achieve a sustained real-word speedup of two on Broadwell, and a speedup of four on KNL. The optimized code will be distributed with each LAMMPS download as part of the USER-INTEL package.

SF

A05: Parallel Prefix Algorithms for the Registration of Arbitrarily Long Electron Micrograph Series

Authors: Marcin Copik (RWTH Aachen University)

Recent advances in the technology of transmission electron microscopy have allowed for a more precise visualization of materials and physical processes, such as metal oxidation. Nevertheless, the quality of information is limited by the damage caused by an electron beam, movement of the specimen or other environmental factors. A novel registration method has been proposed to remove those limitations by acquiring a series of low dose microscopy frames and performing a computational registration on them to understand and visualize the sample. This process can be represented as a prefix sum with a complex and computationally intensive binary operator and a parallelization is necessary to enable processing long series of microscopy images. With our parallelization scheme, the time of registration of results from ten seconds of microscopy acquisition has been decreased from almost thirteen hours to less than seven minutes on 512 Intel IvyBridge cores.

A06: Accelerating Big Data Processing in the Cloud with Scalable Communication and I/O Schemes

Authors: Shashank Gugnani (Ohio State University)

With the advent of cloud computing, the field of Big Data has seen rapid growth. Most cloud providers provide hardware resources such as NVMe SSDs, large memory nodes, and SR-IOV. This opens up the possibility of large-scale high-performance data analytics and provides opportunities to use these resources to develop new designs. Cloud computing provides flexibility, scalability, and reliability, which are important requirements of Big Data frameworks. However, several important requirements are missing, such as performance, scalability, fault-tolerance, and consistency. The focus of this research work revolves around developing communication and I/O designs and concepts which can provide these requirements to Big Data frameworks. Specifically, we explore new ways to provide fault-tolerance and consistency in cloud storage systems, providing scalable and high-performance communication frameworks, and co-designing with Big Data stacks to leverage these features.

A07: Scalable Parallel Scripting in the Cloud

Authors: Benjamin H. Glick (Lewis & Clark College)

It's often complicated, time consuming, but frequently necessary to successfully port complex workflows to multiple high-performance environments. Parsl is a Python-based parallel scripting library that provides a simple model for describing and executing dataflow-based scripts over arbitrary execution resources such as clouds, campus clusters, and high-performance systems. Parsl's execution layer abstracts the differences between providers enabling provisioning and management of compute

nodes for use with a pilot system. In this poster, we describe the development of a new execution provider designed to support Amazon Web Services (AWS) and Microsoft's Azure. This provider supports the transparent execution of implicitly parallel Python-based scripts using elastic cloud resources. We demonstrate that Parsl is capable of executing thousands of applications per second over this elastic execution fabric.

A08: Virtualized Big Data: Reproducing Simulation Output on Demand

Authors: *Salvatore Di Girolamo (ETH Zurich)*

Scientific simulations are being pushed to the extreme in terms of size and complexity of the addressed problems, producing astonishing amount of data. If the data is stored on disk, analysis applications can randomly access simulation output. Yet, storing the massive amounts simulation data is challenging. This is primarily due to the high storage costs and the fact that compute capabilities grow faster than storage capacities and bandwidths. In-situ analysis removes the storage costs but applications lose random access.

We propose to not store the full simulation output data but to produce it on demand. Our system intercepts I/O requests of both analysis tools and simulators, enabling data virtualization. This new paradigm allows us to explore the computation-storage tradeoff, by trading computation power for storage space. Overall, SDaVi offers a viable path towards exa-scale scientific simulations, by exploiting the growing computing power and relaxing the storage capacity requirements.

SF

A09: Ring: Unifying Replication and Erasure Coding to Rule Resilience in KV-Stores

Authors: *Konstantin Taranov (ETH Zurich)*

There is a wide range of storage schemes employed by KV-stores to ensure reliability of stored keys. However, previous implementations do not allow choosing the storage scheme dynamically, thereby forcing developers to commit to a single scheme. Such inefficient data management wastes cluster resources such as memory usage, network load, latency, availability and many others.

To solve this problem, we have designed a strongly consistent key-value store Ring that empowers its users to explicitly manage storage parameters like other resources, such as memory or processor time. The key feature of Ring is that all keys live in the same strongly consistent namespace and a user does not need to specify the resilience when looking up a key or value.

Our poster demonstrates how future applications that manage resilience of key-value pairs consciously can reduce the overall operational cost and improve performance significantly.

A10: Revealing the Power of Neural Networks to Capture Accurate Job Resource Usage from Unparsed Job Scripts and Application Inputs

Authors: *Michael R. Wyatt (University of Delaware)*

Next generation HPC schedulers will rely heavily on accurate information about resource usage of submitted jobs. The information provided by users is often inaccurate and previous prediction models, which rely on parsed job script features, fail to accurately predict for all HPC jobs. We propose a new representation of job scripts and inclusion of application input decks for resource usage predictions with a neural network. Our contributions are a method for representing job scripts as image-like data, an automated method for predicting job resource usage from job script images and input deck features, and validation of our methods with real HPC data. We demonstrate that when job scripts for an application are very similar, our method performs better than other methods. We observe an average decrease in error of 2 node-hours compared to state of the art methods.

A11: Finding a Needle in a Field of Haystacks: Lightweight Metadata Search for Large-Scale Distributed Research Repositories

Authors: *Anna Blue Keleher (University of Maryland)*

Fast, scalable, and distributed search services are commonly available for single nodes, but lead to high infrastructure costs when scaled across tens of thousands of filesystems and repositories, as is the case with Globus. Endpoint-specific indexes may instead be stored on their respective nodes, but while this distributes storage costs between users, it also creates significant query overhead. Our solution provides a compromise by introducing two levels of indexes: a single centralized "second-level index" (SLI) that aggregates and summarizes terms from each endpoint; and many endpoint-level indexes that are referenced by the SLI and used only when needed. We show, via experiments on Globus-accessible filesystems, that the SLI reduces the amount of space needed on central servers by over 96% while also reducing the set of endpoints that need to execute user queries.

SF

A12: Applying Image Feature Extraction to Cluttered Scientific Repositories

Authors: *Emily Herron (Mercer University)*

Over time many scientific repositories and file systems become disorganized, containing poorly described and error-ridden data. As a result, it is often difficult for researchers to discover crucial data. In this poster, we present a collection of image processing modules that collectively extract metadata from a variety of image formats. We implement these modules in Skluma—a system designed to automatically extract metadata from structured and semi-structured scientific formats. Our modules apply several image

metadata extraction techniques that include processing file system metadata, header information, color content statistics, extracted text, feature-based clusters, and predicting tags using a supervised learning model. Our goal is to collect a large number of metadata that may then be used to organize, understand, and analyze data stored in a repository.

SF

A13: Deep Learning with HPC Simulations for Extracting Hidden Signals: Detecting Gravitational Waves

Authors: *Daniel George (National Center for Supercomputing Applications, University of Illinois)*

We introduce Deep Filtering, new machine learning method for end-to-end time-series signal processing, which combines two deep one-dimensional convolutional neural networks for classification and regression to detect and characterize signals much weaker than the background noise. We trained this method with a novel curriculum learning scheme on data derived from HPC simulations and applied it for gravitational wave analysis specifically for mergers of black holes and demonstrated that it significantly outperforms conventional machine learning techniques, is far more efficient than matched-filtering, offering several orders-of-magnitude speed-up, allowing real-time processing of raw big data with minimal resources, and extends the range of detectable signals. This initiates a new paradigm for scientific research which employs massively-parallel numerical simulations to train artificial intelligence algorithms that exploit emerging hardware architectures such as deep-learning-optimized GPUs. Our approach offers a unique framework to enable coincident detection campaigns of gravitational wave sources and their electromagnetic counterparts.

SF

A14: Analysis of Synthetic Graph Generation Methods for Directed Network Graphs

Authors: *Spencer Callicott (Mississippi State University)*

Historically, scientific experiments have been conducted to generate scale-free network graphs based on structure. Metrics used to measure veracity ensure the integrity of a scale-free algorithm given a seed. However, studies do not explore the performance benefits or drawbacks of specific algorithms running on Apache Spark and GraphX. Recognizing the lack of performance benchmarks demands ensuring accuracy through experimenting. This study will utilize the Stochastic Kronecker Graph model to synthetically generate graphs given a seed graph.

A15: Quantifying Compiler Effects on Code Performance and Reproducibility Using FLiT

Authors: *Michael Bentley (University of Utah)*

A busy application developer likes to focus on doing science, but instead is often distracted by the sheer variety of available hardware platforms, their compilers, and associated optimization flags. Exclusive pursuit of speed may jeopardize the reproducibility of scientific experiments. On the other hand, performance is central to HPC. Our previous work provided a unique testing framework called FLiT that helps developers exploit performance without jeopardizing reproducibility. To verify that FLiT is useful for real-world libraries and applications, it was applied to MFEM, a finite element library used in various HPC applications. I show that the compilation with the fastest average runtime for the converted MFEM examples is also bitwise reproducible. For these examples, clang had the fastest average runtimes and the best reproducibility. Our future work aims to enhance the open-source FLiT tool into a strong community resource and to follow up with found compiler oddities.

A16: Diagnosing Parallel I/O Bottlenecks in HPC Applications

Authors: *Peter Z. Harrington (University of California, Santa Cruz)*

HPC applications are generating increasingly large volumes of data (up to hundreds of TBs), which need to be stored in parallel to be scalable. Parallel I/O is a significant bottleneck in HPC applications, and is especially challenging in Adaptive Mesh Refinement (AMR) applications because the structure of output files changes dynamically during runtime. Data-intensive AMR applications run on the Cori supercomputer show variable and often poor I/O performance, but diagnosing the root cause remains challenging. Here we analyze logs from multiple levels of Cori's parallel I/O subsystems, and find bottlenecks during file metadata operations and during the writing of file contents that reduced I/O bandwidth by up to 40x. Such bottlenecks seemed to be system-dependent and not the application's fault. Increasing the granularity of file-system performance data will help provide conclusive causal relationships between file-system servers and metadata bottlenecks.

SF

A17: Toward Capturing Nondeterminism Motifs in HPC Applications

Authors: *Dylan Chapp (University of Delaware)*

High performance MPI applications employ nondeterministic asynchronous communication to achieve greater performance. However, this nondeterminism can significantly hamper debugging. Various software tools have been developed to control nondeterminism in HPC applications, but a high-level application-agnostic taxonomy for this nondeterminism is absent and limits these tools' effectiveness in practice. We propose to address this need by extracting common nondeterministic communication motifs from representative applications.

We present a first step toward capturing nondeterminism motifs by way of a workflow for detecting and summarizing sender nondeterminism in HPC applications.

A18: Understanding the Impact of Fat-Tree Network Locality on Application Performance

Authors: *Philip Taffet (Rice University)*

Network congestion can be a significant cause of performance loss and variability for many message passing programs. However, few studies have used a controlled environment with virtually no other extraneous sources of network traffic to observe the impact of application placement and multi-job interactions on overall performance. We study different placements and pairings for three DOE applications. We observe that for a job size typical for an LLNL commodity cluster, the impact of congestion and poor placement is typically less than 2%, which is less dramatic than on torus networks. In addition, in most cases, the cyclic MPI task mapping strategy increases performance and reduces placement sensitivity despite also increasing total network traffic. We also found that the performance difference between controlled placements and runs scheduled through the batch system was less than 3%.

SF

A19: Performance Analysis of a Parallelized Restricted Boltzmann Machine Artificial Neural Network Using OpenACC Framework and TAU Profiling System

Authors: *Abhishek Kumar (Brookhaven National Laboratory)*

Restricted Boltzmann Machines are stochastic neural networks that create probability distribution based off connection weight between nodes of the hidden and visible layer. The distribution makes the program optimal at classifying large amounts of data, which could be useful in work settings, such as a research lab. The parallelization of these neural networks would allow for the classification of data at a much faster rate than before. Using a high-performance computer it was determined that parallelizing the neural networks could decrease the runtime of the algorithm by over 35% when offloading the work to a GPU through OpenACC. Using Tuning and Analysis Utilities Profiling Systems, it was found that scheduling the program would only be effective if the data size was large enough and an increase in the number of thread blocks used for scheduling would allow for greater performance gains than the number of threads in each thread block.

A20: Correctness Verification and Boundary Conditions for Chapel Iterator-Based Loop Optimization

Authors: *Daniel A. Feshbach (Haverford College)*

We explore two issues of correctness concerning iteration space transformation techniques: data dependencies and boundary conditions. First, we present a data structure which automatically verifies correctness of data dependencies for stencil computations with transformed iteration spaces. This further confirms the viability of Chapel iterators for defining iteration space transformations, by demonstrating that simple tool support can verify data dependencies and assist debugging. Second, we explore the performance and simplicity of three strategies for implementing boundary conditions in transformed iteration spaces: if statements, loop peeling, and an array of coefficients. We find that the coefficient array technique performs the best, often at 70 to 80 percent speed of the benchmark of ignoring the boundary condition. If statements are not far behind, while loop peeling performs much worse. The coefficient array and if statements are indifferent to the transformation technique applied, while loop peeling must be implemented within the transformation.

A21: Runtime Support for Concurrent Execution of Overdecomposed Heterogeneous Tasks

Authors: *Jaemin Choi (University of Illinois)*

With the rise of heterogeneous systems in high performance computing, how we utilize accelerators has become a critical factor in achieving the optimal performance. We explore several issues with using accelerators in Charm++, a parallel programming model that employs overdecomposition. We propose a runtime support scheme that enables concurrent execution of heterogeneous tasks and evaluate its performance. Using a synthetic benchmark that utilizes busy-waiting to simulate workload, we observe that the effectiveness of the runtime support varies with the application characteristics, with a maximum speedup of 4.79x. With a two-dimensional five-point stencil benchmark designed to represent a realistic workload, we obtain up to 2.75x speedup.

A22: Verifying Functional Equivalence Between C and Fortran Programs

Authors: *Wenhao Wu (University of Delaware)*

Software verification is a mature research area with many techniques. These verification approaches can be applied to programs written in different programming languages; nevertheless, most verification tools are only designed for programs written in C or Java. As a result, verification tools are inadequate for other languages, such as Fortran. A high level of software safety is mandatory in most of its application scenarios, which makes verification tools for Fortran programs necessary and significant.

In this poster, the author illustrates the motivation and objectives of the project with examples. Also, this poster shows an extension (as a Fortran program verifier) of an existing verification platform -- CIVL. Additionally, the results of a set of extensive experiments conducted by the author is shown in this poster to indicate that the performance is satisfactory.

A23: Evaluation of Data-Intensive Applications on Intel Knights Landing Cluster

Authors: *Tao Gao (University of Delaware)*

Analyzing and understanding large datasets on high performance computing platforms is becoming more and more important in various scientific domains. MapReduce is the dominant programming model for processing these datasets. Platforms for data processing are empowered by many-core nodes with cutting-edge processing units. Intel Knights Landing (KNL) is the new arrival in the field. However, this new architecture has not been fully evaluated for data-intensive applications. In this poster, we present the

assess of KNL on the performance of three key data-intensive applications based on a high-performance MapReduce programming framework on the latest KNL-cluster, Stampede2. We focus on the impact of different KNL memory models, we compare Stampede2 with other clusters such as Tianhe-2 and Mira, and we measure the scalability of large datasets. We observe how KNL-based clusters are a promising architecture for data-intensive applications. We also identify key aspects to enable more efficient usage of KNL-based clusters.

A24: Comparison of Machine Learning Algorithms and Their Ensembles for Botnet Detection

Authors: *Songhui Ryu (Purdue University)*

A Botnet is a network of compromised devices that is controlled by malicious ‘botmaster’ in order to perform various tasks, such as executing DoS attack, sending SPAM and obtaining personal data etc. As botmasters generate network traffic while communicating with their bots, analyzing network traffic to detect Botnet traffic can be a promising feature of Intrusion Detection System(IDS). Although IDS has been applying various machine learning (ML) techniques, comparison of ML algorithms including their ensembles on Botnet detection has not been figured out yet. In this study, not only the three most popular classification ML algorithms – Naïve Bayes, Decision tree, and Neural network are evaluated, but also the ensemble methods known to strengthen ML algorithms are tested to see if they indeed provide enhanced predictions on Botnet detection. This evaluation is conducted with CTU-13 public dataset, measuring running time of each ML and its f measure and MCC score.

A25: Investigating Performance of Serialization Methods for Networked Data Transfer in HPC Applications

Authors: *Max Yang (Georgia Institute of Technology)*

Cluster-to-user data transfers present challenges with cross-platform endianness (byte-order) compatibility and handling a variety of numeric types, and may occur over suboptimal network links. Two serialization libraries, Protocol Buffers and Conduit, were selected for their ability to handle endianness and their cross-language support, and their performance in both size and speed was measured. It was found that the throughput of Protocol Buffers was significantly more than that of Conduit while exhibiting less protocol overhead. Adding a compression stage after serialization dramatically reduced the size of messages on certain types of data, but had some impact on throughput.

A26: Co-Designing MPI Runtimes and Deep Learning Frameworks for Scalable Distributed Training on GPU Clusters

Authors: *Ammar Ahmad Awan (Ohio State University)*

Deep Learning frameworks like Caffe, TensorFlow, and CNTK have brought forward new requirements and challenges for communication runtimes like MVAPICH2-GDR. These include support for low-latency and high-bandwidth communication of very-large GPU-resident buffers. This support is essential to enable scalable distributed training of Deep Neural Networks on GPU clusters. However, current MPI runtimes have limited support for large-message GPU-based collectives. To address this, we propose the S-Caffe framework; a co-design of distributed training in Caffe and large-message collectives in MVAPICH2-GDR. We highlight two designs for MPI_Bcast, one that exploits NVIDIA NCCL and the other that exploits ring-based algorithms. Further, we present designs for MPI_Reduce that provide up-to 2.5X improvement. We also present layer-wise gradient aggregation designs in S-Caffe that exploit overlap of computation and communication as well as the proposed reduce design. S-Caffe provides a scale-out to 160 GPUs for GoogLeNet training and delivers performance comparable to CNTK for AlexNet training.

A27: High-Performance and Scalable Broadcast Schemes for Deep Learning on GPU Clusters

Authors: *Ching-Hsiang Chu (Ohio State University)*

Broadcast operations are a widely used operation in many streaming and deep learning applications to disseminate large amounts of data on emerging heterogeneous High-Performance Computing (HPC) systems. Further, traditional broadcast schemes are not well optimized for upcoming large-scale Graphics Processing Unit (GPU)-based systems. However, utilizing cutting-edge features of modern HPC technologies such like InfiniBand (IB) and NVIDIA GPUs to enable scalable heterogeneous broadcast operations remains an open challenge.

Toward delivering the best performance for streaming and deep learning workloads, we propose high-performance and scalable broadcast schemes that exploit IB hardware multicast (IB-MCAST) and NVIDIA GPUDirect technology. We present experimental results and find that they indicate improved scalability and up to 68% reduction of latency compared to the state-of-the-art solutions in the benchmark-level evaluation. Furthermore, the proposed design yields up to 24% performance improvement for the popular deep learning framework, Microsoft cognitive toolkit (CNTK), with no application changes.

A28: Exploring Use Cases for Non-Volatile Memories in Support of HPC Resilience

Authors: *Onkar Patil (North Carolina State University)*

Improving resilience and creating resilient architectures is one of the major goals of exascale computing. With the advent of Non-volatile memory technologies, memory architectures with persistent memory regions will be a significant part of future architectures. There is potential to use them in more than one way to benefit different applications. We look to take advantage of this technology to enable more fine-grained and novel methodology that will improve resilience and efficiency of exascale applications. We have developed three modes of memory usage for persistent memory to enable efficient checkpointing in HPC applications. We have developed a simple API that is evaluated with the DGEMM benchmark on a 16-node cluster with independent SSDs on every node.

Our aim is to build on this work and enable static and dynamic runtime systems that will inherently make the HPC applications more fault-tolerant and resistant to errors.

P01: Cache-Blocking Tiling of Large Stencil Codes at Runtime

Authors: Istvan Z. Reguly (Pazmany Peter Catholic University), Gihan R. Mudalige (University of Warwick), Mike B. Giles (University of Oxford)

Stencil codes on structured meshes are well-known to be bound by memory bandwidth. Previous research has shown that compiler techniques that reorder loop schedules to improve temporal locality across loop nests, such as tiling, work particularly well. However in large codes the scope of such analysis is limited by the large number of code paths, compilation units, and run-time parameters. We present how, through run-time analysis of data dependencies across stencil loops enables the OPS domain specific language to tile across a large number of different loops. This lets us tackle much larger applications than previously studied: we demonstrate 1.7-3.5x performance improvement on CloverLeaf 2D, CloverLeaf 3D, TeaLeaf and OpenSBLI, tiling across up to 650 subsequent loopnests accessing up to 30 different state variables per gridpoint with up to 46 different stencils. We also demonstrate excellent strong and weak scalability of our approach on up to 4608 Broadwell cores.

P02: Strassen's Algorithm for Tensor Contraction

Authors: Jianyu Huang (University of Texas), Devin A. Matthews (University of Texas), Robert A. van de Geijn (University of Texas)

Tensor contraction(TC) is an important computational kernel widely used in numerous applications. It is a multi-dimensional generalization of matrix multiplication(GEMM). While Strassen's algorithm for GEMM is well studied in theory and practice, extending it to accelerate TC has not been previously pursued. Thus, we believe this to be the first work to demonstrate how one can in practice speed up tensor contraction with Strassen's algorithm. By adopting a Block-Scatter-Matrix format, a novel matrix-centric tensor layout, we can conceptually view TC as GEMM for general stride storage, with an implicit tensor-to-matrix transformation. This insight enables us to tailor a recent state-of-the-art implementation of Strassen's algorithm to TC, avoiding explicit transpositions(permutations) and extra workspace, and reducing the overhead of memory movement that is incurred. Performance benefits are demonstrated with a performance model as well as in practice on modern single core, multicore, and distributed memory parallel architectures, achieving up to 1.3x speedup.

P03: BEM4I: A Massively Parallel Boundary Element Solver

Authors: Michal Merta (Technical University of Ostrava), Jan Zapletal (Technical University of Ostrava), Michal Kravcenko (Technical University of Ostrava), Lukas Maly (Technical University of Ostrava)

In this work we present a library of parallel solvers based on the boundary element method (BEM). We provide a brief description of BEM and its parallelization, focus on SIMD vectorization and shared- and distributed-memory parallelization by OpenMP and MPI, respectively. Two approaches for distributed parallelization of BEM are discussed - the first one based on a novel parallel adaptive cross approximation (ACA) method, the second one on the boundary element tearing and interconnecting (BETI) domain decomposition method. To demonstrate the efficiency of the library we provide results of numerical experiments on the Xeon and Xeon Phi based clusters.

BP

P04: Unstructured-Grid CFD Algorithms on Many-Core Architectures

Authors: Aaron Walden (NASA Langley Research Center), Eric J. Nielsen (NASA Langley Research Center), Mohammad Zubair (Old Dominion University), John C. Linford (ParaTools), John G. Wohlbier (Engility Corporation), Justin P. Luitjens (Nvidia Corporation), Jason Orender (Old Dominion University), Izaak Beekman (ParaTools), Samuel Khuvis (ParaTools), Sameer S. Shende (ParaTools)

In the field of computational fluid dynamics (CFD), the Navier-Stokes equations are often solved using an unstructured-grid approach to accommodate geometric complexity. Furthermore, turbulent flows encountered in aerospace applications generally require highly anisotropic meshes, driving the need for implicit solution methodologies to efficiently solve the discrete equations. These approaches require frequent construction and solution of large, tightly-coupled systems of block-sparse linear equations.

We explore the transition of two representative CFD kernels from a coarse-grained MPI-based model originally developed for multi-core systems to a shared-memory model suitable for many-core platforms. Results for the Intel Xeon Phi Knights Landing, NVIDIA Pascal P100, and NVIDIA Volta V100 architectures are compared with the aforementioned MPI-based implementation for the multi-core Intel Xeon Broadwell (BWL) processor. We observe substantial speedups over BWL as well as higher performance per dollar MSRP and performance per watt for the many-core architectures.

P05: ooc_cuDNN : A Deep Learning Library Supporting CNNs over GPU Memory Capacity

Authors: Yuki Ito (Tokyo Institute of Technology), Ryo Matsumiya (Tokyo Institute of Technology), Toshio Endo (Tokyo Institute of Technology)

GPUs are widely used to accelerate deep learning with convolutional neural network (CNN). However, since GPU memory capacity is limited, it is difficult to implement efficient programs that compute large CNN on GPU. This poster describes the design and implementation of out-of-core cuDNN (ooc_cuDNN) library, which supports to compute CNN exceeding GPU memory capacity using capacity of CPU memory. ooc_cuDNN is an extension of cuDNN, which is high performance and popular deep learning

library. ooc_cuDNN divides CNN computation based on its performance model for better performance. In addition, ooc_cuDNN provides fused functions combined some computation to reduce extra communication. With ooc_cuDNN, we successfully computed CNN requiring more than 60 GB memory on a single GPU with 16 GB memory. Compared with an in-core case using cuDNN, performance degradation was 13 %.

P06: Large Scale FFT-Based Stress-Strain Simulations with Irregular Domain Decomposition

Authors: Anuva Kulkarni (Carnegie Mellon University), Franz Franchetti (Carnegie Mellon University), Jelena Kovacevic (Carnegie Mellon University)

Large-scale stress-strain simulations involving parallel Fast Fourier Transforms (FFTs) suffer from high memory requirements and high communication overhead. We propose an irregular domain decomposition method to reduce the memory requirement of an FFT-based stress-strain simulation algorithm for composite materials, the Moulinec-Suquet Composite (MSC) - Basic Scheme. This algorithm uses Green's functions to solve a partial differential equation. FFTs are used for convolution of large 3-D tensor fields with the Green's function.

In this preliminary work, we propose a modified algorithm, the MSC-Alternate Scheme, to show that processing the composite with smaller, local FFTs on irregular domains (grains in the material's microstructure) can reduce memory usage without adversely impacting accuracy of the result. Additionally, data models can reduce communication by compressing the data in the domains before the communication step. Our poster presents our proof-of-concept results and charts out the path towards a GPU implementation.

P07: PORTAGE - A Flexible Conservative Remapping Framework for Modern HPC Architectures

Authors: Rao V. Garimella (Los Alamos National Laboratory), Peter D. Crossman (Los Alamos National Laboratory), Gary A. Dilts (Los Alamos National Laboratory), Rajeev S. Erramilli (Los Alamos National Laboratory), Charles R. Ferenbaugh (Los Alamos National Laboratory), Angela M. Herring (Los Alamos National Laboratory), Eugene Kikinon (Los Alamos National Laboratory), Chris M. Malone (Los Alamos National Laboratory), Navamita Ray (Los Alamos National Laboratory), Mike L. Rogers (Los Alamos National Laboratory)

Portage is a massively parallel remapping framework to transfer fields between general polyhedral meshes while conserving integral quantities of interest. The framework also has the capability to remap data between two point clouds. Portage is templated on the component classes required in conservative remapping - search, intersection and interpolation as well as on the mesh and field managers. Applications supply Portage with custom components while the framework takes care of distributed parallelism using MPI and threaded parallelism using NVIDIA Thrust to scale to many thousands of cores. Moreover, the imposition of a functional design on the components used by Portage makes it very amenable to achieve task parallelism with runtime systems such as Legion. Portage has been tested in 2D/3D for remapping between general polygonal and polyhedral meshes and between point clouds. We present scaling results for distributed (MPI) and on-node parallelism (OpenMP) on LANL's HPC machines.

P08: Performance Optimization of Matrix-free Finite-Element Algorithms within deal.II

Authors: Martin Kronbichler (Technical University Munich), Karl Ljungkvist (Uppsala University), Momme Allalen (Leibniz Supercomputing Centre), Martin Ohlerich (Leibniz Supercomputing Centre), Igor Pasichnyk (IBM), Wolfgang A. Wall (Technical University Munich)

We present a performance comparison of highly tuned matrix-free finite element kernels from the deal.II finite element library on three contemporary computer architectures, an NVIDIA P100 GPU, an Intel Knights Landing Xeon Phi, and two multi-core Intel CPUs. The algorithms are based on fast integration on hexahedra using sum factorization techniques. On Cartesian meshes with a relatively high arithmetic intensity, the four architectures provide a surprisingly similar computational throughput. On curved meshes, the kernel is heavily memory bandwidth limited which reveals distinct differences between the architectures: the P100 is twice as fast as KNL, and almost four times as fast as the Haswell and Broadwell CPUs, effectively leveraging the higher memory bandwidth and the favorable shared memory programming model on the GPU.

P09: Adaptive Multistep Predictor for Accelerating Dynamic Implicit Finite-Element Simulations

Authors: Kohei Fujita (University of Tokyo, RIKEN), Tsuyoshi Ichimura (University of Tokyo, RIKEN), Masashi Horikoshi (Intel Corporation), Muneo Hori (University of Tokyo, RIKEN), Lalith Maddegadara (University of Tokyo, RIKEN)

We develop an adaptive multistep predictor for accelerating memory bandwidth-bound dynamic implicit finite-element simulations. We predict the solutions for future time steps adaptively using highly-efficient matrix-vector product kernels with multiple right-hand sides to reduce the number of iterations required in the solver. By applying the method to a conjugate gradient solver with 3 x 3 block Jacobi preconditioning, we were able to achieve a 42% speedup on a Skylake-SP Xeon Gold cluster for a typical earthquake ground motion problem. As the method enables the number of iterations, and thus the communication frequency, to be reduced, the developed solver was able to attain high size-up scalability: 80.6% up to 32,768 compute nodes on the K computer. The developed predictor can also be applied to other iterative solvers and is thus expected to be useful for wide range of dynamic implicit finite-element simulations.

P10: HiCOO: A Hierarchical Sparse Tensor Format for Tensor Decompositions

Authors: Jiajia Li (Georgia Institute of Technology), Jimeng Sun (Georgia Institute of Technology), Richard Vuduc (Georgia Institute of Technology)

This paper proposes a new Hierarchical COOrdinate (HiCOO) format for sparse tensors, which compresses its indices to units of sparse tensor blocks. HiCOO format does not favor one tensor mode over the others, thus can be used as a replacement of the traditional COOrdinate (COO) format. In this paper, we use HiCOO format for the Matriced Tensor Times Khatri-Rao Product (MTTKRP) operation, the most expensive computational core in the popular CANDECOMP/PARAFAC decomposition, then accelerate it on multicore CPU architecture using two parallel strategies for irregular shaped tensors. Parallel MTTKRP using HiCOO format achieves up to $3.5\times$ ($2.0\times$ on average) speedup over COO format and up to $4.3\times$ ($2.2\times$ on average) speedup over CSF format.

P11: Energy-Efficient Transprecision Techniques for Iterative Refinement

Authors: JunKyu Lee (Queen's University Belfast), Hans Vandierendonck (Queen's University Belfast), Dimitrios S. Nikolopoulos (Queen's University Belfast)

This paper presents transprecision techniques for iterative refinement, which utilize various precision arithmetic dynamically according to numeric properties of the algorithm and computational latencies depending on precisions. The transprecision techniques were plugged into a mixed precision iterative refinement on an Intel Xeon E5-2650 2GHz core with MKL 2017 and XBLAS 1.0. The transprecision techniques brought further 2.0-3.4X speedups and 3.0-4.1X energy reductions to a mixed precision iterative refinement when double precision solution accuracy was required for forward error and a matrix size was ranged from 4K to 32K.

P12: Multi-Size Optional Offline Caching Algorithms

Authors: Andrew Y. Choliy (Rutgers University), Max D. Whitmore (Brandeis University), Gruia Calinescu (Illinois Institute of Technology)

The optional offline caching (paging) problem, where all future file requests are known, is a variant of the heavily studied online caching problem. This offline problem has applications in web caching and distributed storage systems. Given a set of unique files with varying sizes, a series of requests for these files, fast cache memory of limited size, and slow main memory, an efficient replacement policy is necessary to decide when it is best to evict some file(s) from the cache in favor of another. It is known that this problem is NP-complete, and few approximation algorithms have been proposed. We propose three new heuristics, as well as a 4-approximation algorithm. We then evaluate each algorithm by the metrics of runtime complexity and proximity to the optimal solutions of many synthetic data sets.

P13: Large-Scale GW Calculations on Pre-Exascale HPC Systems

Authors: Mauro Del Ben (Lawrence Berkeley National Laboratory), Felipe H. da Jornada (University of California, Berkeley), Andrew Canning (Lawrence Berkeley National Laboratory), Nathan Wichmann (Cray Inc), Karthik Raman (Intel Corporation), Ruchira Sasanka (Intel Corporation), Chao Yang (Lawrence Berkeley National Laboratory), Steven G. Louie (Lawrence Berkeley National Laboratory; University of California, Berkeley)

The accurate determination of excitation spectra of materials, such as the electronic band gap, is critical for the design of novel devices, including photovoltaics, transistors, batteries, and LEDs. Many-body perturbation-theory methods, and the ab-initio GW approach in particular, have emerged over the last decades as the gold standard for computing these quantities. However, the ab-initio GW formalism is often limited to systems of at most 100 atoms due to its computational complexity. We present here large scale GW calculations of crystalline defect problems, relevant for the performance of semiconductors, with up to 1000 atoms, on the Cori system at NERSC. We show that the GW method is particularly well suited for exascale/pre-exascale systems. Our implementation, which uses a combination of new algorithms and optimizations targeted at many-core CPU architectures, scales well to the entire Cori system, and obtains a significant fraction of peak performance.

P14: Robust SA-AMG Solver by Extraction of Near-Kernel Vectors

Authors: Naoya Nomura (University of Tokyo), Kengo Nakajima (University of Tokyo), Akihiro Fujii (Kogakuin University)

The smoothed aggregation algebraic multigrid (SA-AMG) method is among the fastest solvers for large-scale linear equations, $Ax=b$. The SA-AMG method achieves good convergence and scalability by damping various wavelength components efficiently. To achieve this damping, this method creates multi-level matrices which are hierarchically smaller in dimension than the original matrix. Moreover, the convergence can be further improved by setting near-kernel vectors p , which satisfy $Ap \approx 0$ and $p \neq 0$. Generally, the same number of near-kernel vectors are used at each level. In the present work, we propose a method that extracts and adds near-kernel vectors at each level. We evaluate the performance of the solver that extracts the near-kernel vectors and adds them at each level. We use the three-dimensional elastic problem and employ up to 512 processes on the FX10 supercomputer system. By using this method, the performance is improved compared with previous work.

P15: Toward Decoupling the Selection of Compression Algorithms from Quality Constraints

Authors: Julian Kunkel (German Climate Computing Center), Anastasiia Novikova (University of Hamburg), Eugen Betke (German Climate Computing Center)

With the Scientific Compression Library (SCIL), we are developing a meta-compressor that allows users to set various quantities

that define the acceptable error and the expected performance behavior. The library then chooses the appropriate chain of algorithms to yield the users requirements. This approach is a crucial step towards a scientifically safe use of much-needed lossy data compression, because it disentangles the tasks of determining scientific ground characteristics of tolerable noise, from the task of determining an optimal compression strategy given target noise levels and constraints. Without changing applications, it allows these codes to utilize future algorithms once they are integrated into the library.

P16: Scaling Analysis of a Hierarchical Parallelization of Large Inverse Multiple-Scattering Solutions

Authors: Mert Hidayetoglu (University of Illinois), Carl Pearson (University of Illinois), Izzat El-Hajj (University of Illinois), Weng Cho Chew (University of Illinois), Levent Gurel (University of Illinois), Wen-Mei Hwu (University of Illinois)

We propose a hierarchical parallelization strategy to improve the scalability of inverse multiple-scattering solutions. The inverse solver parallelizes the independent forward solutions corresponding to different illuminations. For further scaling out on large numbers of computing nodes, each forward solver parallelizes the dense and large matrix-vector multiplications accelerated by the multilevel fast multipole algorithm. An inverse problem involving a large Shepp-Logan phantom is solved on up to 1,024 CPU nodes of the Blue Waters supercomputer in order to demonstrate the strong-scaling efficiency of the proposed parallelization scheme. The results show that parallelizing illuminations has almost perfect scaling efficiency of 95% because of the independent nature of forward-scattering solutions, however, parallelization of MLFMA has 73% efficiency due to MPI communications in MLFMA multiplications. Nevertheless, the proposed strategy improves granularity and allows spreading DBIM solutions on large numbers of nodes.

P17: Fully Non-Blocking Communication-Computation Overlap Using Assistant Cores toward Exascale Computing

Authors: Motoki Nakata (National Institute for Fusion Science), Masanori Nunami (National Institute for Fusion Science), Shinsuke Satake (National Institute for Fusion Science), Yoshihiro Kasai (Fujitsu Ltd), Shinya Maeyama (Nagoya University), Tomo-Hiko Watanabe (Nagoya University), Yasuhiro Idomura (Japan Atomic Energy Agency)

A fully non-blocking optimized Communication-Computation overlap technique using assistant cores (AC), which are independent from the calculation cores, is proposed for the application to the five-dimensional plasma turbulence simulation code with spectral (FFT) and finite-difference schemes, toward exascale supercomputing. The effects of optimization are examined in Fujitsu FX100 (2.62PFlop/s) with 32 ordinary cores and 2 Assistant cores/node, where AC enables us to employ the fully non-blocking MPI communications overlapped by the thread-parallelized calculations with OpenMP Static scheduling with much less overheads. It is clarified that the combination of the non-blocking communications by AC and the static scheduling leads to not only reduction in OpenMP overhead, but also improved load/store and cash performance, where about 22.5% improved numerical performance is confirmed in comparison to the conventional overlap by the master thread communications with dynamic scheduling.

P18: A Parallel Python Implementation of BLAST+ (PPIB) for Characterization of Complex Microbial Consortia

Authors: Amina Jackson (Naval Research Laboratory), William Connor Horne (Naval Research Laboratory), Daniel Beall (Naval Research Laboratory), Kenneth Jiang (Naval Research Laboratory), William Judson Hervey (Naval Research Laboratory)

Technological advancements in analytical instrumentation have enabled large-scale data acquisitions among the ‘-omics’ sciences of genomics, transcriptomics, and proteomics. An essential application among ‘-omics’ disciplines is the Basic Local Alignment Search Tool (BLAST) for functional inference of biomolecules. Though implementations of BLAST+ have been modified to address data volume growth, such improvements have neither been consistently maintained for high performance computing (HPC), nor have they been applied to complex microbiomes. Further, such implementations do not scale well to microbiomes of Naval interest on HPC systems in our hands.

Here, we compare 2 HPC implementations: BLAST+ and a Parallel Python Implementation of BLAST+ (PPIB) for protein functional inference. PPIB enabled functional inference of 2 complex microbiomes, which may be attributed to a combination of MPI and Python multiprocessing to query up to 3,600 proteins simultaneously. In contrast, BLAST+ did not complete functional assignments relative to PPIB at a comparable walltime.

P19: MPI-GIS: An MPI System for Big Spatial Data

Authors: Satish Puri (Marquette University)

In recent times, geospatial datasets are growing in terms of size, complexity and heterogeneity. High performance systems are needed to analyze such data to produce actionable insights in an efficient manner. For polygonal a.k.a vector datasets, operations such as I/O, data partitioning, and communication becomes challenging in a cluster environment.

In this work, we present MPI-GIS equipped with MPI-Vector-IO, a parallel I/O library that we have designed using MPI-IO specifically for irregular polygonal (vector) data formats such as Well Known Text, XML, etc. Our system can perform spatial in-memory indexing and join efficiently for an order of magnitude larger datasets compared to our previous work. It makes MPI aware of spatial data and spatial primitives and provides support for spatial data types embedded within collective computation and communication using MPI message-passing library. It takes less than 2 minutes to scan through 2.7 billion geometries in 96GB file using 160 processes.

P20: Facilitating the Scalability of ParSplice for Exascale Testbeds

Authors: Vinay B. Ramakrishnaiah (University of Wyoming), Jonas L. Landsgesell (University of Stuttgart), Ying Zhou (Loughborough University), Iris Linck (University of Colorado, Denver), Mouad Ramil (National School of Bridges and Roads - ParisTech), Joshua Bevan (University of Illinois), Danny Perez (Los Alamos National Laboratory), Louis J. Vernon (Los Alamos National Laboratory), Thomas D. Swinburne (Los Alamos National Laboratory), Robert S. Pavel (Los Alamos National Laboratory), Christoph Junghans (Los Alamos National Laboratory)

Parallel trajectory splicing (ParSplice) is an attempt to solve the enduring challenge of simulating the evolution of materials over long time scales for complex atomistic systems. A novel version of ParSplice is introduced with features that could be useful in its scaling to exascale architectures. A two-pronged approach is used. First, latent parallelism is exploited by extending support to heterogeneous architectures, including GPUs and KNLs. Second, the efficiency of the Kinetic Monte Carlo predictor is improved, allowing enhanced parallel speculative execution. The key idea in these predictor modifications is to include statistics from higher temperature simulations. The issue of inherent uncertainty in the prediction model was addressed in order to improve the performance, as the current predictor only takes into account the previous observations to formulate the problem. The predictor was also improved by using a hybrid approach of message-passing + multi-threading. (LA-UR-17-26181)

P21: The First Real-Scale DEM Simulation of a Sandbox Experiment Using 2.4 Billion Particles

Authors: Mikito Furuichi (Japan Agency for Marine-Earth Science and Technology), Daisuke Nishiura (Japan Agency for Marine-Earth Science and Technology), Mitsuteru Asai (Kyushu University), Takane Hori (Japan Agency for Marine-Earth Science and Technology)

A novel implementation of the Discrete Element Method (DEM) for a large parallel computer system is presented to simulate a sandbox experiment with realistic particle sizes. To save memory in the pairwise tangential forces and halve the arithmetic costs, interactions are calculated using the action-reaction law. An iterative load-balancer the flexible 2D orthogonal domain decomposition is applied to overcome the load-imbalance problem caused by the Lagrangian nature of DEM. An overlapping communication technique combined with cell-ordering with space-filling curves is also applied to hide the overhead cost because of the MPI communication tasks. We verify our complex parallel implementation with the action-reaction law via a reproducibility test. The parallel scaling test shows good strong, and weak scalabilities up to 2.4 billion particles on the Earth Simulator and the K computer. The world's first real-scaled numerical sandbox simulation successfully captures the characteristics of real observations.

P22: Numerical Simulation of Snow Accretion by Airflow Simulator and Particle Simulator

Authors: Kohei Murotani (Railway Technical Research Institute), Koji Nakade (Railway Technical Research Institute), Yasushi Kamata (Railway Technical Research Institute)

In this research, to take countermeasures for the snow accretion damage, we developed a simulator of realizing the snow accretion process in the following steps. Firstly, air flow analysis is performed by "Airflow simulator" developed by RTRI (Railway Technical Research Institute). Secondly, trajectory of flying snow is calculated by Basset-Boussinesq-Oseen equation using distributed of velocity of air flow. Thirdly, snow accretion analysis is performed by "Particle simulator" developed by RTRI. The shape modified by snow accretion is reflected onto the boundary conditions of the air flow analysis. In this year, snow accretion analysis for simple cubic shapes is performed in order to aim at system development and validation.

P23: AI with Super-Computed Data for Monte Carlo Earthquake Hazard Classification

Authors: Tsuyoshi Ichimura (University of Tokyo, RIKEN), Kohei Fujita (University of Tokyo, RIKEN), Takuma Yamaguchi (University of Tokyo), Muneo Hori (University of Tokyo, RIKEN), Madgededara Lalith (University of Tokyo, RIKEN), Naonori Ueda (RIKEN)

Many problems associated with earthquakes are yet to be solved using heroic computing, which is defined as computing at the largest scale possible using the best supercomputers and algorithms. Thus, a continuous effort has been pursued in HPC to solve these problems. However, even when heroic computing is applied, its practical use is difficult without considering the uncertainties in models. In this study, we constructed an AI methodology that uses super-computed data generated using heroic computing. We applied this AI to an earthquake hazard classification including uncertainty analyses in order to demonstrate its utility. This study can be regarded as an innovative step towards realizing high quality computing for Earthquakes by exploiting the potential of HPC through AI.

BP

P24: A Deployment of HPC Algorithm into Pre/Post-Processing for Industrial CFD on K-Computer

Authors: Keiji Onishi (RIKEN), Niclas Jansson (KTH Royal Institute of Technology), Rahul Bale (RIKEN), Wei-Hsiang Wang (RIKEN), Chung-Gang Li (Kobe University, RIKEN), Makoto Tsubokura (Kobe University, RIKEN)

Pre- and post-processing is still a major problem in industrial computational fluid dynamics (CFD). With the rapid development of computers, physical solvers are getting faster, while pre- remains slow because it's mainly a serial process. A methodology using MPI+OpenMP hybrid parallelization has been proposed to eliminate the manual work required during pre-processing for correcting the surface imperfections of CAD data. Compared to the rapidly increasing amount of data in recent years, the speed-up of visualization is insufficient. We address this limitation of post- by adapting the in-situ visualization to parallelize the post-processing using libsim (Visit) library. The performance of pre-/post- processing is investigated in this work, and we show that the pre-processing time has been reduced from several days in the conventional framework to order of minutes. The post-processing time has been reduced seconds order per frame, and approximately 30% increase of computational time was observed in vehicle aerodynamics cases.

P25: Large-Scale Adaptive Mesh Simulations Through Non-Volatile Byte-Addressable Memory

Authors: Bao Nguyen (Washington State University, Vancouver), Hua Tan (Washington State University, Vancouver), Xuechen Zhang (Washington State University, Vancouver)

Octree-based mesh adaptation has enabled simulations of complex physical phenomena. Existing meshing algorithms were proposed with the assumption that computer memory is volatile. Consequently, for failure recovery, the in-core algorithms need to save memory states as snapshots with slow file I/Os. The out-of-core algorithms store octants on disks for persistence. However, neither of them was designed to leverage unique characteristics of non-volatile byte-addressable memory (NVBM). We propose a novel data structure Persistent Merged octree (PM-octree) for both meshing and in-memory storage of persistent octrees using NVBM. It is a multi-version data structure and can recover from failures using its earlier persistent version stored in NVBM. In addition, we design a feature-directed sampling approach to help dynamically transform the PM-octree layout for reducing NVBM-induced memory write latency.

P26: Optimizing Gravity and Nuclear Physics in FLASH for Exascale

Authors: Hannah Klion (University of California, Berkeley; Oak Ridge National Laboratory), Bronson Messer (Oak Ridge National Laboratory, University of Tennessee), J. Austin Harris (Oak Ridge National Laboratory), Thomas Papatheodore (Oak Ridge National Laboratory)

In a Type Ia supernova, runaway fusion ignites in a white dwarf, causing it to explode. The heavy element yields of these events remain uncertain, and high-performance multiphysics simulations with tools like FLASH are critical for our understanding. Current simulations track approximately a dozen nuclear isotopes, as opposed to the thousands required to completely capture the event's nuclear physics.

Simulating nuclear physics and self-gravity accurately and efficiently is critical for modeling a Type Ia supernova, since supernovae are competitions between energy-releasing nuclear reactions and gravity. Currently, the FLASH nuclear reaction network and self-gravity solver requires substantial inter-node communication. We use non-blocking MPI collectives to overlap communication in the self-gravity calculation with the computation-heavy nuclear burning calculation. We find that speedups from this technique are possible, but are MPI implementation-dependent. We highlight some of the challenges associated with this type of optimization.

P27: Parallelization of the Particle-In-Cell Monte Carlo Collision (PIC-MCC) Algorithm for Plasma Simulation on Intel MIC Xeon Phi Architecture

Authors: Keval Shah (Dhirubhai Ambani Institute of Information and Communication Technology), Anusha Phadnis (Dhirubhai Ambani Institute of Information and Communication Technology), Miral Shah (Dhirubhai Ambani Institute of Information and Communication Technology), Bhaskar Chaudhury (Dhirubhai Ambani Institute of Information and Communication Technology)

The implementation of 2D-3v (2D in space and 3D in velocity space) PIC-MCC (Particle-In-Cell Monte Carlo Collision) method involves the computational solution of Vlasov-Poisson equations. This provides the spatial and temporal evolution of the charged-particle velocity distribution functions in plasma under the effect of self-consistent electromagnetic fields and collisions. Stringent numerical constraints associated with the PIC code makes it computationally prohibitive on CPU.

In our work, parallelization and optimization techniques have been extended to this simulation, along with a novel approach that involves developing a 'self-aware' code that triggers sorting in order to maintain cache-coherence while reducing the total sorting time during iterations.

We present the effect of important numerical parameters on speed-up. Finally, we compare the scalability and performance of the parallelization and optimization strategies on Intel® Xeon™ E5-2630, Xeon Phi™ 5110p and Xeon Phi™ 7250 relative to a serial implementation on Intel® i5.

P28: High-Fidelity Blade-Resolved Wind Plant Modeling

Authors: Andrew C. Kirby (University of Wyoming), Zhi Yang (University of Wyoming), Michael J. Brazell (University of Wyoming), Behzad R. Ahrabi (University of Wyoming), Jay Sitaraman (Parallel Geometric Algorithms LLC), Dimitri J. Mavriplis (University of Wyoming)

Blade-resolved numerical simulations of wind energy applications using full blade and tower models are presented. The computational methodology combines solution technologies in a multi-mesh, multi-solver paradigm through a dynamic overset framework. The coupling of a finite-volume solver and a high-order, hp-adaptive finite-element solver is utilized. Additional technologies including in-situ visualization and atmospheric micro-scale modeling are incorporated into the analysis environment. Validation of the computational framework is performed on the NREL 5MW wind turbine, the unsteady aerodynamics experimental NREL Phase VI turbine, and the Siemens SWT-2.3-93 wind turbine. The power and thrust results of all single turbine simulations agree well with low-fidelity model simulation results and field experiments when available. Scalability of the computational framework is demonstrated using 6, 12, 24, 48, and 96 wind turbine wind plant set-ups including the 48 turbine wind plant known as Lillgrund. Demonstration of the coupling of atmospheric micro-scale and CFD solvers is presented.

P29: A Deep Learning Tool for Fast Simulation

Authors: Sofia Vallecorsa (CERN), Gul rukh Khattak (CERN), Shruti Sharan (CERN)

We present the first application of Volumetric Generative Adversarial Network (VGAN) to High Energy Physics simulation. We generate three dimensional images of particles depositing energy in calorimeters. This is the first time such an approach is taken in HEP where most of data is three dimensional in nature but it is customary to convert it into two dimensional slices. The volumetric approach leads to a larger number of parameters, but two dimensional slicing loses the volumetric dependencies inherent in the dataset. The present work proves the success of handling those dependencies through VGANs. Energy showers are faithfully reproduced in all dimensions and show a reasonable agreement with standard techniques. We also demonstrate the ability to condition training on several parameters such as particle type and energy. This work aims at proving Deep Learning techniques represent a valid fast alternative to standard MonteCarlo approaches and is part of the GEANTV project.

P30: MPI/OpenMP Parallelization of the Hartree-Fock Method for the Second Generation Intel Xeon Phi

Authors: Kristopher Keipert (Iowa State University), Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Michael D'mello (Intel Corporation), Alexander Moskovsky (RSC Technologies), Mark S. Gordon (Iowa State University)

Replication of critical data structures in the MPI-only GAMESS Hartree-Fock algorithm limits the full utilization of the manycore Intel Xeon Phi processor. In this work, modern OpenMP threading techniques are used to implement hybrid MPI/OpenMP algorithms. Two separate implementations that differ by the sharing and replication details of key data structures among threads are considered. The hybrid MPI/OpenMP implementations reduce the memory footprint by approximately 200 times compared to the legacy code. The MPI/OpenMP code was shown to run up to six times faster than the original for a range of molecular system sizes. The implementation details and strategies will be presented for both hybrid algorithms. Benchmark scaling results utilizing up to 3000 Intel Xeon Phi processors will also be discussed.

BP

P31: Understanding the Performance of Small Convolution Operations for CNN on Intel Architecture

Authors: Alexander Heinecke (Intel Corporation), Evangelos Georganas (Intel Corporation), Kunal Banerjee (Intel Corporation), Dhiraj Kalmakar (Intel Corporation), Narayanan Sundaram (Intel Corporation), Anand Venkat (Intel Corporation), Greg Henry (Intel Corporation), Hans Pabst (Intel Corporation)

Convolution layers are prevalent in many classes of deep neural networks, including Convolutional Neural Networks (CNNs) which provide state-of-the-art results for tasks like image recognition, natural language processing, and speech recognition. The computationally expensive nature of a convolution operation has led to the proliferation of implementations including matrix-matrix multiplication formulation, FFT-formulation, Winograd transformation, and direct convolution primarily targeting GPUs. In this paper, we optimize a direct convolution and Winograd implementation for x86 architectures, in particular for Xeon Phi systems, via a dynamic compilation approach. We then show how these JIT optimizations can be integrated in a high-level domain-specific language setting. We shed light on what is possible and what is not possible based on different data-formats and blocking techniques. Our JIT-based Ninja implementation shows close to theoretical peak results on modern x86 architectures, depending on setting and the CPU architecture at hand.

P32: Exploring the Performance of Electron Correlation Method Implementations on Kove XPDs

Authors: Colleen Bertoni (Argonne National Laboratory), Brian Toonen (Argonne National Laboratory), William Allcock (Argonne National Laboratory), Spencer R. Pruitt (Worcester Polytechnic Institute), Mark S. Gordon (Iowa State University)

In electron correlation methods in quantum chemistry, there are often high memory requirements which can reach terabytes for medium-sized molecular systems. For second-order perturbation theory (MP2), the two-electron integral arrays are the main memory bottleneck. Previously the two-electron integrals were recomputed, stored in distributed memory, or stored on disk. A way of storing the arrays which would remove the dependence on compute node memory and large latency associated with using disk is by using an external memory appliance, like Kove's XPD.

In this work, we modified a distributed memory implementation of MP2 to use XPDs instead of distributed memory. We evaluated the performance of our implementation against the distributed memory version for several molecular systems by considering scaling behavior with respect to compute processes and connections to XPDs. In the poster, we present an outline of the MP2 implementation using XPDs and the scaling results.

P33: Massively Parallel Evolutionary Computation for Empowering Electoral Reform: Quantifying Gerrymandering via Multi-objective Optimization and Statistical Analysis

Authors: Wendy K. Cho (National Center for Supercomputing Applications, University of Illinois), Yan Liu (National Center for Supercomputing Applications, University of Illinois)

Important insights into redistricting can be gained by formulating and analyzing the problem within a large-scale optimization framework. Redistricting is an application of the set-partitioning problem that is NP-hard. We design a hybrid metaheuristic as the base search algorithm. With our grant on the Blue Waters supercomputer, we extend our algorithm to the high-performance-computing realm by using MPI to implement an asynchronous processor communication framework. We experimentally demonstrate the effectiveness of our algorithm to utilize multiple processors and to scale to 131,072 processors. The massive computing power allows us to extract new substantive insights that closely mesh with the framework that the Supreme Court has elucidated for electoral reform.

P34: GPU Acceleration for the Impurity Solver in GW+DMFT Packages

Authors: Kwangmin Yu (Brookhaven National Laboratory), Patrick Semon (Brookhaven National Laboratory), Nicholas D'Imperio (Brookhaven National Laboratory)

The combination of dynamical mean field theory (DMFT) and GW (or density functional theory) has become a powerful tool to study and predict properties of real materials with strongly correlated electrons, such as high temperature superconductors. At the core of this combined theory lies the solution of a quantum impurity model, and continuous-time quantum Monte Carlo (CT-QMC) has proven an indispensable algorithm in this respect. However, depending on the material, this algorithm is computationally very expensive, and enhancements are crucial for bringing new materials within reach of GW+DMFT. Based on a CPU implementation, GPU acceleration is added, and two times speedup is achieved. New techniques are invented and implemented to deal with various GPU acceleration environment.

P36: A Novel Feature-Preserving Spatial Mapping for Deep Learning Classification of Ras Structures

Authors: Thomas Corcoran (Lawrence Berkeley National Laboratory), Rafael Zamora-Resendiz (Lawrence Berkeley National Laboratory), Xinlian Liu (Lawrence Berkeley National Laboratory), Silvia Crivelli (Lawrence Berkeley National Laboratory)

A protein's 3D structure determines its functionality, and is therefore a topic of great importance. This work leverages the power of Convolutional Neural Networks (CNNs) to classify proteins and extract features directly from their 3D structures. So far, researchers have been unable to fully exploit 3D structural information with 2D CNNs, partly because it is difficult to encode 3D data into the 2D format that can be ingested by such networks. We designed and implemented a novel method that maps 3D models to 2D data grids as a preprocessing step for 2D CNN use. Our experiments focused on the Ras protein family, which has been linked to various forms of cancer. Our trained CNNs are able to distinguish between two branches within the Ras family, HRas and KRas, which are similar in sequence and structure. Analysis of saliency maps suggests classification is accomplished by detection of structurally and biologically-meaningful sites.

P37: PaSTRI: A Novel Data Compression Algorithm for Two-Electron Integrals in Quantum Chemistry

Authors: Ali Murat Gok (Argonne National Laboratory, Northwestern University), Dingwen Tao (University of California, Riverside), Sheng Di (Argonne National Laboratory), Vladimir Mironov (Lomonosov Moscow State University), Yuri Alexeev (Argonne National Laboratory), Franck Cappello (Argonne National Laboratory)

Integral computations for two-electron repulsion energies are very frequently used applications in quantum chemistry. Computational complexity, energy consumption and the size of the output data generated by these computations scales with $O(N^4)$, where N is the number of atoms simulated. Typically, the same integrals are calculated multiple times. Storing these values and reusing them requires impractical amounts of storage space; whereas recalculating them requires a lot of computations. We propose PaSTRI (Pattern Scaling for Two-electron Repulsion Integrals), a fast novel compression algorithm which makes it possible to calculate these integrals only once, store them, and reuse them at much smaller computational cost than recalculation. PaSTRI is "lossy" compared to floating point numbers, but still maintains the precision level required by the integral computations. PaSTRI is an extension to SZ compressor package as a part of ECP-EZ. PaSTRI achieves 17.5:1 compression ratio whereas vanilla SZ achieves 8.0:1 and ZFP achieves 7.1:1.

P38: Benchmarking Parallelized File Aggregation Tools for Large Scale Data Management

Authors: Tiffany Li (National Center for Supercomputing Applications, University of Illinois), Craig Steffen (National Center for Supercomputing Applications, University of Illinois), Ryan Chui (National Center for Supercomputing Applications, University of Illinois), Roland Haas (National Center for Supercomputing Applications, University of Illinois), Liudmila S. Mainzer (National Center for Supercomputing Applications, University of Illinois)

Large-scale genomic data analyses have given rise to bottlenecks in data management due to the production of many small files. Existing file-archiving utilities, such as tar, are unable to efficiently package large datasets with upward of multiple terabytes and hundreds of thousands of files. To create parallelized and multi-threaded alternatives, ParFu (parallel archiving file utility), MPItar, and ptgz (parallel tar gzip) were developed by the Blue Waters team and the NCSA Genomics team as efficient data management tools, with the ability to perform parallel archiving (and eventually extracting). Scalability was tested for each tool as a function of the number of ranks executed and stripe count on a Lustre filesystem. We used two datasets typically seen in genomic analyses to measure the effects of different file-size distributions. These tests suggest the best user parameters and subsequent costs for usage as efficient replacements of data-packaging tools.

P39: Extremely Large, Wide-Area Power-Line Models

Authors: Ross N. Adelman (US Army Research Laboratory)

The electric and magnetic fields around power lines carry an immense amount of information about the power grid, and can be used to improve stability, balance loads, and reduce outages. To study this, extremely large models of transmission lines over a 49.5-sq-km tract of land near Washington, DC have been built. The terrain is modeled accurately using 1-m-resolution LIDAR data. The models are solved using the boundary element method, and the solvers are parallelized across Army Research Laboratory's Centennial supercomputer using a modified version of the domain decomposition method. The code on each node is accelerated using the fast multipole method and, when available, GPUs. Additionally, larger test models were used to characterize the scaling

properties of the code. The largest test model had 10,020,913,152 elements, and was solved across 1024 nodes in 3.0 hours.

P40: Running Large-Scale Ultrasound Simulations on Piz Daint with 512 Pascal GPUs

Authors: Filip Vaverka (Brno University of Technology), Bradley E. Treeby (University College London), Jiri Jaros (Brno University of Technology)

Ultrasound simulation is a critical component of model-based treatment planning for ultrasound therapy. However, the domains are typically thousands of wavelengths in size, leading to large-scale numerical models with 10s of billions of unknowns. This paper presents a novel local Fourier basis domain decomposition for full-wave ultrasound propagation simulations with a custom bell function which ensures that the numerical error stays below 0.1% while enabling almost ideal strong scaling. Realistic benchmarks with 512 Nvidia P100 GPUs in the best EU supercomputer Piz Daint achieved efficiency between 90 and 100% with a speed-up over 100 and computational cost reduction by a factor of 12 compared to 1024 Haswell cores.

P41: OpenCL-Based High-Performance 3D Stencil Computation on FPGAs

Authors: Hamid Reza Zohouri (Tokyo Institute of Technology), Artur Podobas (Tokyo Institute of Technology), Naoya Maruyama (RIKEN), Satoshi Matsuoka (Tokyo Institute of Technology)

With the recent advancements in OpenCL-based High-Level Synthesis, FPGAs are now more attractive choices for accelerating High Performance Computing workloads. Despite their power efficiency advantage, FPGAs usually fall short in terms of sheer performance against GPUs due to having multiple times lower memory bandwidth and compute performance. In this work, we show that due to the architectural advantage of FPGAs for stencil computation, apart from power efficiency, these devices can also offer comparable performance to high-end GPUs. We achieve this goal using a parameterized OpenCL-based implementation that employs both spatial and temporal blocking, and multiple advanced FPGA-specific optimizations to maximize performance. We show that it is possible to achieve up to 60 GBps and 230 GBps of effective throughput for 3D stencil computation on Intel Stratix V and Arria 10 FPGAs, respectively, which is comparable to a highly-optimized implementation on high-end GPUs.

P42: TRIP: An Ultra-Low Latency, TeraOps/s Reconfigurable Inference Processor for Multi-Layer Perceptrons

Authors: Ahmed Sanaullah (Boston University), Chen Yang (Boston University), Yuri Alexeev (Argonne National Laboratory), Kazutomo Yoshii (Argonne National Laboratory), Martin C. Herboldt (Boston University)

Multi-Layer Perceptron (MLP) is one of the most commonly deployed Deep Neural Networks, representing 61% of the workload in Google data-centers. MLP Inference, a memory bound problem, typically has hard response time deadlines and prefers latency over throughput. In our work, we designed a TeraOps/s Reconfigurable Inference Processor for MLPs (TRIP) on FPGAs that alleviates the memory bottleneck by storing all application specific weights on-chip. It can be deployed in multiple configurations, including host-independent operation. We have shown that TRIP achieves 60x better performance than the current state-of-the-art Google Tensor Processing Unit (TPU) for MLP Inference. It was demonstrated on the cancer patient datasets used in the Candle Exascale Computing Project (ECP).

P43: Deep Packet/Flow Analysis Using GPUs

Authors: Qian Gong (Fermi National Laboratory), Wenji Wu (Fermi National Laboratory), Phil DeMar (Fermi National Laboratory)

Deep packet inspection (DPI) faces severe challenges in high-speed networks as it requires high I/O throughputs and intensive computations. The parallel architecture of GPUs fits exceptionally well for per-packet traffic processing. However, TCP data stream need to be reconstructed in a per-flow level to deliver a consistent content analysis. Since the flow-centric operations are naturally anti-parallel and often require large memory space for buffering out-of-sequence packets, they can be problematic for GPUs. Here, we present a highly efficient DPI framework, which includes a purely GPU-implemented TCP flow tracking and stream reassembly. Instead of buffering till the TCP packets become in sequence, we process the packets in batch with pattern matching states between consecutive batches connected by a Aho-Corasick with a prefix-/suffix- tree method. Evaluation shows that our code can reassemble tens of millions of packets per second and conduct a signature-based DPI at 55 Gbit/s using an NVIDIA K40 GPU.

P44: Increasing Throughput of Multiprogram HPC Workloads: Evaluating a SMT Co-Scheduling Approach

Authors: Elias Lundmark (University West Sweden), Chris Persson (University West Sweden), Andreas de Blanche (University West Sweden, Tetra Pak), Thomas Lundqvist (University West Sweden)

Simultaneous Multithreading (SMT) is a technique that allows for more efficient processor utilization by scheduling multiple threads on a single physical core, thus increasing the instruction level parallelism as it mixes instructions from several threads. Previous research has shown an average throughput increase of around 20% with an SMT level of two, e.g. two threads per core. However, a bad combination of threads can actually result in decreased performance. To be conservative, many HPC-systems have SMT disabled, thus, limiting the number of scheduling slots in the system to one per core. However, for SMT to not hurt performance, we need to determine which threads should share a core. In this poster, we use 30 random SPEC CPU job mixed on a twelve-core Broadwell based node, to study the impact of enabling SMT using two different co-scheduling strategies. The results show that SMT can increase performance especially when using no-same-program co-scheduling.

P45: Campaign Storage: Erasure Coding with GPUs

Authors: Walker Haddock (University of Alabama, Birmingham), Matthew L. Curry (Sandia National Laboratories), Purushotham Bangalore (University of Alabama, Birmingham), Anthony Skjellum (University of Tennessee, Chattanooga)

Cloud computing has developed high capacity, reliable and economical storage systems based on object technology. Los Alamos National Labs has designed the storage systems for Trinity to include a cloud type object storage system as a layer between the Parallel File System (PFS). This pre-archive system has been dubbed "Campaign Storage" with the purpose of storing data products to be quickly accessible during the life of a research project. Data stored on the Campaign Storage can be pulled into the PFS or moved to archive after the data has been curated. Campaign Storage reduces the capacity requirements for PFS storage and reduces the data transfer bandwidth requirements for the archive storage.

We make these contributions to the pre-archive storage layer: * GPU assisted erasure coding * Demonstrating erasure on File Transfer Agents * Reducing erasure recovery costs with "lazy recovery" * Enabling larger erasure coded disk pools

P46: Understanding How OpenCL Parameters Impact on Off-Chip Memory Performance of FPGA Platforms

Authors: Yingyi Luo (Northwestern University), Zheming Jin (Argonne National Laboratory), Kazutomo Yoshii (Argonne National Laboratory), Seda Ogrenci-Memik (Northwestern University)

Reconfigurability has strong potential to achieve higher performance and energy efficiency in the post-Moore era. Field-programmable gate arrays (FPGAs), the most practical reconfigurable architecture today, are becoming more relevant to scientific computing thanks to hardened floating-point circuits and emerging FPGA high-level synthesis (HLS) technology. Most notably, FPGA vendors started supporting OpenCL for FPGA platforms, and some OpenCL-based codes have been ported to FPGAs. However, OpenCL offers no guarantee for performance portability; optimal OpenCL parameters such as global size and local size are different between platforms, which could lead to unfair comparisons. In this study, our objective is twofold: 1) to understand how OpenCL parameters impact off-chip memory access performance of the current generation of OpenCL-FPGA platforms and 2) to find effective OpenCL parameters empirically from microbenchmark results.

P47: Understanding Congestion on Omni-Path Fabrics

Authors: Lauren Gillespie (Southwestern University), Christopher Leap (University of New Mexico), Dan Cassidy (Los Alamos National Laboratory)

High-performance computing systems require high-speed interconnects, such as InfiniBand (IB), to efficiently transmit data. Intel's Omni-Path Architecture (OPA) is a new interconnect similar to IB that is implemented on some of Los Alamos National Laboratory's recent clusters. Both interconnects suffer from degraded performance under heavy network traffic loads, resulting in packet discards. However, unlike IB, OPA specifically calls out these drops in the form of the performance counter, congestion discards. Owing to the relative immaturity of the OPA fabric technology, the correlation between performance degradation and congestion discards has not been fully evaluated to date. This research aims to increase the level of understanding of the effects congestion has on cluster performance by presenting a sufficiently high data injection load to the OPA fabric such that performance degradation is induced and the cause of this performance degradation can be evaluated. LA-UR-17-26341

P48: Prototyping of Offloaded Persistent Broadcast on Tofu2 Interconnect

Authors: Yoshiyuki Morie (RIKEN), Masayuki Hatanaka (RIKEN), Masamichi Takagi (RIKEN), Atsushi Hori (RIKEN), Yutaka Ishikawa (RIKEN)

With the increasing scale of parallel computers, it has become more important to reduce communication time. Overlapping computation and communication is one effective method for hiding communication delay. Although standard non-blocking collective communication is an overlap method, it requires generating a communication command sequence for each collective communication. In contrast, persistent non-blocking collective communication can generate the sequence at initialization and reuse it at the start of collective communication. Moreover, if the sequence can be offloaded to a network device, more efficient execution is possible without using CPU cycles.

In this poster, a persistent non-blocking broadcast is implemented using the offloading functionality of the Tofu2 interconnect on the Fujitsu FX100 supercomputer, the successor to the K computer. We report the performance improvement by offloading persistent non-blocking collective communication in a real machine.

P49: Toward Exascale HPC Systems: Exploiting Advances in High Bandwidth Memory (HBM2) through Scalable All-to-All Optical Interconnect Architectures

Authors: Pouya Fotouhi (University of California, Davis), Roberto Proietti (University of California, Davis), Paolo Grani (University of California, Davis), Mohammad Amin Nazirzadeh (University of California, Davis), S. J. Ben Yoo (University of California, Davis)

As we reach the limits of miniaturization in fabrication processes, the interpretation of Moore's law has changed from doubling the frequency every eighteen months to doubling the core count every three to four years (from 2 cores in 2004 to 16 cores in 2015). To reach exascale-level computation, the communication and data transfers between processors and memory is expected to increase drastically; the on-chip interconnect plays a key role in the overall system latency and energy-efficiency. Therefore, novel solutions providing one order of magnitude higher bandwidth and lower energy consumption than what is possible with current electrical interconnects are needed. This poster discusses an optical interconnected compute node that makes use of embedded photonic interconnects together with emerging high bandwidth memory technologies (such as HBM and HBM2). Two different multi-

processors architectures with different requirements in terms of number of lasers, high-speed SERDES, and memory bandwidth per processors are presented.

P50: Energy-Efficient and Scalable Bio-Inspired Nanophotonic Computing

Authors: Mohammadamin Nazirzadeh (University of California, Davis), Pouya Fotouhi (University of California, Davis), Mohammadsadeqh Shamsabarbeh (University of California, Davis), Roberto Proietti (University of California, Davis), S. J. Ben Yoo (University of California, Davis)

This paper discusses bio-inspired neuromorphic computing utilizing nanophotonic, nanoelectronic, and NEMS technologies integrated into reconfigurable 2D-3D integrated circuits as hierarchical neural networks. The goal is to achieve $\geq 1000\times$ improvements in energy-per-operation compare to the state-of-the-art implementations of neural networks on Von-Neumann based computers. We combine nanophotonic and nanoelectronic technologies to build energy-efficient (~ 10 fJ/b) artificial spiking neurons with required functionality (spiking, integration, thresholding, reset). Photonic interconnects exploiting 2x2 NEMS-MZIs enables distance independent propagation of signal with weighted addition among the neurons as well as possibility of on-line learning capability. Using low-leakage nanophotonic and nanoelectronic devices, and NEMS, the static power consumption of the system can be decreased down to nearly zero. Realizing 2D-3D photonic integrated circuit technologies, the proposed system can overcome the scalability limitations of current neuromorphic computing architectures.

BP

P51: TuPiX-Flow: Workflow-Based Large-Scale Scientific Data Analysis System

Authors: Sunggeun Han (Korea Institute of Science and Technology Information), Jung-Ho Um (Korea Institute of Science and Technology Information), Hyunwoo Kim (Korea Institute of Science and Technology Information), Kyongseok Park (Korea Institute of Science and Technology Information)

With more research being actively conducted on big data, there has been a growing interest towards data-intensive data science in various fields. Against this backdrop, extensive efforts have been made to apply High Performance Computing (HPC) technology such as distributed computing or parallel computing. However, many researchers are unable to fully utilize such technology due to their lack of knowledge, programming skills, and analytical skills for large-scale data. TuPiX-Flow, which provides a workflow-based user interface, enables researchers to easily analyze large-scale data using workflow diagrams for end-to-end analytical processes even without programming knowledge or implementation technology. In addition, large-scale data can be efficiently analyzed in software compatible with HPC, including distributed computing and parallel computing. As a case study of large-scale data analysis using TuPiX-Flow, a model for the detection of red tides surrounding the Korean peninsula was analyzed based on ocean color satellite data.

P52: A Simulation-Based Analysis on the Configuration of Burst Buffer

Authors: Tianqi Xu (Tokyo Institute of Technology), Kento Sato (Lawrence Livermore National Laboratory), Satoshi Matsuoka (Tokyo Institute of Technology)

Burst buffers have been widely deployed in many supercomputer systems to absorb bursty I/O and accelerate I/O performance. Previous work has shown that with burst buffer systems, I/O operations from computer nodes can be greatly accelerated. Different configurations of burst buffers can have huge impact on performance of applications. However, the proper configuration of burst buffers for given systems and workloads still remains an open problem. In this paper, we address this challenge by simulating the behavior of a burst buffer under different buffer sizes with trace logs from a set of HPC applications in a distributed environment. We tuned our simulator with a production level burst buffer system. From the results of the simulation, we found that for most of HPC applications, using a buffer size that is less than half of the total access space of the application can still achieve high performance.

BP

P53: TensorViz: Visualizing the Training of Convolutional Neural Network Using ParaView

Authors: Xinyu Chen (University of New Mexico), Qiang Guan (Los Alamos National Laboratory), Xin Liang (University of California, Riverside), Li-Ta Lo (Los Alamos National Laboratory), Simon Su (US Army Research Laboratory), Trilce Estrada (University of New Mexico), James Ahrens (Los Alamos National Laboratory)

Deep Convolutional Networks have been very successful in visual recognition tasks recently. Previous works visualize learned features at different layers to help people understand how CNNs learn visual recognition tasks. However, they do not help to accelerate the training process. We use ParaView to provides both qualitative and quantitative visualization that help understand the learning procedure, tune the learning parameters, and direct merging and pruning of neural networks.

P54: Investigating Hardware Offloading for Reed-Solomon Encoding

Authors: John W. Dermer (Los Alamos National Laboratory), Gustavo De Leon (Los Alamos National Laboratory; University of California, Berkeley), Tyler S. Rau (Los Alamos National Laboratory)

Reed-Solomon (RS) encoding is a storage scheme which offers better scalability, but requires heavier computation, compared to other models. This presents a problem as it requires users to purchase brawnier CPUs than would be otherwise necessitated. However, Mellanox's ConnectX-4 Infiniband cards have the capability to perform RS encoding on the HCA hardware; removing the need for powerful CPUs to calculate it. We investigated the performance, measured in throughput, between these cards and Intel's ISA-Library, with regard to various block sizes and concurrency. We conclude that the MLX cards encoded faster and more consistently than ISA-L. Furthermore, the ConnectX-5 cards support the Galois Field (GF) 2^8 , this grants compatibility with data

encoded by any system using GF(2⁸) or less, including ISA-L. These cards enable users to substantially increase encoding and decoding throughput by using more cards; additionally enabling the use of less powerful CPUs to achieve similar performance. LA-UR-17-26333

P55: Incorporating Proactive Data Rescue into ZFS Disk Recovery for Enhanced Storage Reliability

Authors: Zhi Qiao (University of North Texas), Song Fu (University of North Texas), Hsing-bung Chen (Los Alamos National Laboratory), Michael Lang (Los Alamos National Laboratory)

As tremendous amount of data are generated every day, storage systems store exabytes of data on hundreds of thousands of disk drives. At such a scale, disk failures become the norm. Data recovery takes longer time due to increased disk capacity. ZFS is a widely used filesystem, providing data recovery from corruption. Many factors may affect ZFS's recovery performance in a production environment. Additionally, disk failure prediction techniques enables ZFS to proactively rescue data prior to disk failures. In this poster, we extensively evaluate the recovery performance with a variety of ZFS configurations. We also compare the performance of different data rescue strategies, including post-failure disk recovery, proactive disk cloning, and proactive data recovery. Our proposed analytic model uses the collected zpool utilization data and system configuration to derive the optimal data rescue strategy that best suits the storage array in the current state.

P56: ZoneTier: A Zone-Based Storage Tiering and Caching Co-Design to Integrate SSDs with Host-Aware SMR Drives

Authors: Xuchao Xie (National University of Defense Technology), Liquan Xiao (National University of Defense Technology), David H.C. Du (University of Minnesota)

Integrating solid state drives (SSDs) and host-aware shingled magnetic recording (HA-SMR) drives can potentially build a cost-effective high-performance storage system. However, existing SSD tiering and caching designs in such a hybrid system are not fully matched with the intrinsic properties of HA-SMR drives due to their handling of non-sequential writes (NSWs) from both workloads and data migration. We propose ZoneTier, a zone-based storage tiering and caching co-design, to effectively control all the NSWs by leveraging the host-aware property of HA-SMR drives. ZoneTier exploits the real-time data layouts of HA-SMR zones to optimize zone placements, reshape NSWs generated from zone demotions to HA-SMR drive preferred sequential writes, and transforms the inevitable NSWs to HA-SMR zones to cleaning-friendly write traffics. Our experiments show that ZoneTier can utilize SSDs with high efficiency, minimize relocation overhead, shorten performance recovery time of HA-SMR drives, and finally accomplish better system performance than existing hybrid storage designs.

P57: Adaptive Tier Selection for NetCDF and HDF5

Authors: Jakob Luettgau (German Climate Computing Center), Eugen Betke (German Climate Computing Center), Olga Perevalova (University of Hamburg), Julian Kunkel (German Climate Computing Center), Michael Kuhn (University of Hamburg)

Scientific applications on supercomputers tend to be I/O-intensive. To achieve portability and performance, data description libraries such as HDF5 and NetCDF are commonly used. Unfortunately, the libraries often default to suboptimal access patterns for reading/writing data to multi-tier distributed storage. This work explores the feasibility of adaptively selecting tiers depending on an application's I/O behavior.

P58: Wharf: Sharing Docker Images across Hosts from a Distributed Filesystem

Authors: Chao Zheng (University of Notre Dame), Lukas Rupprecht (IBM), Vasily Tarasov (IBM), Mohamed Mohamed (IBM), Dimitrios Skourtis (IBM), Amit S. Warke (IBM), Dean Hildebrand (IBM), Douglas Thain (University of Notre Dame)

Due to their portability and less overhead compared to traditional virtual machines, containers are becoming an attractive solution for running HPC workloads. Docker is a popular toolset which enables convenient provisioning and management of containers and their corresponding images. However, Docker does not natively support running on shared storage, a crucial requirement in large-scale HPC clusters which are often diskless or access data via a shared burst buffer layer. This lack of distributed storage support can lead to overhead when running containerized HPC applications. In this work, we explore how Docker images can be served efficiently from a shared distributed storage layer. We implement a distributed layer on top of Docker that allows multiple Docker daemons to access container images from a shared filesystem such as IBM Spectrum Scale or NFS. Our design is independent of the underlying storage layer and minimizes the synchronization overhead between different daemons.

P59: Secure Enclaves: An Isolation-Centric Approach for Creating Secure High-Performance Computing Environments

Authors: Ferrol Aderholdt (Oak Ridge National Laboratory), Susan Hicks (Oak Ridge National Laboratory), Thomas Naughton (Oak Ridge National Laboratory), Lawrence Sorillo (Oak Ridge National Laboratory), Blake Caldwell (University of Colorado, Boulder), James Pogge (Tennessee Technological University), Stephen L. Scott (Tennessee Technological University)

High performance computing environments are used for a wide variety of workloads. These systems may process data at various security levels but in so doing are often enclaved at the highest security posture, which may limit usability or performance. The traditional approach used to provide isolation is effective at the creation of secure enclaves, but poses significant challenges with respect to the use of shared infrastructure in HPC environments. We evaluate the use of system-level (i.e., hypervisor-based) and operating system level (i.e., containers) virtualization as well as software defined networking (SDN) as possible mechanisms for secure, isolation-centric enclaves (secure enclaves). We describe our approach to secure HPC enclaves and provide benchmark results for three focus areas (compute, network and data storage) where isolation mechanisms are most significant.

P60: Managing dbGaP Data with Stratus, a Research Cloud for Protected Data

Authors: Evan F. Bollig (University of Minnesota), Graham Allan (University of Minnesota), Benjamin J. Lynch (University of Minnesota), Yectli Huerta (University of Minnesota), Mathew Mix (University of Minnesota), Brent Swartz (University of Minnesota), Edward A. Munsell (University of Minnesota), Joshua Leibfried (University of Minnesota), Naomi Hospodarsky (University of Minnesota)

Modern research computing needs at academic institutions are evolving. While traditional HPC continues to satisfy most workflows, a new generation of researcher has emerged looking for sophisticated, self-service control of compute infrastructure in a cloud-like environment. Often, these demands are not for their own interest, but nonetheless present due to constraints imposed by data governance and protection policies that cannot be satisfied by traditional HPC.

To cater to these modern users, the Minnesota Supercomputing Institute deployed a cloud service for research computing called Stratus. Stratus is designed expressly to satisfy the requirements set forth by the NIH Genomic Data Sharing (GDS) Policy for dbGaP data. It is powered by the Newton version of OpenStack, and backed by Ceph storage. The service offers three features not available on traditional HPC systems: a) on-demand availability of compute resources; b) long-running jobs (i.e., > 30 days); and c) container-based computing with Docker applications.

P61: Cloud Resource Selection Based on PLS Method for Deploying Optimal Infrastructures for Genomic Analytics Application

Authors: Katsunori Miura (Kitami Institute of Technology), Courtney Powell (Hokkaido University), Masaharu Munetomo (Hokkaido University)

This poster proposes a method for determining infrastructures composed of cloud resources that concurrently meet satisfiability and optimality system requirements, such as computational performance, maximum price payable, and deployment location of a genomic analytics application. The input to the proposed method is a mathematical formula that captures the system requirements given by the user, which is defined in accordance with first-order predicate logic, whereas the output is a set of unit clauses representing infrastructures for deploying the genomic analytics application. In the proposed method, an equivalent transformation algorithm is used to generate valid solutions with respect to system requirements, and a genetic algorithm is used to evaluate the optimality of the solutions.

P62: How To Do Machine Learning on Big Clusters

Authors: Thomas Ashby (IMEC), Tom Vander Aa (IMEC), Stanislav Bohm (Technical University of Ostrava), Vojtech Cima (Technical University of Ostrava), Jan Martinovic (Technical University of Ostrava), Vladimir Chupakhin (Janssen Global Services LLC)

Scientific pipelines, such those in chemogenomics machine learning applications, often compose of multiple interdependent data processing tasks. We are developing HyperLoom - a platform for defining and executing workflow pipelines in large-scale distributed environments. HyperLoom users can easily define dependencies between computational tasks and create a pipeline which can then be executed on HPC systems. The high-performance core of HyperLoom dynamically orchestrates the tasks over available resources respecting task requirements. The entire system was designed to have a minimal overhead and to efficiently deal with varying computational times of the tasks. HyperLoom allows to execute pipelines that contain basic built-in tasks, user-defined Python tasks, tasks wrapping third-party applications or a combination of those.

BP

P63: FleCSPH: a Parallel and Distributed Smoothed Particle Hydrodynamics Framework Based on FleCSI

Authors: Julien Loiseau (University of Reims Champagne-Ardenne), Hyun Lim (Brigham Young University), Ben Bergen (Los Alamos National Laboratory), Nicholas Moss (Los Alamos National Laboratory)

In this poster session, we introduce our new parallel and distributed Smoothed Particle Hydrodynamics implementation, FleCSPH, which is based on the open-source framework FleCSI. This framework provides us the basic data structures and runtime required in our work. We intend to provide a parallel and distributed version of Binary, Quad, and Oct trees data structure, respectively for 1, 2 and 3 dimensions dedicated for SPH problems.

Also, we present various test problems in several dimensions that indicate the flexibility and the scalability of our toolkit. For application and tests we simulate classical physics test cases like Sod Shock Tube, Sedov Blast Wave or Fluid Flows. The aim of this work is to simulate binary compact object mergers such as white dwarfs and neutron stars that address many interesting astrophysical phenomena in the universe.

P64: romeoLAB : HPC Training Platform on HPC facility

Authors: Jean-Matthieu Etancelin (University of Reims Champagne-Ardenne), Arnaud Renard (University of Reims Champagne-Ardenne)

In this pre-exascale era, we are observing a dramatic increase of the necessity of computer science courses dedicated to parallel programming with advanced technologies on hybrid architectures. The full hybrid cluster Romeo has long been used for that purpose in order to train master students and cluster users. The main issue for trainees is the cost of accessing and exploiting a

production facility in a pedagogic context. The use of some specific techniques and software (SSH, workload manager, remote file system, ...) is mandatory without being part of courses prerequisites nor pedagogic objectives. The romeoLAB platform we developed at ROMEO HPC Center is an online interactive pedagogic platform for HPC technologies courses. Its main purpose is to simplify the process of resource usage in order to focus on the taught subjects. This paper presents the context, the romeoLAB architecture and its motivations, usages and pedagogical contents.

P65: CAPES: Unsupervised System Performance Tuning Using Neural Network-Based Deep Reinforcement Learning

Authors: Yan Li (*University of California, Santa Cruz*), Kenneth Chang (*University of California, Santa Cruz*), Oceane Bel (*University of California, Santa Cruz*), Ethan Miller (*University of California, Santa Cruz*), Darrell Long (*University of California, Santa Cruz*)

Parameter tuning is an important task of storage performance optimization. Current practice usually involves numerous tweak-benchmark cycles that are slow and costly. To address this issue, we developed CAPES, a model-less deep reinforcement learning-based unsupervised parameter tuning system driven by a deep neural network (DNN). It is designed to find optimal values for computer systems that have tunable parameters when human tuning can be costly and often cannot achieve optimal performance. CAPES takes periodic measurements of a target computer system's state, and trains a DNN which uses Q-learning to suggest changes to the system's current parameter values. CAPES is minimally intrusive, and can be deployed into a production system to collect training data and suggest tuning actions during the system's daily operation. Evaluation of a prototype on a Lustre file system demonstrates an increase in I/O throughput up to 45% at saturation point.

P66: Analyzing Multi-Layer I/O Behavior of HPC Applications

Authors: Ronny Tschüter (*Technical University Dresden*), Christian Herold (*Technical University Dresden*), Bert Wesarg (*Technical University Dresden*), Matthias Weber (*Technical University Dresden*)

In this work, we present an approach to analyze multi-layer I/O behavior of HPC applications. For a comprehensive analysis it is not sufficient to track I/O operations on a single layer because applications may combine multiple I/O paradigms (e.g., MPI I/O, NetCDF, HDF5). Furthermore, I/O libraries may use another I/O paradigm internally. With our approach, I/O activities can be captured at any layer. This work introduces methodologies to intercept calls to I/O libraries and presents an event design to record applications' I/O activities. We implement our approach in the Score-P measurement system and prove its usability with a study of the Met Office/NERC Cloud Model (MONC) code.

P67: Measuring I/O Behavior on Upcoming Systems with NVRAM

Authors: Christian Herold (*Technical University Dresden*), Ronny Tschüter (*Technical University Dresden*)

Upcoming HPC systems will use NVRAM to address existing I/O bottlenecks. The I/O performance thereby is one of the keys for the exascale challenges. NVRAM introduces a new level in the memory hierarchy and can be utilized by different technologies, e.g. memory mapped files or block transfer operations. Using NVRAM without considering the complete hierarchy may lead to an inefficient usage and bad I/O performance. Therefore, in this work, we evaluate techniques for measuring the I/O behavior of applications that utilize NVRAM in various use cases. We extended the application tracing of our tool Score-P and introduced several metrics as well as events for different use cases of NVRAM.

P68: Continuous Clock Synchronization for Accurate Performance Measurements

Authors: Johannes Ziegenbalg (*Technical University Dresden*), Matthias Weber (*Technical University Dresden*)

Clock synchronization is a key prerequisite for accurate performance analysis. This is particularly true for most HPC systems, due to the lack of a system-wide timer. The effects of different clock properties lead to false measurements and wrong conclusions. Often these errors remain undetected by the user. Different environmental effects cause continuous changes to clock properties over time. This behavior is ignored by the usual post-mortem approach for clock synchronization. In order to improve time measurement accuracy we implemented a method for continuous clock synchronization. In this work, we share our experiences and draw conclusions for tool development.

P69: Portable Methods for Measuring Cache Hierarchy Performance

Authors: Tom Deakin (*University of Bristol*), James Price (*University of Bristol*), Simon McIntosh-Smith (*University of Bristol*)

There has been a recent influx of different processor architecture designs into the market, with many of them targeting HPC applications. When estimating application performance, developers are used to considering the most common figures of merit, such as peak FLOP/s, memory bandwidth, core counts, and so on. In this study, we present a detailed comparison of on-chip memory bandwidths, including single core and aggregate across a node, for a set of next-generation CPUs.

We do this in such a way as to be portable across different architectures and instruction sets. Our study indicates that, while two processors might look superficially similar when only considering the common figures of merit, those two processors might have radically different on-chip memory bandwidth, a fact which may be crucial when understanding observed application performance. Our results and methods will be made available on GitHub to aid the community in evaluating cache bandwidths.

P70: FFT, FMM, and Multigrid on the Road to Exascale: Performance Challenges and Opportunities

Authors: Huda Ibeid (University of Illinois), Luke Olson (University of Illinois), William Gropp (University of Illinois)

FFT, FMM, and multigrid methods are widely used fast and highly scalable solvers for elliptic PDEs. However, emerging systems are introducing challenges in comparison to current petascale computers. The International Exascale Software Project Roadmap identifies several constraints on the design of exascale software. Challenges include massive concurrency, energy efficiency, resilience management, exploiting the high performance of heterogeneous systems, and utilizing the deeper and more complex memory hierarchy expected at exascale. In this study, we perform model-based comparison of the FFT, FMM, and multigrid methods in the context of these constraints and use the performance models to offer predictions about the methods performance on possible exascale system configurations, based on current technology trends.

P71: Is ARM Software Ecosystem Ready for HPC?

Authors: Fabio Banchelli (Barcelona Supercomputing Center), Daniel Ruiz (Barcelona Supercomputing Center), Ying Hao Xu Lin (Barcelona Supercomputing Center), Filippo Mantovani (Barcelona Supercomputing Center)

In recent years, the HPC community has increasingly grown its interest towards the ARM architecture with research projects targeting primarily the deployment of ARM-based clusters. Attention is usually given to hardware platforms, however the availability of a mature software ecosystem and the possibility of running large and complex HPC applications plays a key role in the consolidation process of a new technology.

For this reason in this poster we present a preliminary evaluation of the ARM system software ecosystem, limited here to the ARM HPC Compiler and the ARM Performance Libraries, together with a porting and testing of three fairly complex HPC code suites: QuantumESPRESSO, WRF, and FEniCS.

These codes have been proposed as HPC challenges during the last two editions of the Student Cluster Competition at ISC where all the authors have been involved operating an ARM-based cluster and awarded with the Fan Favorite award.

P72: New Developments for PAPI 5.6+

Authors: Anthony Danalis (University of Tennessee), Heike Jagode (University of Tennessee), Vince Weaver (University of Maine), Yan Liu (University of Maine), Jack Dongarra (University of Tennessee)

The HPC community has relied on PAPI to track low-level hardware operations for over 15 years. In that time, the needs of software developers have changed immensely, and the PAPI team aims to meet these demands through a better understanding of deep and heterogeneous memory hierarchies and finer-grain power-management support.

This poster demonstrates how PAPI enables power-tuning to reduce overall energy consumption without, in many cases, a loss in performance. Furthermore, we discuss efforts to develop microbenchmarks intended to assist application developers who are interested in performance analysis by automatically categorizing and disambiguating performance counters. Finally, the poster illustrates efforts to update PAPI's internal sanity checks, designed to inspect that PAPI's predefined events are in fact measuring the values they claim to measure, and modernize the implementation of critical API functions, e.g., PAPI_read(), and the sampling interface so that more information can be captured and reported with lower overhead.

BP

P73: HPC Production Job Quality Assessment

Authors: Omar Aaziz (New Mexico State University), Jonathan Cook (New Mexico State University)

Users of HPC systems would benefit from more feedback about the quality of their application runs, such as knowing whether or not the performance of a particular run was good, or whether the resources requested were enough, or too much. Such feedback requires more information to be kept regarding production application runs, and requires some analytics to assess any new runs. In this research, we assess the practicality of using job data, system data, and hardware performance counters in a near-zero overhead manner to assess job performance, in particular whether or not the job runtime was in line with expectations from historical application performance. We show over four proxy applications and two real application that our assessment is within 10% of actual performance.

P74: A Methodology for Bridging the Native and Simulated Executions of Parallel Applications

Authors: Ali Mohammed (University of Basel), Ahmed Eleliemy (University of Basel), Florina M. Ciorba (University of Basel)

Simulation is considered as the third pillar of science, following experimentation and theory. Bridging the native and simulated executions of parallel applications is needed for attaining trustworthiness in simulation results. Yet, bridging the native and simulated executions of parallel applications is challenging. This work proposes a methodology for bridging the native and simulated executions of message passing parallel applications on high performance computing (HPC) systems in two steps: Expression of the software characteristics, and representation and verification of the hardware characteristics in the simulation. This work exploits the capabilities of the SimGrid [3] simulation toolkit's interfaces to reduce the effort of bridging the native and simulated executions of a parallel application on an HPC system. For an application from computer vision, the simulation of its parallel execution using straightforward parallelization on an HPC cluster approaches the native performance with a minimum relative percentage difference of 5.6%.

P75: Model-Agnostic Influence Analysis for Performance Data

Authors: *Rahul Sridhar (University of California, Irvine; Lawrence Livermore National Laboratory), Rushil Anirudh (Lawrence Livermore National Laboratory), Jayaraman J. Thiagarajan (Lawrence Livermore National Laboratory), Nikhil Jain (Lawrence Livermore National Laboratory), Todd Gamblin (Lawrence Livermore National Laboratory)*

Execution time of an application is affected by several performance parameters, e.g. number of threads, decomposition, etc. Hence, an important problem in high performance computing is to study the influence of these parameters on the performance of an application. Additionally, quantifying the influence of individual parameter configurations (data samples) on performance also aids in identifying sub-domains of interest in high-dimensional parameter spaces. Conventionally, such analysis is performed using a surrogate model, which introduces its own bias that is often non-trivial to undo, leading to inaccurate results. In this work, we propose an entirely data-driven, model-agnostic influence analysis approach based on recent advances in analyzing functions on graphs. We show that the problem of identifying influential parameters (features) and configurations (samples) can be effectively addressed within this framework.

P76: A Compiler Agnostic and Architecture Aware Predictive Modeling Framework for Kernels

Authors: *William Killian (University of Delaware), Ian Karlin (Lawrence Livermore National Laboratory), David Beckingsale (Lawrence Livermore National Laboratory), John Cavazos (University of Delaware)*

Multi-architecture machines make program characterization for modeling a regression outcome difficult. Determining where to offload compute-dense portions requires accurate prediction models for multiple architectures. To productively achieve portable performance across these diverse architectures, users are adopting portable programming models such as OpenMP and RAJA.

When adopted, portable models make traditional high-level source code analysis inadequate for program characterization. In this poster, we introduce a common microarchitecture instruction format (ComIL) for program characterization. ComIL is capable of representing programs in an architecture-aware compiler-agnostic manner. We evaluate feature extraction with ComIL by constructing multiple regression-objective models for performance (execution time) and correctness (maximum absolute error). These models perform better than the current state of the art -- achieving a mean error rate of only 4.7% when predicting execution time. We plan to extend this work to handle multiple architectures concurrently and evaluate with more representative physics kernels.

P77: AutoTuneTMP: Auto Tuning in C++ With Runtime Template Metaprogramming

Authors: *David Pfander (University of Stuttgart), Dirk Pflüger (University of Stuttgart)*

Maximizing the performance on modern hardware platforms has become more and more difficult, due to different levels of parallelism and complicated memory hierarchies. Auto tuning can help developers to address these challenges by writing code that automatically adjusts to the underlying hardware platform. AutoTuneTMP is a new C++-based auto tuning framework that uses just-in-time compilation to enable runtime-instantiable C++ templates. We use C++ template metaprogramming to provide data structures and algorithms that can be used to develop tunable compute kernels. These compute kernels can be tuned with different optimization strategies. We demonstrate for a first prototype the applicability and usefulness of our approach by tuning 6 parameters of a DGEMM implementation, achieving 68% peak performance on an Intel Skylake processor.

P78: Performance Evaluation of Graph500 Considering CPU-DRAM Power Shifting

Authors: *Yuta Kakibuka (Kyushu University), Yuichiro Yasui (Kyushu University), Takatsugu Ono (Kyushu University), Katsuki Fujisawa (Kyushu University), Koji Inoue (Kyushu University)*

There are power constraints on computer systems which comes from technical, costly or social demands. Power wall is one of the most serious issues for post petascale high-performance computing. A promising solution to tackle this problem is to effectively manage power resources based on the characteristics of workloads. In power constrained computing, the key is to translate the limited power budget into sustained performance effectively. To achieve this goal, assigning the appropriate amount of power budget to each hardware component, or power shifting, is a critical challenge.

In this work, we focus on large-scale graph processing. Graph analysis algorithms are increasing its importance with growing demands of big data analysis. However, the impact of power constraint on the performance of graph processing application is not declared. Our work is the performance evaluation of Graph500 under power constraints to CPU and DRAM.

P79: Porting the Opacity Client Library to a CPU-GPU Cluster Using OpenMP 4.5

Authors: *Jason S. Kimko (College of William and Mary), Michael M. Pozulp (Lawrence Livermore National Laboratory), Riyaz Haque (Lawrence Livermore National Laboratory), Leopold Grinberg (IBM)*

The poster accompanying this summary exhibits our experience porting the Opacity client library to IBM's "Minsky" nodes using OpenMP 4.5. We constructed a GPU-friendly container class that mimics existing library functionality. We benchmarked our implementation on Lawrence Livermore National Laboratory's (LLNL) RZManta, a Minsky cluster. In our benchmarks on a single POWER8 CPU and Tesla P100 GPU, we observed up to a 4x speedup including CPU-GPU data transfers and up to a 30x speedup excluding data transfers. Optimizing to reduce register pressure and increase occupancy may improve speedups. Our results demonstrate a successful and beneficial library port to the CPU-GPU architecture.

P80: Adaptive Loop Scheduling with Charm++ to Improve Performance of Scientific Applications

Authors: Vivek Kale (*University of Southern California*), Harshitha Menon (*Lawrence Livermore National Laboratory*), Karthik Senthil (*University of Illinois*)

Supercomputers today employ a large number of cores on each node. The Charm++ parallel programming system provides an intelligent runtime which has been highly effective at providing dynamic load balancing across nodes of a supercomputer. Modern multi-core nodes present new challenges and opportunities for Charm++. The large degree of over-decomposition required may lead to high overhead. We modified the Charm++ Runtime System (RTS) to assign Charm++ objects to nodes, thus reducing over-decomposition, and spreading work across cores via parallel loops. We modify a library of the Charm++ software suite that supports loop parallelism by adding to it a loop scheduling strategy that maximizes load balance across cores while minimizing data movement. We tune parameters of the RTS and the loop scheduling strategy to improve performance of benchmark codes run on a variety of architectures. Our technique improves performance of a Particle-in-Cell code run on the Blue Waters supercomputer by 17.2%.

BP

P81: Offloading Python Kernels to Micro-Core Architectures

Authors: Nick Brown (*University of Edinburgh*)

Micro-core architectures combine many low memory, low power computing cores together in a single package. These can be used as a co-processor or standalone but due to limited on-chip memory and esoteric nature of the hardware, writing efficient parallel codes for them is challenging. We previously developed ePython, a low memory (24Kb) implementation of Python supporting the rapid development of parallel Python codes and education for these architectures. In this poster we present our work on an offload abstraction to support the use of micro-cores as an accelerator. Programmers decorate specific functions in their Python code, running under any Python interpreter on the host, with our underlying technology then responsible for the low-level data movement, scheduling and execution of kernels on the micro-cores. Aimed at education and fast prototyping, a machine learning code for detecting lung cancer, where computational kernels are offloaded to micro-cores, is used to illustrate the approach.

P82: Performance Evaluation of the NVIDIA Tesla P100: Our Directive-Based Partitioning and Pipelining vs. NVIDIA's Unified Memory

Authors: Xuwen Cui (*Virginia Tech*), Thomas R. W. Scogland (*Lawrence Livermore National Laboratory*), Bronis R. de Supinski (*Lawrence Livermore National Laboratory*), Wu-chun Feng (*Virginia Tech*)

We need simpler mechanisms to leverage the performance of accelerators, such as GPUs, in supercomputers. Programming models like OpenMP offer simple-to-use but powerful directive-based offload mechanisms. By default, these models naively copy data to or from the device without overlapping computation. Achieving performance can require extensive hand-tuning to apply optimizations such as pipelining. Users must manually partition data whenever it exceeds device memory. Our directive-based partitioning and pipelining extension for accelerators overlaps data transfers and kernel computation without explicit user data-splitting. We compare a prototype implementation of our extension to NVIDIA's Unified Memory on the Pascal P100 GPU and find that our extension outperforms Unified Memory on average by 68% for data sets that fit into GPU memory and 550% for those that do not.

P83: Contracts for Message-Passing Programs

Authors: Ziqing Luo (*University of Delaware*), Stephen F. Siegel (*University of Delaware*)

Verification of message-passing parallel programs is challenging because of the state-explosion problem. A procedure-level contract system for parallel programs can help overcome this challenge by verifying contracts individually, in a modular way. A contract system is used to specify intended behaviors of programs. Contracts can be checked at run-time or verified formally. There is a mature theory of contracts for sequential programs, but little work has been done on parallel programs, and even less for message-passing parallel programs. We developed a theory of contracts for message-passing programs and realize this theory in a contract language for programs that use the Message Passing Interface (MPI). We are also developing a verification tool that uses symbolic execution and model checking techniques to prove that MPI programs satisfy their contracts.

P84: PRESAGE: Selective Low Overhead Error Amplification for Easy Detection

Authors: Vishal Chandra Sharma (*University of Utah*), Arnab Das (*University of Utah*), Ian Briggs (*University of Utah*), Ganesh Gopalakrishnan (*University of Utah*), Sriram Krishnamoorthy (*Pacific Northwest National Laboratory*)

Soft-errors remain a vexing challenge. Today's error detectors either come with high false positives or high omissions. Our work focuses on not losing an error sown, but amplifying it so that cheap detection is enabled.

Consider structured address generation in loops where data from a base address plus offset is used. An erroneous offset no longer crashes today's applications thanks to large address spaces; instead, it silently corrupts data (unintended fetch). We relativize address generation using LLVM, thus each new address is not base plus offset but previous relative address plus offset. If one address is corrupted, all future address are corrupted in a chain. This permits efficient loop exit-point detection.

Relativization has low overhead, actually lowered by some ISAs down to zero. These advantages survive crucial compiler memory

access optimizations. We demonstrate 100% SDC detection for a class of benchmarks with respect to structured address protection.

P85: GPU Mekong: Simplified Multi-GPU Programming Using Automated Partitioning

Authors: *Alexander Matz (University of Heidelberg)*

GPU accelerators are pervasively used in the HPC community, because they provide excellent computational performance at a reasonable power efficiency. While programming single-GPU applications is comparatively productive, programming multiple GPUs using data-parallel languages is tedious and error prone as the user has to manually orchestrate data movements and kernel launches.

The Mekong research project is driven by the motivation to improve productivity of multi-GPU systems by compiler based partitioning of single-device data-parallel programs. Key to scalable performance improvement is the resolution of data dependencies between kernels and the orchestration of these kernels. Mekong relies on polyhedral compilation to identify memory access patterns in order to compile a single-GPU application into a multi-GPU application.

In this work, the Mekong project is introduced and its components explained. While the tool is still under development, preliminary results are available and are shortly discussed demonstrating the potential of this approach.

P86: HyGraph: High Performance Graph Processing on Hybrid CPU+GPUs platforms

Authors: *Stijn Heldens (University of Twente), Ana Lucia Varbanescu (University of Amsterdam), Alexandru Iosup (Vrije University Amsterdam)*

Graph analytics is becoming increasingly important in many domains, such as in biology, social sciences, and data mining. Many large-scale graph-processing systems have been proposed, either targeting distributed clusters or GPU-based accelerated platforms. However, little research exists on designing systems for hybrid CPU-GPU platforms, i.e., exploiting both the CPU and the GPU efficiently.

In this work, we present HyGraph, a novel graph-processing system for hybrid platforms which delivers performance by using both the CPU and GPUs concurrently. Its core feature is dynamic scheduling of tasks onto both the CPU and the GPUs, thus providing load balancing, contrary to the state-of-the-art approach based on static partitioning. Additionally, communication overhead is minimized by overlapping computation and communication.

Our results demonstrate that HyGraph outperforms CPU-only and GPU-only solutions, delivering close-to-optimal performance. Moreover, it supports large-scale graphs which do not fit into GPU memory and is competitive against state-of-the-art systems.

P87: EoCoE Performance Benchmarking Methodology for Renewable Energy Applications

Authors: *Paul Gibbon (Forschungszentrum Juelich), Matthieu Haefele (French Alternative Energies and Atomic Energy Commission), Sebastian Luehrs (Forschungszentrum Juelich, Juelich Supercomputing Center)*

An optimisation strategy developed by the Energy Oriented Centre of Excellence (EoCoE) is presented for computational models used in a variety of renewable energy domains. It is found that typical applications in this comparatively new sector exhibit a huge range of HPC maturity, from simple parallelization needs to near-exascale readiness. An essential part of this process has therefore been the introduction of a flexible, quantitative performance assessment of applications using the benchmarking tool JUBE to automatically extract up to 28 different metrics taken with several state-of-the-art performance tools. An initial hands-on workshop to establish this baseline status is consolidated by follow-up actions by joint code-teams comprising members of both developer groups and HPC centres involved with the EoCoE consortium. Examples of early successes achieved with this policy are given, together with an outlook on challenges faced for energy applications with next-generation, pre-exascale architectures.

P88: PetaVision Neural Simulation Toolbox on Intel KNLs

Authors: *Boram Yoon (Los Alamos National Laboratory), Pete Schultz (Los Alamos National Laboratory, New Mexico Consortium), Garrett Kenyon (Los Alamos National Laboratory, New Mexico Consortium)*

We implemented a large-scale neuromorphic algorithm called the Sparse Prediction Machine (SPM), on the Los Alamos Trinity supercomputer. Specifically, we used PetaVision, an open source high-performance neural simulation toolbox, to implement a 4-layer SPM applied to ImageNet video. Various optimization techniques were applied to efficiently utilize up to 8192 KNL nodes. The goal of the SPM is to predict future states of a system from a sequence of previous states, or in the case of video, to predict a subsequent frame from previous frames. In our training, the SPM was able to predict the 8th frame from the preceding 7 frames, including successful separation of foreground and background motion.

P89: Desh: Deep Learning for HPC System Health Resilience

Authors: *Anwesha Das (North Carolina State University), Abhinav Vishnu (Pacific Northwest National Laboratory), Charles Siegel (Pacific Northwest National Laboratory), Frank Mueller (North Carolina State University)*

HPC systems are well known to endure service downtime due to increasing failures. With enhancements in HPC architectures and

design, enabling resilience is extremely challenging due to component scaling and absence of well defined failure indicators. HPC system logs are notorious to be complex and unstructured. Efficient fault prediction to enable proactive recovery mechanisms is the need of the hour to make such systems more robust and reliable. This work addresses such faults in computing systems using a recurrent neural network based technique called LSTM (long short-term memory).

We present our framework Desh : Deep Learning for HPC System Health, which entails a procedure to diagnose and predict failures with acceptable lead times. Desh indicates prospects of indicating failure indicators with enhanced training and classification for generic applicability to other systems. This deep learning based framework gives interesting insights for further work on HPC system reliability.

P90: Global Survey of Energy and Power-Aware Job Scheduling and Resource Management in Supercomputing Centers

Authors: Siddhartha Jana (Intel Corporation), Gregory A. Koenig (Energy Efficient HPC Working Group), Matthias Maiterth (Intel Corporation), Kevin T. Pedretti (Sandia National Laboratories), Andrea Borghesi (University of Bologna), Andrea Bartolini (ETH Zurich), Bilel Hadri (King Abdullah University of Science and Technology), Natalie J. Bates (Energy Efficient HPC Working Group)

The two major driving forces that are leading centers to investigate dynamic power and energy management strategies are (1) limitations to the availability of resources from the electricity service provider, and (2) the desire to spend limited budgets on computational cycles rather than infrastructure requirements such as electricity. In addition, supercomputer systems have increasingly rapid, unpredictable, and large power fluctuations. In addition, electricity service providers may request supercomputing centers to change their timing and/or magnitude of demand to help address electricity supply constraints. To adapt to this new landscape, centers may employ energy and power-aware job scheduling and resource management (EPA-JSRM) strategies to dynamically, and in real-time, control their electricity demand. This poster summarizes the lessons learned from one of the first global surveys of supercomputing centers that are actively investigating such strategies.

P91: Assessing the Availability of Source Code in Computational Physics

Authors: Matthew Krafczyk (National Center for Supercomputing Applications, University of Illinois), Victoria Stodden (University of Illinois), Yantong Zheng (University of Illinois), David Wong (University of Illinois)

Replicability of scientific work based on computation is a subject which has been receiving increased scrutiny recently. One approach to replicating a computational finding is to run the original source code. Availability of source code however is not routine; Only 3/33 computationally based article authors released source code from JASA in 2006, and a 2015 study showed that only 44% of computer science article authors released their source code. The field of Computational Physics has yet to be examined in this way, nor has the effect of author knowledge of such a study been measured.

We present our findings regarding the availability of source code in recent articles of the Journal of Computational Physics as well as how author knowledge of the study affects their willingness to make source code available. This work extends current reproducibility efforts being explored by the ACM, SIGHPC, and the SC conference community.

P92: Characterization and Comparison of Application Resilience for Serial and Parallel Executions

Authors: Kai Wu (University of California, Merced), Qiang Guan (Los Alamos National Laboratory, Ultrascale Systems Research Center), Nathan DeBardeleben (Los Alamos National Laboratory, Ultrascale Systems Research Center), Dong Li (University of California, Merced)

Soft error of exascale application is a challenge problem in modern HPC. In order to quantify an application's resilience and vulnerability, the application-level fault injection method is widely adopted by HPC users. However, it is not easy since users need to inject a large number of faults to ensure statistical significance, especially for parallel version program. Normally, parallel execution is more complex and requires more hardware resources than its serial execution. Therefore, it is essential that we can predict error rate of parallel application based on its corresponding serial version. In this poster, we characterize fault pattern in serial and parallel executions. We find first there are same fault sources in serial and parallel execution. Second, parallel execution also has some unique fault sources compared with serial executions. Those unique fault sources are important for us to understand the difference of fault pattern between serial and parallel executions.

P93: Spacehog: Evaluating the Costs of Dedicated Resources to In Situ Analysis

Authors: Rebecca Kreitinger (University of New Mexico), Scott Levy (Sandia National Laboratories), Kurt B. Ferreira (Sandia National Laboratories), Patrick Widener (Sandia National Laboratories)

Using in situ analytics requires that computational resources be shared between the simulation and the analysis. With space-sharing, there is a possibility for contention over these shared resources such as memory, memory bandwidth, network bandwidth, or filesystem bandwidth. In our analysis, we explore the sensitivity of different applications with a set of microbenchmarks that are representative of analytics that may be used with scientific simulation. These tasks are modeled using a library called libspacehog. The experimentation consisted of examining three different dimensions of how simulation workloads might be space-shared with analysis codes. The results indicate that contention does need to be considered when applying in situ analytic techniques and can be of greater concern than simply the number of analysis processes or overall process density. This research provides an explanation on how the application's performance is affected by space-sharing to further understand in situ analytic techniques.

P94: Fully Hierarchical Scheduling: Paving the Way to Exascale Workloads

Authors: Stephen Herbein (University of Delaware), Tapasya Patki (Lawrence Livermore National Laboratory), Dong H. Ahn (Lawrence Livermore National Laboratory), Don Lipari (Lawrence Livermore National Laboratory), Tamara Dahlgren (Lawrence Livermore National Laboratory), David Domyancic (Lawrence Livermore National Laboratory), Michela Taufer (University of Delaware)

Exascale workloads, such as uncertainty quantification (UQ), represent an order of magnitude increase in both scheduling scale and complexity. Batch schedulers with their decades-old, centralized scheduling model will fail to address the needs of these new workloads. To address these upcoming challenges, we claim that HPC schedulers must transition from the centralized to the fully hierarchical scheduling model. In this work, we assess the impact of the fully hierarchical model on both a synthetic stress test and a real-world UQ workload. We observe over a 100x increase in scheduler scalability on the synthetic stress test and a 37% decrease in the runtime of real-world UQ workloads under the fully hierarchical model. Our empirical results demonstrate that the fully hierarchical scheduling model can overcome the limitations of existing schedulers to meet the needs of UQ and other exascale workloads.

BP

P95: GEOPM: A Scalable Open Runtime Framework for Power Management

Authors: Siddhartha Jana (Intel Corporation), Asma H. Al-rawi (Intel Corporation), Steve S. Sylvester (Intel Corporation), Christopher M. Cantalupo (Intel Corporation), Brad Geltz (Intel Corporation), Brandon Baker (Intel Corporation), Jonathan M. Eastep (Intel Corporation)

The power scaling challenges associated with exascale systems is a well-known issue. In this work, we introduce the Global Extensible Open Power Manager (GEOPM): a tree-hierarchical, open source runtime framework we are contributing to the HPC community to foster increased collaboration and accelerated progress toward software-hardware co-designed energy management solutions that address exascale power challenges and improve performance and energy efficiency in current systems. Through its plugin extensible architecture, GEOPM enables rapid prototyping of new energy management strategies. Different plugins can be tailored to the specific performance or energy efficiency priorities of each HPC center. To demonstrate the potential of the framework, this work develops an example plugin for GEOPM. This power rebalancing plugin targets power-capped systems and improves efficiency by minimizing job time-to-solution within a power budget. Our results demonstrate up to 30% improvements in the time-to-solution of CORAL system procurement benchmarks on a Xeon Phi cluster.

P96: Correcting Detectable Uncorrectable Errors in Memory

Authors: Grzegorz Pawelczak (University of Bristol), Simon McIntosh-Smith (University of Bristol)

With the expected decrease in Mean Time Between Failures, Fault Tolerance has been identified as one of the major challenges for exascale computing. One source of faults are soft errors caused by cosmic rays, which can cause bit corruptions to the data held in memory. Current solutions for protection against these errors include Error Correcting Codes, which can detect and/or correct these errors. When an error that can be detected but not corrected occurs, a Detectable Uncorrectable Error (DUE) results, and unless checkpoint-restart is used, the system will usually fail. In our work we present a probabilistic method of correcting DUEs which occur in the part of the memory where the program instructions are stored. We devise a correction technique for DUEs for the ARM A64 instruction set which combines extended Hamming code with Cyclic Redundancy Check code to provide near 100% Successful Correction Rate of DUEs.

P97: Profile Guided Kernel Optimization for Individual Container Execution on Bare-Metal Container

Authors: Kuniyasu Suzuki (National Institute of Advanced Industrial Science and Technology), Hidetaka Koie (National Institute of Advanced Industrial Science and Technology), Ryousei Takano (National Institute of Advanced Industrial Science and Technology)

Container technologies become popular on supercomputers as well as in data centers. They use a container image as a package of an application, which makes easy to customize the computing environment. Unfortunately, they are not allowed to change the kernel. It means that an application cannot get the benefit of kernel optimization. Especially, Profile Guided Kernel Optimization (PGKO) is not applied.

Bare-Metal Container (BMC) tries to solve this problem. BMC utilizes remote machine management technologies (IPMI, Intel AMT, and WakeupOnLAN) to run a container image on a remote machine with a suitable Linux kernel. It enables to use PGKO easily, because the trial execution to get a profile and the optimized execution are executed automatically. Furthermore, BMC easily changes the target machine, and the user can compare the effects. We measured the performance of PGKO on big data workloads (Apache and Redis) on Xeon and i7 and found the difference.

P98: Energy Efficiency in HPC with Machine Learning and Control Theory

Authors: Connor Imes (University of Chicago), Steven Hofmeyr (Lawrence Berkeley National Laboratory), Henry Hoffmann (University of Chicago)

Performance and power management in HPC has historically favored a race-to-idle approach in order to complete applications as quickly as possible, but this is not energy-efficient on modern systems. As we move toward exascale and hardware over-provisioning, power management is becoming more critical than ever for HPC system administrators, opening the door for more balanced approaches to performance and power management. We propose two projects to address balancing application performance and system power consumption in HPC during application runtime, using closed loop feedback designs based on the

Self-Aware Computing Model to observe, decide, and act.

P99: The Intersection of Big Data and HPC: Using Asynchronous Many Task Runtime Systems for HPC and Big Data

Authors: Joshua Daniel Suetterlein (*Pacific Northwest National Laboratory*), Joshua Landwehr (*Trovares Inc*), Andres Marquez (*Pacific Northwest National Laboratory*), Joseph Manzano (*Pacific Northwest National Laboratory*), Kevin Barker (*Pacific Northwest National Laboratory*), Guang Gao (*University of Delaware*)

Although the primary objectives of the HPC and Big data fields seem disparate, HPC is beginning to suffer from a growing size of its workloads and the limitation of its techniques to handle large amount of data. This places interesting research challenges for both HPC and Big Data on how to marriage both fields together. This poster presents a case study which uses Asynchronous Many Task Runtimes (AMTs) as an exploratory vehicle to highlight possible solutions to these challenges. AMTs presents the unique opportunity for better load balancing, reconfigurable schedulers and data layouts that can take advantage of introspection frameworks, and the ability to exploit a massive amount of concurrency. We use the Performance Open Community Runtime (P-OCR) as a vehicle to port MapReduce operators to the HPC realm. We conduct experiments with both strong and weak scaling experimental format using WordCount and TeraSort as our kernels.

Thursday, November 16th

Room: EXDO Event Center | 1399 35th St, Denver, CO 80205

6:00 pm - 9:00 pm

Technical Program Reception

Technical Program Reception

HPC Connects You to an Out of This World Experience! To thank our Technical Program attendees and to encourage continued global and inter-galactic interactions, SC17 is hosting a conference reception for all Technical Program attendees at the [EXDO Event Center](#), Denver's most versatile event space in the heart of the city's vibrant River North Art District.

This year's event features a sci-fi theme, with a virtual reality lounge, a green-screen photo booth, an Emulator Screen DJ, a live painting demonstration, and much more.

Even the food will be out of this world - with a "Mad Scientist" salad station, "Fire-in-the-Sky" flame wall, "Close Encounters of the Third" kind pasta station, "Lost-in-the-Pacific" station and of course "Alien Pod" s'mores. **Guests are encouraged to come dressed as their favorite sci-fi character!** So plan ahead.

A Tech Program badge, event ticket, and government-issued photo ID are required to attend this event. Busing will be provided to / from the Convention Center from the F Lobby from 5:30pm until 9:00pm. Note, you will need your badge and your event ticket to get on the shuttle.

Scientific Visualization & Data Analytics Showcase [\(back to top\)](#)

Tuesday, November 14th

Room: Mile High Prefunction
8:30 am - 5:00 pm

Scientific Visualization and Data Analytics Showcase Posters

Visualization of Decision-Making Support (DMS) Information for Responding to a Typhoon-Induced Disaster

Authors: Dongmin Jang (Korea Institute of Science and Technology Information), Jin-Hee Yuk (Korea Institute of Science and Technology Information), Junghyun Park (Korea Institute of Science and Technology Information), Jooneun An (Korea Institute of Science and Technology Information), Minsu Joh (Korea Institute of Science and Technology Information)

A high-resolution coupled atmosphere, ocean, and inundation (flood) modeling and simulation system was developed for scientific, accurate, fast, and efficient forecasting of typhoon-induced disasters. This is based on the KISTI decision-making support system (K-DMSS). Our prediction system consists of a typhoon, surge/wave, and flooding prediction and analysis systems (TPAS, SPAS, and FPAS). In this research, we simulated Typhoon 'CHABA' (1618), which was ranked third among the most intense tropical cyclones and was the most powerful typhoon in the Republic of Korea (South Korea) in 2016. The CHABA-induced storm surge and inundation were simulated using our prediction and analysis system. To understand intuitively the changes of physical phenomena and damage caused by the typhoon, numerical data sets produced by the prediction and analysis systems were visualized using VAPOR (Visualization and Analysis Platform for Ocean, atmosphere, and solar Researchers, and which is a part of the K-DMSS) as one of the visualization systems.

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Visualizations of a High-Resolution Global-Regional Nested, Ice-Sea-Wave Coupled Ocean Model System

Authors: Kangyou Zhong (Sun Yat-Sen University), Danya Xu (Sun Yat-Sen University), Changsheng Chen (University of Massachusetts, Dartmouth; Sun Yat-Sen University), Yutong Lu (Sun Yat-Sen University), Wenjie Dong (Sun Yat-Sen University), Jiang Li (Sun Yat-Sen University)

A multi-scale, global-regional nested ocean modeling system based on the unstructured grid Finite Volume Community Ocean Model (FVCOM) has been deployed on the Tianhe-2 supercomputer providing 24/7 marine forecasting since September 2016. The modeling system is part of the Sun Yat-Sen University Community Integrated Model (SYCIM) project for developing a new generation Earth System Model to explore the physical mechanisms of climate change. With a horizontal resolution up to ~17 m, this high-resolution modeling system can properly resolve the complex dynamical interactions of estuary, near shore coast, continental shelf, and deep ocean basins. The submitted animation shows the modeled global surface waves distribution pattern and propagation in the world ocean basins and in the South China Sea. Additionally, the variations of the global sea surface temperature, Arctic and Antarctic sea ice thickness and extension simulated by SYCIM is also presented. The animation can not only be used to visualize the big data for scientific research but also as a good example to demonstrate dynamical variations of the ocean to non-scientific communities. We hope this animation can help to arouse public attention on some issues such as global warming, polar sea ice melting, ocean environment, and marine hazards.

Milky Way Analogue Isolated Disk Galaxy

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications,

University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois), Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Sriv (University of Illinois)

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center; University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Room: Mile High Prefunction
5:15 pm - 7:00 pm

Scientific Visualization & Data Analytics Showcase Reception

Visualization of Decision-Making Support (DMS) Information for Responding to a Typhoon-Induced Disaster

Authors: Dongmin Jang (Korea Institute of Science and Technology Information), Jin-Hee Yuk (Korea Institute of Science and Technology Information), Junghyun Park (Korea Institute of Science and Technology Information), Jooneun An (Korea Institute of Science and Technology Information), Minsu Joh (Korea Institute of Science and Technology Information)

A high-resolution coupled atmosphere, ocean, and inundation (flood) modeling and simulation system was developed for scientific, accurate, fast, and efficient forecasting of typhoon-induced disasters. This is based on the KISTI decision-making support system (K-DMSS). Our prediction system consists of a typhoon, surge/wave, and flooding prediction and analysis systems (TPAS, SPAS, and FPAS). In this research, we simulated Typhoon 'CHABA' (1618), which was ranked third among the most intense tropical cyclones and was the most powerful typhoon in the Republic of Korea (South Korea) in 2016. The CHABA-induced storm surge and inundation were simulated using our prediction and analysis system. To understand intuitively the changes of physical phenomena and damage caused by the typhoon, numerical data sets produced by the prediction and analysis systems were visualized using VAPOR (Visualization and Analysis Platform for Ocean, atmosphere, and solar Researchers, and which is a part of the K-DMSS) as one of the visualization systems.

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Visualizations of a High-Resolution Global-Regional Nested, Ice-Sea-Wave Coupled Ocean Model System

Authors: Kangyou Zhong (Sun Yat-Sen University), Danya Xu (Sun Yat-Sen University), Changsheng Chen (University of Massachusetts, Dartmouth; Sun Yat-Sen University), Yutong Lu (Sun Yat-Sen University), Wenjie Dong (Sun Yat-Sen University), Jiang Li (Sun Yat-Sen University)

A multi-scale, global-regional nested ocean modeling system based on the unstructured grid Finite Volume Community Ocean Model (FVCOM) has been deployed on the Tianhe-2 supercomputer providing 24/7 marine forecasting since September 2016. The modeling system is part of the Sun Yat-Sen University Community Integrated Model (SYCIM) project for developing a new generation Earth System Model to explore the physical mechanisms of climate change. With a horizontal resolution up to ~17 m, this high-resolution modeling system can properly resolve the complex dynamical interactions of estuary, near shore coast, continental shelf, and deep ocean basins. The submitted animation shows the modeled global surface waves distribution pattern and propagation in the world ocean basins and in the South China Sea. Additionally, the variations of the global sea surface temperature, Arctic and Antarctic sea ice thickness and extension simulated by SYCIM is also presented. The animation can not only be used to visualize the big data for scientific research but also as a good example to demonstrate dynamical variations of the ocean to non-scientific communities. We hope this animation can help to arouse public attention on some issues such as global warming, polar sea ice melting, ocean environment, and marine hazards.

Milky Way Analogue Isolated Disk Galaxy

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois), Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Srivier (University of Illinois)

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center; University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Wednesday, November 15th

Room: Mile High Prefunction
8:30 am - 5:00 pm

Scientific Visualization & Data Analytics Showcase Posters

Visualization of Decision-Making Support (DMS) Information for Responding to a Typhoon-Induced Disaster

Authors: Dongmin Jang (Korea Institute of Science and Technology Information), Jin-Hee Yuk (Korea Institute of Science and Technology Information), Junghyun Park (Korea Institute of Science and Technology Information), Jooneun An (Korea Institute of Science and Technology Information), Minsu Joh (Korea Institute of Science and Technology Information)

A high-resolution coupled atmosphere, ocean, and inundation (flood) modeling and simulation system was developed for scientific, accurate, fast, and efficient forecasting of typhoon-induced disasters. This is based on the KISTI decision-making support system (K-DMSS). Our prediction system consists of a typhoon, surge/wave, and flooding prediction and analysis systems (TPAS, SPAS, and FPAS). In this research, we simulated Typhoon 'CHABA' (1618), which was ranked third among the most intense tropical cyclones and was the most powerful typhoon in the Republic of Korea (South Korea) in 2016. The CHABA-induced storm surge and inundation were simulated using our prediction and analysis system. To understand intuitively the changes of physical phenomena and damage caused by the typhoon, numerical data sets produced by the prediction and analysis systems were visualized using

VAPOR (Visualization and Analysis Platform for Ocean, atmosphere, and solar Researchers, and which is a part of the K-DMSS) as one of the visualization systems.

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: *Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)*

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Visualizations of a High-Resolution Global-Regional Nested, Ice-Sea-Wave Coupled Ocean Model System

Authors: *Kangyou Zhong (Sun Yat-Sen University), Danya Xu (Sun Yat-Sen University), Changsheng Chen (University of Massachusetts, Dartmouth; Sun Yat-Sen University), Yutong Lu (Sun Yat-Sen University), Wenjie Dong (Sun Yat-Sen University), Jiang Li (Sun Yat-Sen University)*

A multi-scale, global-regional nested ocean modeling system based on the unstructured grid Finite Volume Community Ocean Model (FVCOM) has been deployed on the Tianhe-2 supercomputer providing 24/7 marine forecasting since September 2016. The modeling system is part of the Sun Yat-Sen University Community Integrated Model (SYCIM) project for developing a new generation Earth System Model to explore the physical mechanisms of climate change. With a horizontal resolution up to ~17 m, this high-resolution modeling system can properly resolve the complex dynamical interactions of estuary, near shore coast, continental shelf, and deep ocean basins. The submitted animation shows the modeled global surface waves distribution pattern and propagation in the world ocean basins and in the South China Sea. Additionally, the variations of the global sea surface temperature, Arctic and Antarctic sea ice thickness and extension simulated by SYCIM is also presented. The animation can not only be used to visualize the big data for scientific research but also as a good example to demonstrate dynamical variations of the ocean to non-scientific communities. We hope this animation can help to arouse public attention on some issues such as global warming, polar sea ice melting, ocean environment, and marine hazards.

Milky Way Analogue Isolated Disk Galaxy

Authors: *Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois), Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)*

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: *David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Sriver (University of Illinois)*

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center; University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Room: 210-212

10:30 am - 12:00 pm

Scientific Visualization & Data Analytics Showcase

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Milky Way Analogue Isolated Disk Galaxy

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois),

Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Srivir (University of Illinois)

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center; University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Thursday, November 16th

Room: Mile High Prefunction
8:30 am - 5:00 pm

Scientific Visualization & Data Analytics Showcase Posters

Visualization of Decision-Making Support (DMS) Information for Responding to a Typhoon-Induced Disaster

Authors: Dongmin Jang (Korea Institute of Science and Technology Information), Jin-Hee Yuk (Korea Institute of Science and Technology Information), Junghyun Park (Korea Institute of Science and Technology Information), Jooneun An (Korea Institute of Science and Technology Information), Minsu Joh (Korea Institute of Science and Technology Information)

A high-resolution coupled atmosphere, ocean, and inundation (flood) modeling and simulation system was developed for scientific, accurate, fast, and efficient forecasting of typhoon-induced disasters. This is based on the KISTI decision-making support system (K-DMSS). Our prediction system consists of a typhoon, surge/wave, and flooding prediction and analysis systems (TPAS, SPAS, and FPAS). In this research, we simulated Typhoon 'CHABA' (1618), which was ranked third among the most intense tropical cyclones and was the most powerful typhoon in the Republic of Korea (South Korea) in 2016. The CHABA-induced storm surge and inundation were simulated using our prediction and analysis system. To understand intuitively the changes of physical phenomena and damage caused by the typhoon, numerical data sets produced by the prediction and analysis systems were visualized using VAPOR (Visualization and Analysis Platform for Ocean, atmosphere, and solar Researchers, and which is a part of the K-DMSS) as one of the visualization systems.

Comprehensive Visualization of Large-Scale Simulation Data Linked to Respiratory Flow Computations on HPC Systems

Authors: Andreas Lintermann (RWTH Aachen University, Juelich Aachen Research Alliance), Sonja Habbinga (Forschungszentrum Juelich), Jens Henrik Goebbert (Forschungszentrum Juelich)

Conditioning large-scale simulation data for comprehensive visualizations to enhance intuitive understanding of complex physical phenomena is a challenging task. This is corroborated by the fact that the massive amount of data produced by such simulations exceeds the human horizon of perception. It is therefore essential to distill the key features of such data to derive at new knowledge on an abstract level.

Furthermore, presenting scientific data to a wide public audience, especially if the scientific content is of high societal interest, i.e., as it is the case for fine dust pollution, is not only difficult from a visualization but also from an information transfer point of view. Impressive visual and contextual presentation are hence key to an effective knowledge transfer of complicated scientific data and the involved methods to arrive at such data.

In this paper such an approach is presented for highly-dense simulation data stemming from HPC simulations of inspiratory flows in the human respiratory tract. The simulations are performed using a coupled lattice-Boltzmann/Lagrange method and aim at understanding the microscopic interactions of flow and particle dynamics in highly intricate anatomically correct geometries. As such, they deliver insights on the impact of particulate matter on the human body.

Visualizations of a High-Resolution Global-Regional Nested, Ice-Sea-Wave Coupled Ocean Model System

Authors: Kangyou Zhong (Sun Yat-Sen University), Danya Xu (Sun Yat-Sen University), Changsheng Chen (University of Massachusetts, Dartmouth; Sun Yat-Sen University), Yutong Lu (Sun Yat-Sen University), Wenjie Dong (Sun Yat-Sen University), Jiang Li (Sun Yat-Sen University)

A multi-scale, global-regional nested ocean modeling system based on the unstructured grid Finite Volume Community Ocean Model (FVCOM) has been deployed on the Tianhe-2 supercomputer providing 24/7 marine forecasting since September 2016. The modeling system is part of the Sun Yat-Sen University Community Integrated Model (SYCIM) project for developing a new generation Earth System Model to explore the physical mechanisms of climate change. With a horizontal resolution up to ~17 m, this high-resolution modeling system can properly resolve the complex dynamical interactions of estuary, near shore coast, continental shelf, and deep ocean basins. The submitted animation shows the modeled global surface waves distribution pattern and propagation in the world ocean basins and in the South China Sea. Additionally, the variations of the global sea surface temperature, Arctic and Antarctic sea ice thickness and extension simulated by SYCIM is also presented. The animation can not only be used to visualize the big data for scientific research but also as a good example to demonstrate dynamical variations of the ocean to non-scientific communities. We hope this animation can help to arouse public attention on some issues such as global warming, polar sea ice melting, ocean environment, and marine hazards.

Milky Way Analogue Isolated Disk Galaxy

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Nathan J. Goldbaum (National Center for Supercomputing Applications, University of Illinois), Mark R. Krumholz (Australian National University; University of California, Santa Cruz), John C. Forbes (Harvard University, Harvard-Smithsonian Center for Astrophysics)

This visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications shows the evolution of a simulated analogue for the Milky Way galaxy over the course of 50 million years.

Simulation and Visual Representation of Tropical Cyclone-Ocean Interactions

Authors: David Bock (National Center for Supercomputing Applications, University of Illinois), Hui Lui (University of Illinois), Ryan L. Srivier (University of Illinois)

The winds of a tropical cyclone (TC) can induce vigorous ocean vertical mixing bringing cold water to the ocean surface and injecting warm water down into the ocean interior. The result of such interactions can alter the ocean heat budget and transport having implications for large-scale circulations within the Earth system. To better understand and analyze these implications, we computationally simulated the effect of TC winds on the ocean. We present results from ocean model simulations forced with tropical cyclone winds that are extracted from a fully coupled Earth system model simulation. Results are compared with a control simulation without TC winds. Differences between the TC-forcing run and the control run can reveal the effect of TC wind mixing on the ocean. The unique spatial and temporal relationships between the simulation results present unique challenges to effective visual representation. Using a variety of different visualization techniques, a custom software system is used to design and develop several visualization sequences to help better understand and analyze the simulation results.

First Light in the Renaissance Simulation Visualization: Formation of the Very First Galaxies in the Universe

Authors: Donna J. Cox (National Center for Supercomputing Applications, University of Illinois), Robert M. Patterson (National Center for Supercomputing Applications, University of Illinois), Stuart A. Levy (National Center for Supercomputing Applications, University of Illinois), Jeffrey D. Carpenter (National Center for Supercomputing Applications, University of Illinois), AJ Christensen (National Center for Supercomputing Applications, University of Illinois), Kalina M. Borkiewicz (National Center for Supercomputing Applications, University of Illinois), Brian W. O'Shea (Michigan State University), John H. Wise (Georgia Institute of Technology), Hao Xu (University of California, San Diego), Michael L. Norman (San Diego Supercomputer Center; University of California, San Diego)

This two-part visualization by the Advanced Visualization Lab at the National Center for Supercomputing Applications starts shortly after the Big Bang, and shows the evolution of the first galaxies in the universe over the first 400 million years, in increments of about 4 million years. The second part of the visualization stops time at the 400 million year mark, and flies the viewer through the data, breaking down the different variables that are being visualized - filaments of dense gas, pockets of elevated temperature, metals, ionized gas, and ultraviolet light.

Visualizing Silicene Growth Through Island Migration and Coalescence

Authors: Joseph A. Insley (Argonne National Laboratory, Northern Illinois University), Mathew J. Cherukara (Argonne National Laboratory), Badri Narayanan (Argonne National Laboratory), Henry Chan (Argonne National Laboratory), Subramanian Sankaranarayanan (Argonne National Laboratory)

Massively parallel molecular dynamics simulations carried out on the Argonne Leadership Computing Facility's supercomputer, Mira, are providing insight into materials that are vital to the improved design and functionality of the next generation of electronic devices. Silicene has a number of desirable properties, which could make it ideal for use in such devices. These simulations identify the elementary steps involved in the formation and evolution of monolayers of silicene on an iridium substrate. In this work, we present the visualization of the various stages of silicene nucleation and growth identified in these studies.

Physical Signatures of Cancer Metastasis

Authors: Anne Dara Bowen (Texas Advanced Computing Center, University of Texas), Abdul N. Malmi-Kakkada (University of Texas), Ayat Mohammed (Texas Advanced Computing Center, University of Texas)

Metastasis is the development of secondary malignant growths at a distance from a primary site of cancer. Most of human deaths (90%) from cancer are due to metastasis. In order to study physical signatures of metastasis, detailed analyses of individual cell trajectories are performed. The compilation of animated and still visualizations selected for this movie summarize the key findings from the investigation using their model, and identify complex spatial and time dependent cell migration patterns.

Student Cluster Competition [\(back to top\)](#)

Monday, November 13th

Room: Booth 796 - Exhibit Floor

7:30 pm - 8:00 pm

Student Cluster Competition Kick-Off

Student Cluster Competition

The Student Cluster Competition (SCC) was developed in 2007 to immerse undergraduate and high school students in high performance computing. Student teams design and build small clusters with hardware and software vendor partners, learn designated scientific applications, apply optimization techniques for their chosen architectures, and compete in a non-stop, 48-hour challenge. During the competition, each team races to complete a real-world scientific workload while impressing interview judges and conference attendees like you with their HPC knowledge.

For the competition, teams of up to six students partner with vendors to design and build cutting-edge clusters from commercially-available components, and teams work with application experts to prepare for the competition. With a catch! - Clusters must not exceed a 3000-watt power limit.

The winning team will have the highest aggregate score on these three criteria:

- Throughput, correctness, and understanding of real application workload
- Demonstrated understanding and performance of benchmark and mystery applications
- HPC presentation, conference engagement, and interviews. Recognition will be given for the highest HPL result.

The SCC committee would like to thank those who have given their money, time and hardware to support the SCC.

- Platinum Level: Microsoft, Geist
- Gold Level: SAIC
- Silver Level: ARM, Google
- Bronze Level: NVIDIA, Schlumberger

Room: Booth 796 - Exhibit Floor

8:00 pm - 9:00 pm

Student Cluster Competition

Student Cluster Competition

The Student Cluster Competition (SCC) was developed in 2007 to immerse undergraduate and high school students in high performance computing. Student teams design and build small clusters with hardware and software vendor partners, learn designated scientific applications, apply optimization techniques for their chosen architectures, and compete in a non-stop, 48-hour challenge. During the competition, each team races to complete a real-world scientific workload while impressing interview judges and conference attendees like you with their HPC knowledge.

For the competition, teams of up to six students partner with vendors to design and build cutting-edge clusters from commercially-available components, and teams work with application experts to prepare for the competition. With a catch! - Clusters must not exceed a 3000-watt power limit.

The winning team will have the highest aggregate score on these three criteria:

- Throughput, correctness, and understanding of real application workload
- Demonstrated understanding and performance of benchmark and mystery applications
- HPC presentation, conference engagement, and interviews. Recognition will be given for the highest HPL result.

The SCC committee would like to thank those who have given their money, time and hardware to support the SCC.

- Platinum Level: Microsoft, Geist
- Gold Level: SAIC
- Silver Level: ARM, Google
- Bronze Level: NVIDIA, Schlumberger

Tuesday, November 14th

Room: Booth 796 - Exhibit Floor
10:00 am - 6:00 pm

Student Cluster Competition

Student Cluster Competition

The Student Cluster Competition (SCC) was developed in 2007 to immerse undergraduate and high school students in high performance computing. Student teams design and build small clusters with hardware and software vendor partners, learn designated scientific applications, apply optimization techniques for their chosen architectures, and compete in a non-stop, 48-hour challenge. During the competition, each team races to complete a real-world scientific workload while impressing interview judges and conference attendees like you with their HPC knowledge.

For the competition, teams of up to six students partner with vendors to design and build cutting-edge clusters from commercially-available components, and teams work with application experts to prepare for the competition. With a catch! - Clusters must not exceed a 3000-watt power limit.

The winning team will have the highest aggregate score on these three criteria:

- Throughput, correctness, and understanding of real application workload
- Demonstrated understanding and performance of benchmark and mystery applications
- HPC presentation, conference engagement, and interviews. Recognition will be given for the highest HPL result.

The SCC committee would like to thank those who have given their money, time and hardware to support the SCC.

- Platinum Level: Microsoft, Geist
- Gold Level: SAIC
- Silver Level: ARM, Google
- Bronze Level: NVIDIA, Schlumberger

Wednesday, November 15th

Room: Booth 796 - Exhibit Floor
10:00 am - 4:30 pm

Student Cluster Competition

Student Cluster Competition

The Student Cluster Competition (SCC) was developed in 2007 to immerse undergraduate and high school students in high performance computing. Student teams design and build small clusters with hardware and software vendor partners, learn designated scientific applications, apply optimization techniques for their chosen architectures, and compete in a non-stop, 48-hour challenge. During the competition, each team races to complete a real-world scientific workload while impressing interview judges

and conference attendees like you with their HPC knowledge.

For the competition, teams of up to six students partner with vendors to design and build cutting-edge clusters from commercially-available components, and teams work with application experts to prepare for the competition. With a catch! - Clusters must not exceed a 3000-watt power limit.

The winning team will have the highest aggregate score on these three criteria:

- Throughput, correctness, and understanding of real application workload
- Demonstrated understanding and performance of benchmark and mystery applications
- HPC presentation, conference engagement, and interviews. Recognition will be given for the highest HPL result.

The SCC committee would like to thank those who have given their money, time and hardware to support the SCC.

- Platinum Level: Microsoft, Geist
- Gold Level: SAIC
- Silver Level: ARM, Google
- Bronze Level: NVIDIA, Schlumberger

Room: Booth 796 - Exhibit Floor

4:30 pm - 5:00 pm

Student Cluster Competition Finale

Student Cluster Competition

The Student Cluster Competition (SCC) was developed in 2007 to immerse undergraduate and high school students in high performance computing. Student teams design and build small clusters with hardware and software vendor partners, learn designated scientific applications, apply optimization techniques for their chosen architectures, and compete in a non-stop, 48-hour challenge. During the competition, each team races to complete a real-world scientific workload while impressing interview judges and conference attendees like you with their HPC knowledge.

For the competition, teams of up to six students partner with vendors to design and build cutting-edge clusters from commercially-available components, and teams work with application experts to prepare for the competition. With a catch! - Clusters must not exceed a 3000-watt power limit.

The winning team will have the highest aggregate score on these three criteria:

- Throughput, correctness, and understanding of real application workload
- Demonstrated understanding and performance of benchmark and mystery applications
- HPC presentation, conference engagement, and interviews. Recognition will be given for the highest HPL result.

The SCC committee would like to thank those who have given their money, time and hardware to support the SCC.

- Platinum Level: Microsoft, Geist
- Gold Level: SAIC
- Silver Level: ARM, Google
- Bronze Level: NVIDIA, Schlumberger

Student Job/Opportunity Fair [\(back to top\)](#)

Wednesday, November 15th

Room: 702-704-706

10:00 am - 3:00 pm

Student/Postdoc Job Fair

Student/Postdoc Job Fair

All Students attending SC are invited to attend the Student/Postdoc Job Fair. Hiring organizations will be present from 10am-3pm, and students may stop by at any time.

Students@SC [\(back to top\)](#)

Sunday, November 12th

Room: 703

9:00 am - 10:00 am

Welcome to Students@SC

Students@SC17 Welcome and Opening Session *Bruce Loftis (Independent)*

The Students@SC program works to attract new and diverse groups of students to the conference each year with the goal of sparking new conversations, new connections, and new ideas. However, we understand that it can be difficult to navigate the conference and experience all that SC17 has to offer. This session will help to orient students to what the conference has to offer and how best to make the most of their experience.

Room: 703

10:30 am - 12:00 pm

Student Résumé Workshop

Student Résumé Workshop *Scott Michael (Indiana University)*

Students will have an opportunity to consult with professionals from industry and academia to get feedback on their résumés and curriculum vitae in advance of the student job fair.

Room: 703

1:30 pm - 3:00 pm

Future Trends in HPC

Future Trends in HPC *David Hancock (Indiana University), Rangan Sukumar (Cray Inc)*

Three talks on the future of computing in HPC. Cloud computing, Data analytics, and Accelerators.

Room: 703

3:30 pm - 5:00 pm

Professional Persona Workshop

Building Your Professional Persona *Patty Lopez (Intel Corporation)*

This session addresses the dos and don'ts of building a professional image. Topics will include your elevator pitch, in-person communications and web presence (personal pages and social media), social media postings, dissemination of technical contributions, and professionalism.

Monday, November 13th

Room: 703

8:30 am - 10:00 am

Student Résumé Workshop

Student Résumé Workshop *Scott Michael (Indiana University)*

Students will have an opportunity to consult with professionals from industry and academia to get feedback on their résumés and curriculum vitae in advance of the student job fair.

Room: 703

10:30 am - 12:00 pm

Career Panel

Careers in HPC *Elizabeth Jessup (University of Colorado, Boulder), Verónica Vergara (Oak Ridge National Laboratory)*

This panel discussion will highlight a variety of career options and career paths that are open to students. Panelists will share their experiences in developing their careers in several different areas of HPC.

Room: 703

1:30 pm - 3:00 pm

Professional Persona Workshop

Building Your Professional Persona *Patty Lopez (Intel Corporation)*

This session addresses the dos and don'ts of building a professional image. Topics will include your elevator pitch, in-person communications and web presence (personal pages and social media), social media postings, dissemination of technical contributions, and professionalism.

Room: 701

2:00 pm - 3:00 pm

Experiencing HPC for Undergraduates: Opening Reception

Experiencing HPC for Undergraduates Opening Reception

This session is a "meet and greet" for students in the HPC for Undergraduate program and the HPC for Undergraduates Chair.

Room: 701

3:00 pm - 5:00 pm

Experiencing HPC for Undergraduates: Orientation

Experiencing HPC for Undergraduates Orientation *Erik Saule (University of North Carolina, Charlotte)*

This session will provide an introduction to HPC for the participants in the HPC for Undergraduates Program. Topics will include an introduction to MPI, shared memory programming, domain decomposition and the typical structure of scientific programming.

Room: Four Seasons Ballroom

3:30 pm - 5:00 pm

Mentor-Protégé

Mentor-Protégé

The Mentor-Protégé program connects newcomers at SC with experienced conference attendees. This event serves as an opportunity for mentor-protégé pairs to meet each other and others. Matches are made well before the conference. Attendance is by ticket only.

Tuesday, November 14th

Room: 701

10:30 am - 12:00 pm

Experiencing HPC for Undergraduates: Introduction to HPC Research

Experiencing HPC for Undergraduates: Introduction to HPC Research *Manish Parashar (Rutgers University), John Shalf (Lawrence Berkeley National Laboratory), Martin Schulz (Technical University Munich), Chris Johnson (University of Utah)*

A panel of leading practitioners in HPC will introduce the various aspects of HPC, including architecture, applications, programming models and tools.

The speakers are: *Programming : Manish Parashar, Rutgers, the State University of New Jersey *Architecture : John Shalf, Lawrence Berkeley National Laboratory *Tools : Martin Schulz, Technical University of Munich *Applications : Chris Johnson, University of Utah

Wednesday, November 15th

Room: 702-704-706

10:00 am - 3:00 pm

Student/Postdoc Job Fair

Student/Postdoc Job Fair

All Students attending SC are invited to attend the Student/Postdoc Job Fair. Hiring organizations will be present from 10am-3pm, and students may stop by at any time.

Room: 701

10:30 am - 12:00 pm

Experiencing HPC for Undergraduates: Graduate Student Perspective

Experiencing HPC for Undergraduates: Graduate Student Perspective Danilo Carastan-Santos (*Federal University of ABC, Santo André, Brazil*), Amir Gholami (*University of California, Berkeley*), Amalee Wilson (*University of Alabama, Birmingham*), Yasodhadevi Nachimuthu (*Portland State University*), Kellon Belfon (*Stony Brook University*), Klaudius Scheufele (*University of Stuttgart*)

This session will be held as a panel discussion. Current graduate students, some of whom are candidates for the Best Student Paper Award in the Technical Papers program at SC17, will discuss their experiences in being a graduate student in an HPC discipline. They will also talk about the process of writing their award-nominated papers.

Thursday, November 16th

Room: 701

10:30 am - 12:00 pm

Experiencing HPC for Undergraduates: Careers in HPC

Experiencing HPC for Undergraduates: Careers in HPC Steve Poole (*Open Source Software Solutions*), Swann Perarneau (*Argonne National Laboratory*), Daniel Katz (*National Center for Supercomputing Applications, University of Illinois*), Dorian Arnold (*Emory University*), Sunita Chandrasekaran (*University of Delaware*), Michael Wolfe (*Nvidia Corporation*)

This panel will feature a number of distinguished members of the HPC research community discussing their varied career paths. The panel includes representatives from industry, government labs and universities. The session will include ample time for questions from the audience.

Test of Time [\(back to top\)](#)

Tuesday, November 14th

Room: Mile High Ballroom
1:30 pm - 3:00 pm

Test of Time Award Special Lecture

Why Iteration Space Tiling?
Michael Wolfe (Portland Group)

Tiling subdivides the multidimensional iteration space of a nested loop into blocks or tiles. Tiles become a natural candidate as the unit of work for parallel task scheduling, where scheduling and synchronization is done at tile granularity, reducing overhead at some loss of potential parallelism. Tiling allows separation of concerns between local optimizations (vectorization, cache footprint) and global optimization (parallelization, scheduling for locality). Tiling is well-known and has been included in many compilers and code transformation systems. The talk will explore the basic contribution of the SC1989 paper to the current state of iteration space tiling.

Tutorial [\(back to top\)](#)

Sunday, November 12th

Room: 201

8:30 am - 12:00 pm

Performance Tuning of Scientific Codes with the Roofline Model

Presenters: *Tuomas S. Koskela (Lawrence Berkeley National Laboratory), Aleksandar Ilic (Institute of Systems and Computer Engineering, Portugal), Zakhar A. Matveev (Intel Corporation), Samuel W. Williams (Lawrence Berkeley National Laboratory), Philippe Thierry (Intel Corporation)*

The Roofline performance model offers insightful and intuitive method for extracting key characteristics of HPC applications and comparing them against peak upper-bounds of modern platforms and micro-architectures. Its capability to abstract the complexity of modern non-uniform memory hierarchies and identify the most profitable optimization techniques had made Roofline-based analysis increasingly popular in HPC community. Although different flavors of the roofline model have been developed to deal with various definitions of memory data traffic (by some of the contributors and presenters of this tutorial), there is still need for a more systematic methodology when applying them to analyze the efficiency of applications in multi/many-core systems. The tutorial aims to bridge this gap by exposing the fundamental aspects behind different Roofline modeling principles and providing several practical use-case scenarios to show their usability for application optimization. This tutorial presents a unique and solid combination of novel methodologies applied to optimize a representative set of open science use cases, while highly practice-oriented topics and labs are given by the lead methodology researchers or main designer of Intel's Roofline automation tools. The tutorial presenters have a long experience working with the Roofline model and have presented several Roofline-based tutorials.

Room: 210-212

8:30 am - 12:00 pm

Migratable Objects and Task-Based Parallel Programming with Charm++

Presenters: *Laxmikant Kale (University of Illinois), Michael Robson (University of Illinois), Ronak Buch (University of Illinois)*

Modern supercomputers present several challenges to effectively program parallel applications: exposing concurrency, optimizing data movement, controlling load imbalance, addressing heterogeneity, handling variations in application behavior, tolerating system failures, etc. By leveraging Charm++, application developers have been able to successfully address these challenges and efficiently run their code on large supercomputers. The foundational concepts underlying Charm++ are overdecomposition, asynchrony, migratability, and adaptivity. A Charm++ program specifies collections of interacting objects, which are assigned to processors dynamically by the runtime system.

Charm++ provides an asynchronous, message-driven, task-based programming model with migratable objects and an adaptive runtime system that controls execution. It automates communication overlap, load balance, fault tolerance, checkpoints for split-execution, power management, MPI interoperation, and promotes modularity.

This tutorial will begin with an explanation of the foundational concepts, followed by a detailed exposition of the syntax and specific capabilities, with several simple example programs, as well as real-world application case studies. We will illustrate how users can write Charm++ programs in C++ which can scale across a range of supercomputers and scales.

Room: 301

8:30 am - 12:00 pm

Essential HPC Finance Practice: Total Cost of Ownership (TCO), Internal Funding, and Cost-Recovery Models

Presenters: *Andrew Jones (Numerical Algorithms Group), Owen G. M. Thomas (Red Oak Consulting)*

The tutorial provides an impartial, practical, non-sales focused guide to the financial aspects of HPC facilities and service. It presents a rare and invaluable opportunity for HPC managers, practitioners, and stakeholders to learn more about calculating and using TCO models; along with the pros and cons of different internal cost recovery and funding models.

Well-managed TCO, return on investment and cost recovery models can be hugely beneficial to HPC managers and operators by demonstrating the value of HPC to the organization, driving the continuation and growth of HPC investment. They can also help uncover practical improvements to deliver better services to users.

Attendees will benefit from exploration of the main issues, pros and cons of differing approaches, practical tips, hard-earned experience and potential pitfalls. After the tutorial, attendees will be in a stronger position to calculate and use TCO within their

organizations, and to design and use internal cost-recovery models.

The tutorial is based on experience across a diverse set of real world cases in various countries, in both private and public sectors, with projects of all sizes and shapes. The material will be updated from the SC16 base, with increased consideration of cloud vs in-house HPC.

Room: 401

8:30 am - 12:00 pm

Compression of Scientific Data

Presenters: *Franck Cappello (Argonne National Laboratory), Peter Lindstrom (Lawrence Livermore National Laboratory)*

Large-scale numerical simulations and experiments are generating very large datasets that are difficult to analyze, store and transfer. This problem will be exacerbated for future generations of systems. Data reduction becomes a necessity in order to reduce as much as possible the time lost in data transfer and storage. Lossless and lossy data compression are attractive and efficient techniques to significantly reduce data sets while being rather agnostic to the application. This tutorial will review the state of the art in lossless and lossy compression of scientific data sets, discuss in detail two lossy compressors (SZ and ZFP) and introduce compression error assessment metrics. The tutorial will also cover the characterization of data sets with respect to compression and introduce Z-checker, a tool to assess compression error. More specifically the tutorial will introduce motivating examples as well as basic compression techniques, cover the role of Shannon Entropy, the different types of advanced data transformation, prediction and quantization techniques, as well as some of the more popular coding techniques. The tutorial will use examples of real world compressors (GZIP, JPEG, FPZIP, SZ, ZFP, etc.) and datasets coming from simulations and instruments to illustrate the different compression techniques and their performance.

Room: 203

8:30 am - 5:00 pm

Cloud Computing for Science and Engineering

Presenters: *Ian Foster (Argonne National Laboratory, University of Chicago), Dennis Gannon (Indiana University), Vani Mandava (Microsoft)*

This tutorial introduces the working scientist, engineer or student to cloud computing. It surveys the technology that underpins the cloud, new approaches to technical problems enabled by the cloud, and the concepts required to integrate cloud services into scientific work. It covers managing data in the cloud, and how to program data services; computing in the cloud, from deploying single virtual machines or containers to supporting basic interactive science experiments to gathering clusters of machines to do data analytics; using the cloud as a platform for automating analysis procedures, machine learning, and analyzing streaming data; building your own cloud with open source software; and cloud security. The tutorial works best if you have Python and Jupyter installed on your laptop, but it can be enjoyed otherwise.

The tutorial is supported by a book, *Cloud Computing for Science and Engineering*, to be published by MIT Press in September 2017. The text of the book is also available online as a reference at <https://Coud4SciEng.org>, along with numerous Jupyter notebooks and other code samples. This site is the online resource for the book. It contains the Jupyter notebooks that are described in the book as well as additional material including lecture slides.

Room: 205

8:30 am - 5:00 pm

Deep Learning for Science in the Cloud

Presenters: *Cassie Kozyrkov (Google), Kevin D. Kissell (Google)*

This tutorial is designed for researchers, computer scientists, and software engineers who have a minimal background in machine learning and data science, but have an interest in applying machine learning techniques to scientific research. It will cover the basic principles and guide the students through a series of hands-on exercises, building, training, and testing machine learning models on the Google Cloud Platform, with a particular emphasis on scientific data sets.

Room: 207

8:30 am - 5:00 pm

Parallel Computing 101

Presenters: *Quentin F. Stout (University of Michigan), Christiane Jablonowski (University of Michigan)*

This tutorial provides a comprehensive overview of parallel computing, emphasizing those aspects most relevant to the user. It is

suitable for new users, managers, students and anyone seeking an overview of parallel computing. It discusses software and hardware/software interaction, with an emphasis on standards, portability, and systems that are widely available.

The tutorial surveys basic parallel computing concepts, using examples selected from multiple engineering, scientific, and data analysis problems. These examples illustrate using MPI on distributed memory systems; OpenMP on shared memory systems; MPI+OpenMP on hybrid systems; CUDA and compiler directives on GPUs; and Hadoop and Spark on big data. It discusses numerous parallelization and load balancing approaches, and software engineering and performance improvement aspects, including the use of state-of-the-art tools.

The tutorial helps attendees make intelligent decisions by covering the primary options that are available, explaining how they are used and what they are most suitable for. Extensive pointers to web-based resources are provided to facilitate follow-up studies.

Room: 302
8:30 am - 5:00 pm

Advanced Manycore Programming (KNL)

Presenters: *Jerome Vienne (University of Texas), Si Liu (University of Texas), Victor Eijkhout (University of Texas), Feng Chen (University of Texas)*

As processors continue to eke out more performance at the cost of complexity, an understanding of the underlying hardware is essential to developing code that runs well on new platforms such as the KNL. To take advantage of these features, application development now requires the consideration of at least three different levels of parallelism (MPI, threading, SIMD), proper task/thread placement, and the allocation of limited resources such as high bandwidth memory.

This tutorial is designed for experienced programmers familiar with OpenMP who wish to familiarize themselves with Intel's next generation manycore processor, the 2nd Generation Intel Xeon Phi "Knights Landing" (KNL). We will start by discussing the evolution of manycore processing and provide an overview of the KNL hardware and its various memory modes. Then, we will briefly show the use of reports and directives to improve vectorization and the implementation of proper memory access. We will next focus on new Intel VTune Amplifier XE capabilities that allow for in-depth memory access analysis and hybrid code profiling, as well as Intel Advisor capabilities for vectorization analysis. Hands-on exercises will be executed on the KNL-upgraded Stampede system at the Texas Advanced Computing Center (TACC).

Room: 303
8:30 am - 5:00 pm

OpenMP Common Core: A "Hands-On" Exploration

Presenters: *Timothy Mattson (Intel Corporation), Alice Koniges (Lawrence Berkeley National Laboratory), Yun (Helen) He (Lawrence Berkeley National Laboratory), Barbara Chapman (Stony Brook University)*

OpenMP is the de facto standard for writing parallel applications for shared memory computers. Born 20 years ago in 1997, it runs on just about every shared memory platform on the market. It's also very complicated. We created OpenMP to be the "easy API" for the general application programmer. With a specification running to over 300 pages, OpenMP has grown into an intimidating API viewed by many as for "experts only".

Most OpenMP programmers, however, use around 16 items from the specification. We call these 16 constructs the "OpenMP Common Core". By focusing on the common core, we make OpenMP what it was always meant to be; an easy API for parallel application programmers.

In this hands-on tutorial, we explore the common core of OpenMP. We utilize active learning through a carefully selected set of exercises, so students will master the common core and learn to apply it to their own problems. Students will use their own laptops (with Windows, Linux, or OS/X) to access remote systems that support OpenMP (a remote SMP server). Alternatively, students can load an OpenMP compiler onto their laptops before the tutorial. Information about OpenMP compilers is available at www.openmp.org.

Room: 402
8:30 am - 5:00 pm

Node-Level Performance Engineering

Presenters: *Georg Hager (University of Erlangen-Nuremberg), Gerhard Wellein (University of Erlangen-Nuremberg)*

The advent of multi- and manycore chips has led to a further opening of the gap between peak and application performance for many scientific codes. This trend is accelerating as we move from petascale to exascale. Paradoxically, bad node-level performance helps to "efficiently" scale to massive parallelism, but at the price of increased overall time to solution. If the user cares about time to solution on any scale, optimal performance on the node level is often the key factor. We convey the architectural

features of current processor chips, multiprocessor nodes, and accelerators, as far as they are relevant for the practitioner. Peculiarities like SIMD vectorization, shared vs. separate caches, bandwidth bottlenecks, and ccNUMA characteristics are introduced, and the influence of system topology and affinity on the performance of typical parallel programming constructs is demonstrated. Performance engineering and performance patterns are suggested as powerful tools that help the user understand the bottlenecks at hand and to assess the impact of possible code optimizations. A cornerstone of these concepts is the roofline model, which is described in detail, including useful case studies, limits of its applicability, and possible refinements.

Room: 403

8:30 am - 5:00 pm

Large Scale Visualization with ParaView

Presenters: *Kenneth Moreland (Sandia National Laboratories), W. Alan Scott (Sandia National Laboratories), David E. DeMarle (Kitware Inc), Joe Insley (Argonne National Laboratory), John Patchett (Los Alamos National Laboratory), Jon Woodring (Los Alamos National Laboratory)*

ParaView is a powerful open-source turnkey application for analyzing and visualizing large data sets in parallel. Designed to be configurable, extendible, and scalable, ParaView is built upon the Visualization Toolkit (VTK) to allow rapid deployment of visualization components. This tutorial presents the architecture of ParaView and the fundamentals of parallel visualization. Attendees will learn the basics of using ParaView for scientific visualization with hands-on lessons. The tutorial features detailed guidance in visualizing the massive simulations run on today's supercomputers, an introduction to scripting ParaView and in using ParaView's Catalyst extensions for in situ workflows. Attendees should bring laptops to install ParaView and follow along with the demonstrations.

Room: 404

8:30 am - 5:00 pm

Advanced OpenMP: Performance and 4.5 Features

Presenters: *Christian Terboven (RWTH Aachen University), Michael Klemm (Intel Corporation), Ruud van der Pas (Oracle), Eric J. Stotzer (Texas Instruments), Bronis R. de Supinski (Lawrence Livermore National Laboratory)*

With the increasing prevalence of multicore processors, shared-memory programming models are essential. OpenMP is a popular, portable, widely supported and easy-to-use shared-memory model. Developers usually find OpenMP easy to learn. However, they are often disappointed with the performance and scalability of the resulting code. This disappointment stems not from shortcomings of OpenMP but rather with the lack of depth with which it is employed. Our "Advanced OpenMP Programming" tutorial addresses this critical need by exploring the implications of possible OpenMP parallelization strategies, both in terms of correctness and performance.

We assume attendees understand basic parallelization concepts and know the fundamentals of OpenMP. We focus on performance aspects, such as data and thread locality on NUMA architectures, false sharing, and exploitation of vector units. All topics are accompanied with extensive case studies and we discuss the corresponding language features in-depth. The second half of the tutorial is dedicated to the directives for attached compute accelerators and the corresponding optimization techniques, such as asynchronous data transfer and kernel execution. Throughout all topics we present the recent additions of OpenMP 4.5 and extensions that have been subsequently adopted by the OpenMP Language Committee.

Room: 405

8:30 am - 5:00 pm

Hands-On Practical Hybrid Parallel Application Performance Engineering

Presenters: *Christian Feld (Juelich Supercomputing Center), Markus Geimer (Juelich Supercomputing Center), Sameer Shende (University of Oregon), Ronny Tschüter (Technical University Dresden), Brian Wylie (Juelich Supercomputing Center)*

This tutorial presents state-of-the-art performance tools for leading-edge HPC systems founded on the community-developed Score-P instrumentation and measurement infrastructure, demonstrating how they can be used for performance engineering of effective scientific applications based on standard MPI, OpenMP, hybrid combination of both, and increasingly common usage of accelerators. Parallel performance tools from the Virtual Institute – High Productivity Supercomputing (VI-HPS) are introduced and featured in hands-on exercises with Scalasca, Vampir, and TAU. We present the complete workflow of performance engineering, including instrumentation, measurement (profiling and tracing, timing and PAPI hardware counters), data storage, analysis, and visualization. Emphasis is placed on how tools are used in combination for identifying performance problems and investigating optimization alternatives. Using their own notebook computers with a provided HPC Linux [<http://www.hpclinux.org>] OVA image containing all of the necessary tools (running within a virtual machine), participants will conduct exercises on the Stampede system at TACC where remote access to Intel Xeon Phi (KNL) nodes will be provided for the hands-on sessions. This will help to prepare participants to locate and diagnose performance bottlenecks in their own parallel programs.

Room: 406
8:30 am - 5:00 pm

Fault-Tolerance for High Performance and Distributed Computing: Theory and Practice

Presenters: *George Bosilca (University of Tennessee), Aurélien Bouteiller (University of Tennessee), Thomas Hérault (University of Tennessee), Yves Robert (ENS Lyon)*

Resilience is a critical issue for large-scale platforms, and this tutorial provides a comprehensive survey of fault-tolerant techniques for high-performance and distributed computing, with a fair balance between practice and theory. This tutorial is organized along four main topics:

- (i) An overview of failure types (software/hardware, transient/fail-stop), and typical probability distributions (Exponential, Weibull, Log-Normal);
- (ii) General-purpose techniques, which include several checkpoint and rollback recovery protocols, replication, prediction, and silent error detection;
- (iii) Application-specific techniques, such as ABFT for grid-based algorithms or fixed-point convergence for iterative applications, user-level checkpointing in memory;
- (iv) Practical deployment of fault tolerance techniques with User Level Fault Mitigation (a proposed MPI standard extension). Relevant examples based on ubiquitous computational solver routines will be protected with a mix of checkpoint-restart and advanced recovery techniques in a hands-on session.

The tutorial is open to all SC17 attendees who are interested in the current status and expected promise of fault-tolerant approaches for scientific applications. There are no audience prerequisites: background will be provided for all protocols and probabilistic models. However, basic knowledge of MPI will be helpful for the hands-on session.

Room: 407
8:30 am - 5:00 pm

Linear Algebra Libraries for High-Performance Computing: Scientific Computing with Multicore and Accelerators

Presenters: *Jakub Kurzak (University of Tennessee), Jack Dongarra (University of Tennessee), Michael Heroux (Sandia National Laboratories), James Demmel (University of California, Berkeley)*

Today, a desktop with a multicore processor and a GPU accelerator can already provide a TeraFlop/s of performance, while the performance of the high-end systems, based on multicores and accelerators, is already measured in tens of PetaFlop/s. This tremendous computational power can only be fully utilized with the appropriate software infrastructure, both at the low end (desktop, server) and at the high end (supercomputer installation). Most often a major part of the computational effort in scientific and engineering computing goes in solving linear algebra subproblems. After providing a historical overview of legacy software packages, the tutorial surveys the current state-of-the-art numerical libraries for solving problems in linear algebra, both dense and sparse. MAGMA, PETSc, Trilinos, SuperLu, Hypre, and other software packages are discussed. The tutorial also highlights recent advances in algorithms that minimize communication, i.e., data motion, which is much more expensive than arithmetic.

Room: 201
1:30 pm - 5:00 pm

Big Data Meets HPC: Exploiting HPC Technologies for Accelerating Big Data Processing and Management

Presenters: *Dhabaleswar K. (DK) Panda (Ohio State University), Xiaoyi Lu (Ohio State University)*

The convergence of HPC, Big Data, and Deep Learning is the next game-changing business opportunity. Apache Hadoop, Spark, gRPC/TensorFlow, and Memcached are becoming standard building blocks in handling Big Data oriented processing and mining. Recent studies have shown that default designs of these components can not efficiently leverage the features of modern HPC clusters, like Remote Direct Memory Access (RDMA) enabled high-performance interconnects, high-throughput parallel storage systems (e.g. Lustre), Non-Volatile Memory (NVM). In this tutorial, we will provide an in-depth overview of the architecture of Hadoop, Spark, gRPC/TensorFlow, and Memcached.

We will examine the challenges in re-designing networking and I/O components of these middleware with modern interconnects, protocols (such as InfiniBand, RoCE) and storage architectures. Using the publicly available software packages in the High-Performance Big Data project (HiBD, <http://hibd.cse.ohio-state.edu>), we will provide case studies of the new designs for several Hadoop/Spark/gRPC/TensorFlow/Memcached components and their associated benefits. Through these, we will also examine the interplay between high-performance interconnects, storage (HDD, NVM, and SSD), and multi-core platforms to achieve the best solutions for these components and applications on modern HPC clusters.

We also present in-depth case-studies with modern Deep Learning tools (e.g., Caffe, TensorFlow, DL4J, BigDL) with RDMA-

enabled Hadoop, Spark, and gRPC.

Room: 210-212
1:30 pm - 5:00 pm

Power-Aware High Performance Computing: Challenges and Opportunities for Application and System Developers

Presenters: *Dieter Kranzmueller (Ludwig Maximilian University of Munich), David Lowenthal (University of Arizona), Barry Rountree (Lawrence Livermore National Laboratory), Martin Schulz (Technical University Munich)*

Power and energy consumption are critical design factors for next generation large-scale HPC systems. The costs for energy are shifting budgets from investment to operating costs, and more and more often the size of systems is determined by its power needs.

As a consequence, the US Department of Energy (DOE) has set ambitious limits in terms of power consumption for their first exascale system, and many other funding agencies around the world have expressed similar goals. Yet, with today's HPC architectures and systems, these goals are still far out of reach: they will only be achievable through a complex set of mechanisms at all levels of hardware and software, including buildings and infrastructure; all of these aspects will additionally and directly impact the application developer. On future HPC systems, running a code efficiently (as opposed to purely with high performance) will be a major requirement for every user.

In this tutorial, we will discuss the challenges caused by power and energy constraints, review available approaches in hardware and software, highlight impacts on HPC center and infrastructure design and operations, and ultimately show how this change in paradigm from "cycle awareness" to "power awareness" will impact HPC users and their work.

Room: 301
1:30 pm - 5:00 pm

Extracting Value from HPC: Business Cases, Planning, and Investment

Presenters: *Andrew Jones (Numerical Algorithms Group), Owen G. M. Thomas (Red Oak Consulting)*

The tutorial provides an impartial, practical, non-sales focused guide to enabling an organization to deliver value by investing in HPC. A well-managed business case process can be hugely beneficial to HPC managers by demonstrating the value of HPC to the organization, driving the continuation and growth of HPC investment.

The tutorial is applicable to either a first investment or an upgrade of existing capability. It is most relevant to organizations with a clear purpose (e.g., industry) or those with a clear service mission (e.g., academic HPC facilities). It presents a rare opportunity for HPC managers, practitioners, and stakeholders to learn more about the steps involved in considering HPC investments, identifying the value, building a business case, engaging stakeholders, securing funding, requirements capture, market survey, strategic choices, and more. Attendees will benefit from exploration of the main issues, pros and cons of differing approaches, practical tips, hard-earned experience and potential pitfalls.

The tutorial is based on experience across a diverse set of real world cases in various countries, in both private and public sectors, with projects of all sizes and shapes.

Room: 401
1:30 pm - 5:00 pm

The Legion Programming Model

Presenters: *Alex Aiken (Stanford University)*

Legion is a programming model designed for portable, scalable and high performance applications on contemporary and future supercomputers. In this tutorial, participants will be introduced to Regent, a high-level programming language for the Legion programming model. The tutorial will be organized around teaching Regent "from the ground up", beginning with the motivation for task-based programming, simple examples and hands-on exercises, and working up to advanced programming concepts. The examples and exercises using Regent will allow participants to progress quickly from being introduced to the basics of task-based programming to writing parts of and thoroughly understanding a non-trivial, self-contained Regent application at the end of the tutorial. The cloud-based exercises will allow participants to run experiments on a cluster of multicore machines with GPUs. Tools for visualizing the functional behavior and performance of Regent applications will show participants the effect of different choices in mapping tasks and data onto complex machines. Overall, the tutorial will provide participants with an overview of task-based programming and performance tuning, as well as a starting point for developing their own task-based applications.

Monday, November 13th

Room: 201
8:30 am - 12:00 pm

Expressing Heterogeneous Parallelism in C++ with Intel Threading Building Blocks

Presenters: *Michael Voss (Intel Corporation), James Reinders (James Reinders Consulting LLC), Pablo Reble (Intel Corporation), Rafael Asenjo (University of Malaga)*

Due to energy constraints, high performance computing platforms are becoming increasingly heterogeneous, achieving greater performance per watt through the use of hardware that is tuned to specific computational kernels or application domains. It can be challenging for developers to match computations to accelerators, choose models for targeting those accelerators, and then coordinate the use of those accelerators in the context of their larger applications.

This tutorial starts with a survey of heterogeneous architectures and programming models, and discusses how to determine if a computation is suitable for a particular accelerator. Next, Intel® Threading Building Blocks (Intel® TBB), a widely used, portable C++ template library for parallel programming is introduced. Intel TBB was first developed in 2006 as a shared-memory parallel programming library, but has since been extended to allow developers to coordinate the use of accelerators such as integrated and discrete GPUs, attached devices such as Intel® Xeon Phi co-processors, and FPGAs in to their parallel C++ applications.

Attendees will be given a hands-on opportunity to use TBB to create parallel implementations of a sample code. They will first build a shared-memory implementation and then a heterogeneous implementation, running their samples on a mix of CPUs and accelerators.

Room: 203
8:30 am - 12:00 pm

HPC via HTTP: Portable, Scalable Computing Using App Containers and the Agave API

Presenters: *Rion Dooley (University of Texas), Kathy Traxler (Louisiana State University), Steven Brandt (Louisiana State University), John Fonner (Texas Advanced Computing Center, University of Texas)*

Supercomputing matters. So does user experience. Standing between the mainstream adoption of supercomputing and a new generation of users is the reality that the entry cost to using these systems, both in terms of dollars and in time spent learning the technology, has not significantly changed in the last 20 years. The rise of cloud computing only complicates the learning curve further. Over the last 6 years, the authors have been addressing this gap through the development of a Science-as-a-Service platform enabling users to go from their desktop, to their local data center, to the cloud, and back without sacrificing their existing tool chain or user experience.

In this tutorial, we combine best practices and lessons learned while on-boarding the last 70k new users to TACC's data center through the Agave Platform. Participants will walk through the process of scaling their application from a local environment to the Jetstream academic cloud and to an HPC at the Texas Advanced Computing Center. They will learn to use multiple container technologies to harmonize app execution between cloud and HPC resources, and they will learn to use modern APIs to orchestrate job execution, capture provenance information, and foster collaboration.

Room: 205
8:30 am - 12:00 pm

InfiniBand, Omni-Path, and High-Speed Ethernet for Dummies

Presenters: *Dhabaleswar K. (DK) Panda (Ohio State University), Hari Subramoni (Ohio State University)*

InfiniBand (IB), Omni-Path, and High-speed Ethernet (HSE) technologies are generating a lot of excitement toward building next generation High-End Computing (HEC) systems including clusters, datacenters, file systems, storage, cloud computing and Big Data (Hadoop, Spark, HBase, and Memcached) environments. RDMA over Converged Enhanced Ethernet (RoCE) technology is also emerging.

This tutorial will provide an overview of these emerging technologies, their offered architectural features, their current market standing, and their suitability for designing HEC systems. It will start with a brief overview of IB, Omni-Path, and HSE. An in-depth overview of the architectural features of IB, Omni-Path, and HSE (including iWARP and RoCE), their similarities and differences, and the associated protocols will be presented. Next, an overview of the OpenFabrics stack which encapsulates IB, HSE, and RoCE (v1/v2) in a unified manner will be presented. An overview of the libfabrics stack will also be provided. Hardware/software solutions and the market trends behind IB, Omni-Path, HSE, and RoCE will be highlighted. Finally, sample performance numbers of these technologies and protocols for different environments will be presented.

Room: 207
8:30 am - 12:00 pm

Getting Started with the Burst Buffer: Using DataWarp Technology

Presenters: *George S. Markomanolis (King Abdullah University of Science and Technology), Deborah Bard (Lawrence Berkeley National Laboratory)*

The long-awaited Burst Buffer technology is now being deployed on major supercomputing systems. In this tutorial, we will introduce the Burst Buffers deployed at the two latest supercomputers at NERSC (Cori) and KAUST (Shaheen II) based on the Cray DataWarp, and discuss in detail our experience with Burst Buffers from both a system and a user's perspective. Both KAUST and NERSC have been supporting BB projects for more than a year, and have developed a wealth of experience using these resources efficiently. For this tutorial, we combine the knowledge and experience of staff from both sites to provide attendees with an effective understanding of how to optimally use BB technology. We focus on optimizing massively parallel I/O for SSDs, a relatively new problem compared to well-established optimizations for parallel I/O to disk-based file systems. The tutorial will conclude with a live demonstration of a complex workflow executed on the Cray DataWarp, including simulation, analysis and visualization. We will also include user applications as case studies; we solicit user applications using this form: <https://goo.gl/M3dUv9>.

Room: 210-212
8:30 am - 12:00 pm

Mastering Tasking with OpenMP

Presenters: *Christian Terboven (RWTH Aachen University), Michael Klemm (Intel Corporation), Sergi Mateo Bellido (Barcelona Supercomputing Center), Xavier Teruel (Barcelona Supercomputing Center), Bronis R. de Supinski (Lawrence Livermore National Laboratory)*

With the increasing prevalence of multi-core processors, shared-memory programming models are essential. OpenMP is a popular, portable, widely supported and easy-to-use shared-memory model. Since version 3.0 released in 2008, OpenMP offers tasking to support the creation of composable parallel software blocks and the parallelization of irregular algorithms. Developers usually find OpenMP easy to learn. However, mastering the tasking concept of OpenMP requires a change in the way developers reason about the structure of their code and how to expose the parallelism of it. Our tutorial addresses this critical aspect by examining the tasking concept in detail and presenting patterns as solutions to many common problems.

We assume attendees understand basic parallelization concepts and know the fundamentals of OpenMP. We present the OpenMP tasking language features in detail and focus on performance aspects, such as introducing cut-off mechanisms, exploiting task dependencies, and preserving locality. All aspects are accompanied by extensive case studies. Throughout all topics we present the recent additions of OpenMP 4.5 and extensions that have been subsequently adopted by the OpenMP Language Committee.

Room: 301
8:30 am - 12:00 pm

HPC Acquisition and Commissioning

Presenters: *Andrew Jones (Numerical Algorithms Group), Owen G. M. Thomas (Red Oak Consulting), W. Terry Hewitt (WTH Associates Ltd)*

This tutorial will guide attendees through the process of purchasing and deploying a HPC system. It will cover the whole process from engaging with stakeholders in securing funding, requirements capture, market survey, specification of the tender/request for proposal documents, engaging with suppliers, evaluating proposals, and managing the installation.

Attendees will learn how to specify what they want, yet enable the suppliers to provide innovative solutions beyond their specification both in technology and in the price; how to demonstrate to stakeholders that the solution selected is best value for money; and the common risks, pitfalls and mitigation strategies essential to achieve an on-time and on-quality installation process. The tutorial has 3 parts: the procurement process including RFP; bid evaluation, benchmarks, clarification processes; contracting, project management, commissioning, acceptance testing.

The presenters have been involved in numerous major HPC procurements since 1990, as service managers, bidders to funding agencies, as customers and as advisors. The presenters are from the UK but the tutorial will be applicable to HPC procurements anywhere. The tutorial is based on experience across a diverse set of real world cases in various countries, in private and public sectors.

Room: 303
8:30 am - 12:00 pm

Better Scientific Software

Presenters: *David E. Bernholdt (Oak Ridge National Laboratory), Anshu Dubey (Argonne National Laboratory), Michael A. Heroux (Sandia National Laboratories, St. John's University), Alicia Klinvex (Sandia National Laboratories)*

The computational science and engineering (CSE) community is in the midst of an extremely challenging period created by the confluence of disruptive changes in computing architectures, demand for greater scientific reproducibility, and new opportunities for greatly improved simulation capabilities, especially through coupling physics and scales. Computer architecture changes require new software design and implementation strategies, including significant refactoring of existing code. Reproducibility demands require more rigor across the entire software endeavor. Code coupling requires aggregate team interactions including integration of software processes and practices. These challenges demand large investments in scientific software development and improved practices. Focusing on improved developer productivity and software sustainability is both urgent and essential.

This tutorial will provide information and hands-on experience with software practices, processes, and tools explicitly tailored for CSE. Goals are improving the productivity of those who develop CSE software and increasing the sustainability of software artifacts. We discuss practices that are relevant for projects of all sizes, with emphasis on small teams, and on aggregate teams composed of small teams. Topics include software licensing, effective models, tools, and processes for small teams (including agile workflow management), reproducibility, and scientific software testing (including automated testing and continuous integration).

Room: 406

8:30 am - 12:00 pm

In Situ Analysis and Visualization with SENSEI

Presenters: *E. Wes Bethel (Lawrence Berkeley National Laboratory), Andrew C. Bauer (Kitware Inc), Brad Whitlock (Intelligent Light), Matthew Wolf (Oak Ridge National Laboratory), Burlen Loring (Lawrence Berkeley National Laboratory), Silvio Rizzi (Argonne National Laboratory)*

A key trend facing extreme-scale computational science is the widening gap between computational and I/O rates, and the challenge that follows is how to best gain insight from simulation data when it is increasingly impractical to save it to persistent storage for subsequent visualization and analysis. One approach to this challenge is centered around the idea of in situ processing, where visualization and analysis processing is performed while data is still resident in memory.

The SENSEI community in situ data interface is an API that promotes code portability and reusability. From the simulation view, a developer can instrument their code with the SENSEI API and then make use of any number of in situ infrastructures. From the methods view, a developer can write an in situ method using the SENSEI API, then expect it to run in any number of in situ infrastructures.

This tutorial presents the fundamentals of in situ data analysis and visualization leveraging this generic interface. Attendees will learn the basics of in situ analysis and visualization while being exposed to advanced analysis such as time-dependent autocorrelation and interactive monitoring and steering. We demonstrate the infrastructure coupling using ADIOS, ParaView Catalyst, GLEAN and VisIt Libsim.

Room: 302

8:30 am - 5:00 pm

Scalable Parallel Programming Using OpenACC for Multicore, GPUs, and Manycore

Presenters: *Michael Wolfe (Nvidia Corporation)*

You will learn how modern supercomputers are organized and how highly parallel nodes will affect the design of your applications. Whether you use multicore, manycore, or GPU-accelerated nodes, the same basic concepts apply to your algorithm and data structure design: parallelism management and data management. For parallelism management, you will learn what kinds of parallelism can be productively exploited on different architectures. You will also learn how to write programs that use parallelism on different architectures effectively. Data management means managing the implicit or exposed memory hierarchy. On a multicore or manycore node, this includes improving cache utilization. On nodes with high bandwidth memory (manycore and GPU), data management includes minimizing traffic between the levels of the exposed memory hierarchy.

This tutorial will focus on the OpenACC API. Additional information about other programming models (OpenMP, CUDA, OpenCL, MPI) will be included where appropriate. The tutorial will include prepackaged hands on sessions. We plan to provide access to three types of systems: X86 multicore with attached GPUs, supporting both multicore and GPU programming; Power multicore with attached GPUs, similarly; Xeon Phi KNL manycore, supporting manycore programming. This allows attendees to experiment with and port between different systems.

Room: 402

8:30 am - 5:00 pm

How To Analyze the Performance of Parallel Codes 101

Presenters: *Martin Schulz (Technical University Munich), James Galarowicz (Krell Institute), Donald Maghrak (Krell Institute), David Montoya (Los Alamos National Laboratory), Jennifer Green (Los Alamos National Laboratory)*

Performance analysis is an essential step in the development of HPC codes. It will even gain in importance with the rising complexity of machines and applications that we are seeing today. Many tools exist to help with this analysis, but the user is too often left alone with interpreting the results.

We will provide a practical road map for the performance analysis of HPC codes and will provide users step-by-step advice on how to detect and optimize common performance problems, covering both on-node performance and communication optimization as well as issues on threaded and accelerator-based architectures. Throughout this tutorial, we will show live demos using Open|SpeedShop, a comprehensive and easy to use tool set. Additionally, at the end of each section we will provide hands-on exercises for attendees to try out the new techniques learned. All techniques will, however, apply broadly to any tool, and we will point out alternative tools where useful.

Room: 403
8:30 am - 5:00 pm

Parallel I/O in Practice

Presenters: *Rob Latham (Argonne National Laboratory), Rob Ross (Argonne National Laboratory), Brent Welch (Google), Katie Antypas (Lawrence Berkeley National Laboratory)*

I/O on HPC systems is a black art. This tutorial sheds light on the state-of-the-art in parallel I/O and provides the knowledge necessary for attendees to best leverage I/O resources available to them. We cover the entire I/O software stack including storage and parallel file systems at the lowest layer, the role of burst buffers (NVRAM), intermediate layers (such as MPI-IO), and high-level I/O libraries (such as HDF-5). We emphasize ways to use these interfaces that result in high performance and tools for generating insight into these stacks. Benchmarks on real systems are used throughout to show real-world results.

In the first third of the tutorial we cover the fundamentals of parallel I/O. We discuss storage technologies, both present and near-future. Our parallel file systems material covers general concepts and gives examples from Lustre, GPFS, PanFS, HDF5, Ceph, and Cloud Storage.

Our second third takes a more application-oriented focus. We examine the upper library layers of the I/O stack, covering MPI-IO, Parallel netCDF, and HDF5. We discuss interface features, show code examples, and describe how application calls translate into PFS operations.

Finally we discuss tools for capturing and understanding I/O behavior.

Room: 404
8:30 am - 5:00 pm

Advanced MPI Programming

Presenters: *Pavan Balaji (Argonne National Laboratory), William Gropp (University of Illinois), Torsten Hoefler (ETH Zurich), Rajeev Thakur (Argonne National Laboratory)*

The vast majority of production parallel scientific applications today use MPI and run successfully on the largest systems in the world. At the same time, the MPI standard itself is evolving to address the needs and challenges of future extreme-scale platforms as well as applications. This tutorial will cover several advanced features of MPI, including new MPI-3 features, that can help users program modern systems effectively. Using code examples based on scenarios found in real applications, we will cover several topics including efficient ways of doing 2D and 3D stencil computation, derived datatypes, one-sided communication, hybrid (MPI + shared memory) programming, topologies and topology mapping, and neighborhood and nonblocking collectives. Attendees will leave the tutorial with an understanding of how to use these advanced features of MPI and guidelines on how they might perform on different platforms and architectures.

Room: 405
8:30 am - 5:00 pm

Application Porting and Optimization on GPU-Accelerated POWER Architectures

Presenters: *Dirk Pleiter (Forschungszentrum Juelich), Andreas Herten (Forschungszentrum Juelich), Archana Ravindar (IBM), Jiri Kraus (Nvidia Corporation), Christoph Hagleitner (IBM), Bronson Messer (Oak Ridge National Laboratory)*

The POWER processor has re-emerged as a technology for supercomputer architectures. One major reason is the tight integration of processor and GPU accelerator through the new NVLink technology. Two major sites in the US, ORNL and LLNL, have already decided to have their pre-exascale systems being based on this new architecture.

This tutorial will give an opportunity to obtain in-depth knowledge and experience with GPU-accelerated POWER nodes. It focuses on porting applications to a single node and covers the topics architecture, compilers, performance analysis and tuning, and multi-

GPU programming. The tutorial will include an overview of the new NVLink-based node architectures, lectures on first-hand experience in porting to this architecture, and exercises using tools to focus on performance.

Room: 407
8:30 am - 5:00 pm

Managing HPC Software Complexity with Spack

Presenters: *Todd Gamblin (Lawrence Livermore National Laboratory), Gregory Becker (Lawrence Livermore National Laboratory), Massimiliano Culpo (Swiss Federal Institute of Technology in Lausanne), Gregory L. Lee (Lawrence Livermore National Laboratory), Matthew LeGendre (Lawrence Livermore National Laboratory), Mario Melara (National Energy Research Scientific Computing Center), Adam J. Stewart (Argonne National Laboratory, University of Illinois)*

HPC software is becoming increasingly complex. The largest applications require over 100 dependency libraries, and they combine interpreted languages like Python with lower-level C, C++, and Fortran libraries. To achieve good performance, developers must tune for multiple compilers, build options, and implementations of dependency libraries like MPI, BLAS, and LAPACK. The space of possible build configurations is combinatorial, and developers waste countless hours rebuilding software instead of producing new scientific results.

This tutorial focuses on Spack, our open-source tool for HPC package management. Spack uses concise package recipes written in Python to automate builds with arbitrary combinations of compilers, MPI versions, and dependency libraries. With Spack, users can install over 1,800 community-maintained packages without knowing how to build them; developers can efficiently automate builds of tens or hundreds of dependency libraries; and HPC center staff can deploy many versions of software for thousands of users. We provide a thorough introduction to Spack's capabilities: basic software installation, creating new packages, using Spack for HPC software development, and advanced multi-user deployment. We provide detailed use cases from our experiences in production HPC environments. Most sessions involve hands-on demonstrations, so attendees should bring a laptop computer.

Room: 201
1:30 pm - 5:00 pm

Interactive HPC: Using C++ and HPX Inside Jupyterhub to Write Performant Portable Parallel Code

Presenters: *Hartmut Kaiser (Louisiana State University), Steven R. Brandt (Louisiana State University), Kevin A. Huck (University of Oregon), Alice Koniges (Lawrence Berkeley National Laboratory), Bryce A. Lebach (Lawrence Berkeley National Laboratory)*

Interactive computing has taken on new importance in many branches of science, particularly in exploratory phases, where researchers are still playing with ideas and code. Jupyter notebooks have emerged as the dominant tool for interactive computing. With the advent of Cling, C++ can be used interactively in the Jupyter framework.

Additionally, the C++14/17 standard brings significant new parallel programming capability to the C++ language with the introduction of a unified interface for asynchronous programming using futures. This style of programming enables fine-grained constraint-based parallelism, and avoids many load-balancing issues. HPX built on C++14/17 extends it to distributed operations and increases its composability. By conforming to the standard, developers comfortable with the modern language can easily write scalable, portable parallel applications.

In this tutorial, we give a simple, Jupyter-based method to learn modern HPC-C++14/17 parallel programming, and then we show how to adapt C++ programs to modern massively parallel environments using HPX. Using hands-on examples and tool demonstrations, we show how HPX allows users to apply asynchronous concepts and active messaging techniques to gain performance and productivity in a variety of applications. The interactive environment provided by Jupyter has been effective for university students to quickly learn programming concepts.

Room: 203
1:30 pm - 5:00 pm

Programming Your GPU with OpenMP: A Hands-On Introduction

Presenters: *Timothy Mattson (Intel Corporation), Simon McIntosh-Smith (University of Bristol)*

OpenMP 1.0 was released in 1997 when the primary concern was symmetric multiprocessors. Over time, hardware has evolved with more complex memory hierarchies forcing us to embrace NUMA machines and work to understand how OpenMP fits in with distributed memory systems.

Current trends in hardware bring co-processors such as GPUs into the fold. A modern platform is often a heterogeneous system with CPU cores, GPU cores, and other specialized accelerators. OpenMP has responded by adding directives that map code and data onto a device. We refer to this family of directives as the target directives.

In this hands-on tutorial, we will explore these directives as they apply to programming GPUs. We assume people know the

fundamentals of OpenMP (perhaps by taking the OpenMP Common Core tutorial) so we can focus on deeply understanding the target directives and their use in complex application programs. We expect students to use their own laptops (with Windows, Linux, or OS/X) to connect to remote servers with GPUs, but the best option is for students to load an OpenMP compiler onto their laptops before the tutorial. Information about OpenMP compilers is available at www.openmp.org.

Room: 205
1:30 pm - 5:00 pm

InfiniBand, Omni-Path, and High-Speed Ethernet: Advanced Features, Challenges in Designing, HEC Systems and Usage

Presenters: *Dhabaleswar K. (DK) Panda (Ohio State University), Hari Subramoni (Ohio State University)*

As InfiniBand (IB), Omni-Path, and High-Speed Ethernet (HSE) technologies mature, they are being used to design and deploy various High-End Computing (HEC) systems: HPC clusters with GPGPUs and Xeon Phi supporting MPI, Storage and Parallel File Systems, Cloud Computing systems with SR-IOV Virtualization, Grid Computing systems, and Deep Learning systems. These systems are bringing new challenges in terms of performance, scalability, portability, reliability and network congestion. Many scientists, engineers, researchers, managers and system administrators are becoming interested in learning about these challenges, approaches being used to solve these challenges, and the associated impact on performance and scalability.

This tutorial will start with an overview of these systems. Advanced hardware and software features of IB, Omni-Path, HSE, and RoCE and their capabilities to address these challenges will be emphasized. Next, we will focus on Open Fabrics RDMA and Libfabric programming, and network management infrastructure and tools to effectively use these systems. A common set of challenges being faced while designing these systems will be presented. Finally, case studies focusing on domain-specific challenges in designing these systems (including the associated software stacks), their solutions and sample performance numbers will be presented.

Room: 207
1:30 pm - 5:00 pm

Container Computing for HPC and Scientific Workflows

Presenters: *Lisa Gerhardt (Lawrence Berkeley National Laboratory), Shane Canon (Lawrence Berkeley National Laboratory)*

Container computing is revolutionizing the way applications are developed and delivered. It offers opportunities that never existed before for significantly improving efficiency of scientific workflows and easily moving these workflows from the laptop to the supercomputer. Tools like Docker and Shifter enable a new paradigm for scientific and technical computing. However, to fully unlock its potential, users and administrators need to understand how to utilize these new approaches. This tutorial will introduce attendees to the basics of creating container images, explain best practices, and cover more advanced topics such as creating images to be run on HPC platforms using Shifter.

The tutorial will also explain how research scientists can utilize container-based computing to accelerate their research and how these tools can boost the impact of their research by enabling better reproducibility and sharing of their scientific process without compromising security. This is an extension of the highly successful tutorial presented at SC16 that was attended by more than 100 people. The 2016 tutorial was very highly rated with 2.8 / 3 stars for “would recommend” and 4.3 / 5 stars for overall quality (compared to an average of 4.2 for all tutorials).

Room: 210-212
1:30 pm - 5:00 pm

Automating Research Data Management with Globus

Presenters: *Vas Vasiliadis (University of Chicago), Steve Tuecke (University of Chicago)*

Globus is an established service that is widely used for managing research data in campus computing centers, national laboratories, and HPC facilities. While the interactive web browser interface can address simple file transfer and sharing scenarios, there are many researchers who require some level of automation in order to cope with large volumes of repetitive tasks. Common use cases such as replicating data across systems at multiple institutions, staging data captured from an instrument for analysis, and enabling access to reference data for a large user community, cannot be handled at scale by ad hoc methods.

Over the past year, Globus has released a set of robust APIs that facilitate integration of data management tasks into new and existing research workflows. In this tutorial, we will describe and demonstrate how researchers, campus computing administrators, and application developers, can access these advanced data management capabilities in an automated fashion, using existing organizational credentials. A combination of presentation and hands-on exercises will result in attendees building simple, yet fully functional, programs that can be used in their own work.

Room: 301

1:30 pm - 5:00 pm

Data Center Design and Planning for HPC Folks

Presenters: *Sharan Kalwani (DataSwing), Michael Thomas (Environmental Systems Design Inc), Bernie Woytek (Gensler Inc)*

High Performance Computing continues to rise and is no longer the domain of the special centers. It is inevitable that it will place demands on facilities and provide for the care and feeding of the external resources needed. It is common for HPC admins to get tasked at the last minute, with planning the data center needs specific to HPC. This tutorial addresses that since there is no other course available to impart the basic HPC-centric facilities planning. We know how to run codes efficiently but not to how to manage the surrounding infrastructure planning process. Data center design is critical as otherwise HPC associated expensive resources can be hobbled by lack of proper data center facilities.

In this tutorial, the authors propose to train users on how to begin the planning process, how to take stock of the current environmental demands, space, cooling, electricity, security, operations, and maintenance as well as foresee potential problems, point out methods and sources helpful to diagnose shortcoming in existing or incomplete data centers and develop data and insight necessary to distinguish and document correct and incorrect implementations.

Room: 303

1:30 pm - 5:00 pm

Scalable HPC Visualization and Data Analysis Using VisIt

Presenters: *Cyrus Harrison (Lawrence Livermore National Laboratory), Kevin Griffin (Lawrence Livermore National Laboratory), David Pugmire (Oak Ridge National Laboratory), Robert Sisneros (National Center for Supercomputing Applications, University of Illinois)*

Visualization and data analysis are essential components of the scientific discovery process. Scientists and analysts running HPC simulations rely on visualization and analysis tools for data exploration, quantitative analysis, visual debugging, and communication of results. This half-day tutorial will provide SC17 attendees with a practical introduction to mesh-based HPC visualization and analysis using VisIt, an open source parallel scientific visualization and data analysis platform. We will provide an introduction to HPC visualization practices and couple this with hands-on experience using VisIt.

This tutorial includes: 1) An introduction to visualization techniques for mesh-based simulations. 2) A guided tour of VisIt. 3) End-to-end demonstration visualizing a CFD Blood Flow simulation. 4) Information on representing simulation mesh data for in situ processing and parallel IO.

This tutorial builds on the past success of VisIt tutorials, updated for SC17 with new content focused on solutions that allow attendees to represent simulation data for both in situ and IO output. Attendees will gain practical knowledge and recipes to help them effectively use VisIt to analyze data from their own simulations.

Room: 406

1:30 pm - 5:00 pm

Kokkos: Enabling Manycore Performance Portability for C++ Applications and Domain Specific Libraries/Languages

Presenters: *H. Carter Edwards (Sandia National Laboratories), Christian Trott (Sandia National Laboratories), Fernanda Foertter (Oak Ridge National Laboratory)*

The Kokkos library enables applications and domain specific libraries/libraries to implement intra-node thread scalable algorithms that are performance portable across diverse manycore architectures. Kokkos uses C++ template meta-programming, as opposed to compiler extensions or source-to-source translators, to map user code onto architecture-targeted mechanisms such as OpenMP, Pthreads, and CUDA. Kokkos' execution mapping inserts users' parallel code bodies into well-defined parallel patterns and then uses an architecture-appropriate scheduling to execute the computation. Kokkos' data mapping implements polymorphic layout multidimensional arrays that are allocated in architecture-abstracted memory spaces with a layout (e.g., row-major, column-major, tiled) appropriate for that architecture. By integrating execution and data mapping into a single programming model Kokkos eliminates the contemporary array-of-structures versus structure-of-arrays dilemma from user code. Kokkos' programming model consists of the following extensible abstractions: execution spaces where computations execute, execution policies for scheduling computations, parallel patterns, memory spaces where data is allocated, array layouts mapping multi-indices onto memory, and data access intent traits to portably map data accesses to architecture-specific mechanisms such as GPU texture cache. Tutorial participants will learn Kokkos' programming model through lectures, hands on exercises, and presented examples.

Workshop [\(back to top\)](#)

Sunday, November 12th

Room: 501

9:00 am - 12:30 pm

Correctness 2017: First International Workshop on Software Correctness for HPC Applications

Introduction - 1st International Workshop on Software Correctness for HPC Applications (Correctness 2017)

Ensuring the correctness of high-performance computing (HPC) applications is one of the fundamental challenges that developers and users of these applications face today. An application is correct when it performs what a user expects with respect to a specification. Given today's complex HPC software stack, correctness is very difficult to achieve---the use of combined parallel programming models (e.g., MPI+OpenMP), complex compiler optimizations/transformations, floating-point precision issues, and unanticipated scale-dependent behavior, are some of the challenges to achieve correctness. As emerging task-based programming models and heterogeneous architectures become more predominant in HPC, the level of non-determinism in applications increase, which makes the isolation of errors much harder. The aim of this workshop is to bring together researchers and developers to present and discuss novel ideas to address the problem of correctness in HPC.

A Verification Language for High Performance Computing *Stephen Siegel (University of Delaware)*

Break

Verifying Concurrency in an Adaptive Ocean Circulation Model

Alper Altuntas (National Center for Atmospheric Research)

We present a model checking approach for verifying the correctness of concurrency in numerical models. The forms of concurrency we address are from (1) coupled modeling where distinct components, e.g., ocean, wave, and atmospheric, exchange interface conditions during runtime, and (2) multi-instance modeling where local variations of the same numerical model are executed concurrently to minimize common (and therefore redundant) computations. We present general guidelines for representing these forms of concurrency in an abstract verification model and then apply them to an adaptive ocean circulation model that determines the geographic extent and severity of coastal floods. The ocean model employs multi-instance concurrency: a collection of engineering design and failure scenarios are concurrently simulated using patches, regions of a grid that grow and shrink based on the hydrodynamic changes induced by each scenario. We show how concurrency inherent in the simulation model can be represented in a verification model to ensure correctness and to automatically generate safe synchronization arrangements.

Quality Assurance and Error Identification for the Community Earth System Model

Allison H. Baker (National Center for Atmospheric Research)

Earth system models are valuable tools for furthering our understanding of past, present, and future climate states. Because these models tend to be large and complex as well as in a state of near constant development, quality assurance (and subsequent debugging) are critical pieces in the development cycle. Here, we describe our multi-year effort to better evaluate the quality and "correctness" of the Community Earth System Model (CESM), a widely-used climate model. Our approach depends on an initial coarse-grain ensemble-based consistency test to determine code correctness, which has already proved successful in practice. The additional capability desired is a means of easily tracing a coarse-grain failure to its root cause, and we discuss our strategy and promising efforts to date toward that goal.

A Family of Provably Correct Algorithms for Exact Triangle Counting

Tze Meng Low (Carnegie Mellon University)

In recent years, there has been renewed interest in casting graph algorithms in the language of linear algebra. By replacing the computations with appropriate operations over different semi-rings, different graph algorithms can be cast as a sequence of linear algebra operations. In this paper, we show that this recent trend allows us to leverage the Formal Linear Algebra Methods Environment (FLAME) methodology to formally specify, derive and implement a family of graph algorithms for counting the number of triangles (3-cliques) in a graph. In addition, we introduce a new back-end for the FLAME Application Programming Interface (API) that computes over sparse matrices stored in the compressed sparse row (CSR) format so that the correctness of the derived algorithm can be translated to the implementation.

Verifying MPI Applications with SimGridMC

Martin Quinson (ENS Rennes)

SimGridMC (also dubbed Mc SimGrid) is a stateful Model Checker for MPI applications. It is integrated to SimGrid, a framework mostly dedicated to predicting the performance of distributed applications. We describe the architecture of McSimGrid, and show how it copes with the state space explosion problem using Dynamic Partial Order Reduction and State Equality algorithms. As case studies we show how SimGrid can enforce safety and liveness properties for MPI applications, as well as global invariants over communication patterns.

Runtime Correctness Checking for Emerging Programming Paradigms

Joachim Protze (RWTH Aachen University)

With rapidly increasing concurrency, the HPC community is looking for new parallel programming paradigms to make best use of current and up-coming machines. Under the Japanese CREST funding program, the post-petascale HPC project developed the XMP programming paradigm, a pragma-based partitioned global address space (PGAS) approach. Good tool support for debugging and performance analysis is crucial for the productivity and therefore acceptance of a new programming paradigm.

In this work, we investigate which properties of a parallel programming language specification may help tools to highlight correctness and performance issues or help to avoid common issues in parallel programming in the first place. In this paper, we exercise these investigations on the example of XMP. We also investigate the question how to improve the reusability of existing correctness and performance analysis tools.

Verifying the Floating-Point Computation Equivalence of Manually and Automatically Differentiated Code

Markus Schordan (Lawrence Livermore National Laboratory)

The semantics of floating-point computations are known to be difficult to verify. Software verification tools often provide little or no support for floating-point semantics, making it difficult to prove the correctness of an optimized variant of a program involving floating-point computations. In this paper we present an approach for verifying the equivalence of two program variants involving non-trivial floating-point operations. The selected test case for our approach are two variants of a differentiated code - one automatically generated, the other manually written. The verification technique operates at the source level, therefore we also investigate the generated assembly code variants and reason on a set of selected compiler options and architectures to guarantee that the correctness proof also holds for the generated binaries.

Towards Self-Verification in Finite Difference Code Generation

Jan Huckelheim (Imperial College, London)

Code generation from domain-specific languages is becoming increasingly popular as a method to obtain optimised low-level code that performs well on a given platform and for a given problem instance. Ensuring the correctness of generated codes is crucial. At the same time, testing or manual inspection of the code is problematic, as the generated code can be complex and hard to read. Moreover, the generated code may change depending on the problem type, domain size, or target platform, making conventional code review or testing methods impractical. As a solution, we propose the integration of formal verification tools into the code generation process. We present a case study in which the CIVL verification tool is combined with the Devito finite difference framework that generates optimised stencil code for PDE solvers from symbolic equations. We show a selection of properties of the generated code that can be automatically specified and verified during the code generation process. Our approach allowed us to detect a previously unknown bug in the Devito code generation tool.

Adjourn

Room: 601

9:00 am - 12:30 pm

4th International Workshop on HPC User Support Tools (HUST-17)

Supercomputing centers exist to drive scientific discovery by supporting researchers in computational science fields. To make users more productive in the complex HPC environment, HPC centers employ user support teams. These teams serve many roles, from setting up accounts, to consulting on math libraries and code optimization, to managing HPC software stacks. Often, support teams struggle to adequately support scientists. HPC environments are extremely complex, and combined with the complexity of multi-user installations, exotic hardware, and maintaining research software, supporting HPC users can be extremely demanding.

With the fourth HUST workshop, we will continue to provide a necessary forum for system administrators, user support team members, tool developers, policy makers and end users. We will provide a forum to discuss support issues and we will provide a publication venue for current support developments. Best practices, user support tools, and any ideas to streamline user support at supercomputing centers are in scope.

Room: 702

9:00 am - 12:30 pm

MCHPC2017: Workshop on Memory Centric Programming for HPC

Opening Remarks: MCHPC'17: Workshop on Memory Centric Programming for HPC

Yonghong Yan (University of South Carolina)

Memory systems have been becoming increasingly complex in recent years. Performance optimization has shifted from computing to data access. This increase of memory complexity also demands significant system support, from compiler/runtime and tools to modeling and new programming paradigms.

Memory-centric programming refers to the notion and techniques of exposing the hardware memory system and its hierarchy, which include NUMA regions, shared and private caches, scratch pad, 3-D stack memory, and non-volatile memory, to the programmer for extreme performance programming via portable abstraction and APIs for explicit memory allocation, data movement and consistency enforcement between memories. The concept has been gradually adopted in main stream programming interfaces, for example and to name a few, the use of place to represent memory regions in OpenMP and X10, locale in Chapel, the use of shared modifier in CUDA or cache modifier in OpenACC for representing scratch-pad SRAM for GPUs, the memkind library and the recent effort for OpenMP memory management for supporting 3-D stack memory (HBM or HMC), and PMEM library for persistent memory programming. The MCHPC workshop aims to provide a community forum for discussing and sharing topics related to memory-centric programming, including programming and optimization practices, application development, programming interfaces and compiler/runtime innovations that focus on reducing data access delay in a memory hierarchy, on improving bandwidth utilization in both private and shared memory and cache, and on reducing energy consumption of memory systems for parallel execution. More details of the workshop can be found from <http://passlab.github.io/mchpc/mchpc2017.html>.

Keynote: Compiler and Runtime Challenges for Memory Centric Programming

Vivek Sarkar (Georgia Institute of Technology)

It is widely recognized that a major disruption is under way in computer hardware as processors strive to extend, and go beyond, the end-game of Moore's Law. This disruption will include new forms of processor and memory hierarchies, including near-memory computation structures. In this talk, we summarize compiler and runtime challenges for memory centric programming, based on past experiences with the X10 project at IBM and the Habanero project at Rice University and Georgia Tech. A key insight in addressing compiler challenges is to expand the state-of-the-art in analyzing and transforming explicitly-parallel programs, so as to encourage programmers to write forward-scalable layout-independent code rather than hardwiring their programs to specific hardware platforms and specific data layouts. A key insight in addressing runtime challenges is to focus on asynchrony in both computation and data movement, while supporting both in a unified and integrated manner. A cross-cutting opportunity across compilers and runtimes is to broaden the class of computation and data mappings that can be considered for future systems. Based on these and other insights, we will discuss recent trends in compilers and runtime systems that point the way towards possible directions for addressing the challenges of memory centric programming.

Break

Invited Talk: Persistent Memory: The Value to HPC and the Challenges

In this talk, Andy will describe the emerging Persistent Memory technology and how it can be applied to HPC-related use cases. Andy will also discuss some of the challenges using Persistent Memory, and the ongoing work the ecosystem is doing to mitigate those challenges.

Bit Contiguous Memory Allocation for Processing In Memory

John Leidel (Tactical Computing Laboratories)

Given the recent resurgence of research into processing in or near memory systems, we find an ever increasing need to augment traditional system software tools in order to make efficient use of the PIM hardware abstractions. One such architecture, the Micron In-Memory Intelligence (IMI) DRAM, provides a unique processing capability within the sense amp stride of a traditional 2D DRAM architecture. This accumulator processing circuit has the ability to compute both horizontally and vertically on pitch within the array. This unique processing capability requires a memory allocator that provides physical bit locality in order to ensure numerical consistency.

In this work we introduce a new memory allocation methodology that provides bit contiguous allocation mechanisms for horizontal and vertical memory allocations for the Micron IMI DRAM device architecture. Our methodology drastically reduces the complexity by which to find new, unallocated memory blocks by combining a sparse matrix representation of the array with dense continuity vectors that represent the relative probability of finding candidate free blocks. We demonstrate our methodology using a set of pathological and standard benchmark applications in both horizontal and vertical memory modes.

Beyond 16GB: Out-of-Core Stencil Computations

Istvan Z. Reguly (Pazmany Peter Catholic University)

Stencil computations are a key class of applications, widely used in the scientific computing community, and a class that has

particularly benefited from performance improvements on architectures with high memory bandwidth. Unfortunately, such architectures come with a limited amount of fast memory, which is limiting the size of the problems that can be efficiently solved. In this paper, we address this challenge by applying the well-known cache-blocking tiling technique to large scale stencil codes implemented using the OPS domain specific language, such as CloverLeaf 2D, CloverLeaf 3D, and OpenSBLI. We introduce a number of techniques and optimisations to help manage data resident in fast memory, and minimise data movement. Evaluating our work on Intel's Knights Landing Platform as well as NVIDIA P100 GPUs, we demonstrate that it is possible to solve 3 times larger problems than the on-chip memory size with at most 15% loss in efficiency.

NUMA Distance for Heterogeneous Memory

Michael Lang (Los Alamos National Laboratory)

Experience with Intel Xeon Phi suggests that NUMA alone is inadequate for assignment of pages to devices in heterogeneous memory systems. We argue that this is because NUMA is based on a single distance metric between all domains (i.e., number of devices "in between" the domains), while relationships between heterogeneous domains can and should be characterized by multiple metrics (e.g., latency, bandwidth, capacity). We therefore propose elaborating the concept of NUMA distance to give better and more intuitive control of placement of pages, while retaining most of the simplicity of the NUMA abstraction. This can be based on minor modification of the Linux kernel, with the possibility for further development by hardware vendors.

Evaluating GPGPU Memory Performance Through the C-AMAT Model

Hariharan Devarajan (Illinois Institute of Technology)

General Purpose Graphics Processing Units (GPGPU) have become a popular platform to accelerate computing. However, while they provide additional computing powers, GPGPU have put even more pressure on the already behindhand memory systems. Memory performance is an identified performance killer of GPGPU. Evaluating, understanding, and improving GPGPU data access delay is an imperative research issue of high-performance computing. In this study, we utilize the newly proposed C-AMAT (Concurrent Average Memory Access Time) model to measure the memory performance of GPGPU. We first introduce a GPGPU-specialized measurement design of C-AMAT. Then the modern GPGPU simulator, GPGPU-Sim, is used to carry the performance study. Finally, the performance results are analyzed.

Room: 502-503-504

9:00 am - 5:30 pm

7th Workshop on Python for High-Performance and Scientific Computing

Python is an established, high-level programming language with a large community in academia and industry. Scientists, engineers, and educators use Python for data science, high-performance computing, and distributed computing. Since Python is extremely easy to learn with a very clean syntax, it's well-suited for education in scientific computing. Programmers are much more productive by using Python.

The workshop will bring together researchers and practitioners using Python in all aspects of data science and high performance computing. The goal is to present Python applications, to discuss general topics regarding the use of Python, and to share experiences using Python in scientific computing. While Python is extremely strong in supporting human productivity, it still lacks in computational performance compared to 'traditional' HPC languages such as Fortran or C. We especially encourage authors to submit novel research in improving performance of Python applications as well as research on productivity of development with Python.

More information: <http://www.dlr.de/sc/pyhpc2017>

Room: 505

9:00 am - 5:30 pm

ESPM2 2017: Third International Workshop on Extreme Scale Programming Models and Middleware

ESPM2'17: Opening Remarks

Hari Subramoni (Ohio State University), Karl Schulz (Intel Corporation), Dhableswar Panda (Ohio State University)

Next generation architectures and systems being deployed are characterized by high concurrency, low memory per-core, and multiple levels of hierarchy and heterogeneity. These characteristics bring out new challenges in energy efficiency, fault-tolerance and, scalability. It is commonly believed that software has the biggest share of the responsibility to tackle these challenges. In other words, this responsibility is delegated to the next generation programming models and their associated middleware/runtimes. This workshop focuses on different aspects of programming models such as task-based parallelism (Charm++, OCR, X10, HPX, etc), PGAS (OpenSHMEM, UPC, CAF, Chapel, UPC++, etc.), BigData (Hadoop, Spark, etc), Deep Learning (Caffe, Microsoft CNTK, Google TensorFlow), directive-based languages (OpenMP, OpenACC) and hybrid MPI+X, etc. It also focuses on their associated middleware (unified runtimes, interoperability for hybrid programming, tight integration of MPI+X, and support for accelerators) for next generation systems and architectures. The ultimate objective of the ESPM2 workshop is to serve as a forum that brings together researchers from academia and industry working in the areas of programming models, runtime systems, compilation and

languages, and application developers.

Challenges in Programming Extreme Scale Systems

William Gropp (University of Illinois)

After over two decades of relative architectural stability for distributed memory parallel computers, the end of Dennard scaling and the looming end of Moore's "law" is forcing major changes in computing systems. Can the community continue to use programming systems such as MPI for extreme scale systems? Does the growing complexity of compute nodes require new programming approaches? This talk will discuss some of the issues, with emphasis on internode and intra-node programming systems and the connections between them.

Morning Break

Addressing Global Data Dependencies in Heterogeneous Asynchronous Runtime Systems on GPUs

Bradley Peterson (University of Utah, Scientific Computing and Imaging Institute)

Large-scale parallel applications with complex global data dependencies beyond those of reductions pose significant scalability challenges in an asynchronous runtime system. Internodal challenges include identifying the all-to-all communication of data dependencies among the nodes. Intranodal challenges include gathering together these data dependencies into usable data objects while avoiding data duplication. This paper addresses these challenges within the context of a large-scale, industrial coal boiler simulation using the Uintah asynchronous many-task runtime system on GPU architectures. We show significant reduction in time spent analyzing data dependencies through refinements in our dependency search algorithm. Multiple task graphs are used to eliminate subsequent analysis when task graphs change in predictable and repeatable ways. Using a combined data store and task scheduler redesign reduces data dependency duplication ensuring that problems fit within host and GPU memory. These modifications did not require any changes to application code or sweeping changes to the Uintah runtime system. We report results running on the DOE Titan system on 119K CPU cores and 7.5K GPUs simultaneously. Our solutions can be generalized to other task dependency problems with global dependencies among thousands of nodes which must be processed efficiently at large scale.

HPX Smart Executors

Zahra Khatami (Louisiana State University)

The performance of many parallel applications depends on looplevel parallelism. However, manually parallelizing all loops may result in degrading parallel performance, as some of them cannot scale desirably to a large number of threads. In addition, the overheads of manually tuning loop parameters might prevent an application from reaching its maximum parallel performance. We illustrate how machine learning techniques can be applied to address these challenges. In this research, we develop a framework that is able to automatically capture the static and dynamic information of a loop. Moreover, we advocate a novel method by introducing HPX smart executors for determining the execution policy, chunk size, and prefetching distance of an HPX loop to achieve higher possible performance by feeding static information captured during compilation and runtime-based dynamic information to our learning model. Our evaluated execution results show that using these smart executors can speed up the HPX execution process by around 12% – 35% for the Matrix Multiplication, Stream and 2D Stencil benchmarks compared to setting their HPX loop's execution policy/parameters manually or using HPX auto-parallelization techniques.

Integrating OpenMP into the Charm++ Programming Model

Seonmyeong Bak (University of Illinois)

The recent trend of rapid increase in the number of cores per chip has resulted in vast amounts of on-node parallelism. These high core counts result in hardware variability that introduces imbalance. Applications are also becoming more complex themselves, resulting in dynamic load imbalance. Load imbalance of any kind can result in loss of performance and decrease in system utilization. In this paper, we propose a new integrated runtime system that adds OpenMP shared-memory parallelism to the Charm++ distributed programming model to improve load balancing on distributed systems. Our proposal utilizes an infrequent periodic assignment of work to cores based on load measurement, in combination with tasks created via OpenMP's parallel loop construct from each core to handle load imbalance. We demonstrate the benefits of using this integrated runtime system on the LLNL ASC proxy application Lassen, achieving speedups of 50% over runs without any load balancing and 10% over existing distributed-memory-only balancing schemes in Charm++.

Chapel-on-X: Exploring Tasking Runtimes for PGAS Languages

Akihiro Hayashi (Rice University)

With the shift to exascale computer systems, the importance of productive programming models for distributed systems is increasing. Partitioned Global Address Space (PGAS) programming models aim to reduce the complexity of writing distributed-memory parallel programs by introducing global operations on distributed arrays, distributed task parallelism, directed synchronization, and mutual exclusion. However, a key challenge in the application of PGAS programming models is the improvement of compilers and runtime systems. In particular, one open question is how runtime systems meet the requirement of exascale systems, where a large number of asynchronous tasks are executed.

While there are various tasking runtimes such as Qthreads, OCR, and HCLib, there is no existing comparative study on PGAS tasking/ threading runtime systems. To explore runtime systems for PGAS programming languages, we have implemented OCR-based and HCLib-based Chapel runtimes and evaluated them with an initial focus on tasking and synchronization implementations.

The results show that our OCR and HCLib-based implementations can improve the performance of PGAS programs compared to the existing Qthreads backend of Chapel.

Lunch (participants on their own)

Automatic Risk-Based Selective Redundancy for Fault-Tolerant Task-Parallel HPC Applications

Omer Subasi (Pacific Northwest National Laboratory)

Silent data corruption (SDC) and fail-stop errors are the most hazardous error types in high-performance computing (HPC) systems. In this study, we present an automatic, efficient and lightweight redundancy mechanism to mitigate both error types. We propose partial task-replication and checkpointing for task-parallel HPC applications to mitigate silent and fail-stop errors. To avoid the prohibitive costs of complete replication, we introduce a lightweight selective replication mechanism. Using a fully automatic and transparent heuristics, we identify and selectively replicate only the reliability-critical tasks based on a risk metric. Our approach detects and corrects around 70% of silent errors with only 5% average performance overhead. Additionally, the performance overhead of the heuristic itself is negligible.

Verification of the Extended Roofline Model for Asynchronous Many Task Runtimes

Joshua Suetterlein (University of Delaware)

Asynchronous Many Task (AMT) runtimes promise application designers the ability to better utilize novel hardware resources and to take advantages of the idle times that might arise from the discrepancies due to mismatches between software and hardware components. To foresee possible problems between hardware and software components (described as mismatches), designers usually use models to predict and analyze application behaviors. However, current models are ill suited for the AMT crowd because of its dynamic behavior and agility. To this effect, we developed an extended roofline model that aims to provide upper bounds on execution for AMT frameworks. This work focuses on the validation and error characterization of this model using different statistical techniques and a large set of experiments to evaluate and characterize its error and its sources. We found out that in the worst case, the error can grow to an order of magnitude, however there are several techniques to increase the model accuracy given a machine configuration.

Afternoon Break

Extending the Open Community Runtime with External Application Support

Jiri Dokulil (University of Vienna)

The Open Community Runtime specification prescribes the way a task-parallel application has to be written, in order to give the runtime system the ability to automatically migrate work and data, provide fault tolerance, improve portability, etc. These constraints prevent an application from efficiently starting a new process to run another external program. We have designed an extension of the specification which provides exactly this functionality in a way that fits the task-based model. The bulk of our work is devoted to exploring the way the task-parallel application can interact with an external application without having to resort to using files on a physical drive for data exchange. To eliminate the need to make changes to the external application, the data is exposed via a virtual file system using the filesystem-in-userspace architecture.

Effective Programming Models for Deep Learning at Scale

Daniel Holmes (University of Edinburgh), Michael Houston (Nvidia Corporation), Prabhat M (Lawrence Berkeley National Laboratory), Jeffrey M. Squyres (Cisco Systems), Rick Stevens (Argonne National Laboratory)

Artificial intelligence (AI) has been an interesting research topic for many decades but has struggled to enter mainstream use. Deep Learning (DL) is one form of AI that has recently become more practicable and useful because of dramatic increases in the computational power and in the amount of training data available. Research labs are already using Deep Learning to progress scientific investigations in numerous fields. Commercial enterprises are starting to make product development and marketing decisions based on machine learning models. However, there is a worrying skills gap between the hype and the reality of getting business benefit from Deep Learning. To address this, we need to answer some urgent questions. What practical programming techniques (specifically, programming models and middleware options) should we be teaching new recruits into this area? What existing knowledge and experience (from HPC or elsewhere) should existing practitioners be leveraging? Do traditional big-iron supercomputers and HPC software techniques (including MPI or PGAS) have a place in this vibrant new sphere or is all about high-level scripting, complex workflows, and elastic cloud resources?

ESPM2'17: Closing Remarks

Hari Subramoni (Ohio State University), Karl Schulz (Intel Corporation), Dhableswar K. Panda (Ohio State University)

Room: 506

9:00 am - 5:30 pm

Workshop for Open Source Supercomputing

As we approach the end of lithographic/Dennard scaling, the HPC community needs a way to continue performance scaling. One way of providing that scaling is an increase in the number and diversity of specialized architectures tailored for specific applications. To accelerate the architecture specification and verification of these new architectures, more rapid prototyping methods are needed. At the same time, these new architectures need software stacks and programming models to be able to actually use these new designs.

In concert with this increase in architecture heterogeneity, there has been a consistent march towards development of open source based hardware and software solutions for each of these components to be used in lieu of existing closed source solutions.

We present the Workshop for Open Source Supercomputing. This is meant to be a workshop for exploring and collaborating on building an HPC system using open-source hardware and system software IP. The goal of this workshop is to engage the HPC community and explore open source solutions for constructing an HPC system -- from silicon to applications.

Room: 507

9:00 am - 5:30 pm

DataCloud 2017: The Eighth International Workshop on Data-Intensive Computing in the Clouds

Introduction - The Eighth International Workshop on Data-Intensive Computing in the Clouds

Applications and experiments in all areas of science are becoming increasingly complex and more demanding in terms of their computational and data requirements. Some applications generate data volumes reaching hundreds of terabytes and even petabytes. As scientific applications become more data intensive, the management of data resources and dataflow between the storage and compute resources is becoming the main bottleneck. Analyzing, visualizing, and disseminating these large data sets has become a major challenge and data intensive computing is now considered as the “fourth paradigm” in scientific discovery after theoretical, experimental, and computational science.

The eighth international workshop on Data-intensive Computing in the Clouds (DataCloud 2017) will provide the scientific community a dedicated forum for discussing new research, development, and deployment efforts in running data-intensive computing workloads on Cloud Computing infrastructures. The DataCloud 2017 workshop will focus on the use of cloud-based technologies to meet the new data intensive scientific challenges that are not well served by the current supercomputers, grids or compute-intensive clouds. We believe the workshop will be an excellent place to help the community define the current state, determine future goals, and present architectures and services for future clouds supporting data intensive computing.

Submarine: A Subscription-Based Data Streaming Framework for Integrating Large Facilities and Advanced Cyberinfrastructure

Daniel Balouek-Thomert (Rutgers University)

Large scientific facilities provide researchers with instrumentation, data, and data products that can accelerate scientific discovery. However, increasing data volumes coupled with limited local computational power prevents researchers from taking full advantage of what these facilities can offer. Many researchers looked into using commercial and academic cyberinfrastructure (CI) to process this data. Nevertheless, there remains a disconnect between large facilities and cyberinfrastructure that requires researchers to be actively part of the data processing cycle. The increasing complexity of cyberinfrastructure and data scale necessitates new data delivery models, those that can autonomously integrate large-scale scientific facilities and cyberinfrastructure to deliver real-time data and insights. In this paper, we present our initial efforts using the Ocean Observatories Initiative project as a use case. In particular, we present a subscription-based data streaming service for data delivery that leverages the Apache Kafka data streaming platform. We also show how our solution can automatically integrate large-scale facilities with cyberinfrastructure services for automated data processing.

Enosis: Bridging the Semantic Gap between File-Based and Object-Based Data Models

Anthony Kougkas (Illinois Institute of Technology)

File and block storage are well-defined concepts in computing and have been used as common components of computer systems for decades. Big data has led to new types of storage. The predominant data model in cloud storage is the object-based storage and it is highly successful. Object stores follow a simpler API with get() and put() operations to interact with the data. A wide variety of data analysis software have been developed around objects using their APIs. However, object storage and traditional file storage are designed for different purpose and for different applications. Organizations maintain file-based storage clusters and a high volume of existing data are stored in files. Moreover, many new applications need to access data from both types of storage.

In this paper, we first explore the key differences between object-based and the more traditional file-based storage systems. We have designed and implemented several file-to-object mapping algorithms to bridge the semantic gap between these data models. Our evaluation shows that by achieving an efficient such mapping, our library can grant 2x-27x higher performance against a naive one-to-one mapping and with minimal overheads. Our study exposes various strengths and weaknesses of each mapping strategy and frames the extended potential of a unified data access system.

Open Ethernet Drive: Evolution of Energy-Efficient Storage Technology

Hariharan Devarajan (Illinois Institute of Technology)

An Open Ethernet Drive, also known as an OED, is a new technology that embeds a low-powered processor, a fixed-size memory and an Ethernet card on a disk drive (e.g. HDD or SSD). All major storage vendors have introduced their respective implementations with similar architectural designs. As the technology progresses into its second generation, the performance characteristics have improved substantially. In this study, we first demonstrate the differences between two generations of the OED technology. We run a variety of benchmarks and applications to thoroughly evaluate the performance of this device and its compatibility with the current ecosystem. Furthermore, we investigate the performance and energy footprint of the OED technology when used as a storage server and as an I/O accelerator. Evaluation results show that OED technology can be a reliable, scalable, energy and cost efficient storage solution. It is a viable replacement for storage servers offering competitive performance while consuming only 10% of the power that a typical storage node needs. Finally, this study shows that OED's can be used as I/O accelerators capable of executing data-intensive computations (such as sorting, compression/decompression, etc.) on local data, whereby the expensive data movement is minimized resulting in low power consumption.

Break

Multiple Stream Job Performance Optimization with Source Operator Graph Transformations

Miyuru Dayarathna (WSO2 Inc)

Multiple distributed stream queries which are executed on stream processing systems need to be fine tuned to the compute cluster in order to harness the full potential of the hardware they run on. In this paper we describe an automatic technique for conducting such stream query optimization in the presence of multiple stream jobs. During this auto-tuning process we identify the structure of each program and conduct automatic program transformation to generate optimized unified streaming jobs. The operators on the unified secondary sample application are grouped into PEs considering their performance characteristics and the stream graph topology structure to produce high performance stream query network. We implemented this multiple stream query optimization technique on a mechanism called Tahitica. We demonstrate our approach's ability for producing optimized stream query performance by comparing it to naive deployments using two real world stream processing applications in the domains of healthcare and search advertising. Our stream query optimization approach reported 7.1% throughput performance improvement compared to a naive deployment.

Reliable Access to Massive Restricted Texts: Experience-Based Evaluation

Zong Peng (Indiana University)

Libraries are seeing growing numbers of digitized textual corpora with restrictions on their content. Probing and mining these massive corpora, of interest to scholars, can be cumbersome because of size, granularity, access restrictions, and organization. Efficient management of such a collection especially under failures depends on the primary storage system. In this paper, we identify the requirements for managing a massive text corpus based on experience in managing the 5.5 billion pages of the HathiTrust digital library. Using the requirements, we compare candidate storage solutions, and using a combination of experimental evaluation and comparison, to identify an optimum choice.

Seamless Infrastructure Customization and Performance Optimization for Time-Critical Services in Data Infrastructures

Spiros Koulouzis (University of Amsterdam), Zhiming Zhao (University of Amsterdam)

The increasing volumes of data being produced, curated, and made available by research infrastructure (RI) initiatives in the environmental science domain require services to optimise the delivery and staging of data on behalf of investigators and other users of scientific data. Specialised data services for managing data lifecycle, for creating and delivering data products, and for customised data processing, play a crucial role in RIs to serve their user communities. We describe our experiences identifying the time-critical requirements of environmental scientists making use of ICT research support infrastructure. We present a microservice based infrastructure optimisation suite called the Dynamic Real-time Infrastructure Planner (DRIP). We provide a case study whereby DRIP is used to optimise runtime service quality for a data subscription service provided by the Euro-Argo RI using EGI FedCloud and EUDAT's B2SAFE service.

Room: 603

9:00 am - 5:30 pm

Workshop on Exascale MPI (ExaMPI)

The aim of workshop is to bring together researchers and developers to present and discuss innovative algorithms and concepts in the Message Passing programming model and to create a forum for open and potentially controversial discussions on the future of MPI in the Exascale era. Possible workshop topics include innovative algorithms for collective operations, extensions to MPI, including datacentric models, scheduling/routing to avoid network congestion, "fault-tolerant" communication, interoperability of MPI and PGAS models, integration of task-parallel models in MPI, and use of MPI in large scale simulations.

Room: 605
9:00 am - 5:30 pm

ISAV 2017: In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization

Introduction - ISAV 2017: In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization

E. Wes Bethel (Lawrence Berkeley National Laboratory)

The considerable interest in the HPC community regarding in situ analysis and visualization is due to several factors. First is an I/O cost savings, where data is analyzed/visualized while being generated, without first storing to a filesystem. Second is the potential for increased accuracy, where fine temporal sampling of transient analysis might expose some complex behavior missed in coarse temporal sampling. Third is the ability to use all available resources, CPU's and accelerators, in the computation of analysis products.

The workshop brings together researchers, developers and practitioners from industry, academia, and government laboratories developing, applying, and deploying in situ methods in extreme-scale, high performance computing. The goal is to present research findings, lessons learned, and insights related to developing and applying in situ methods and infrastructure across a range of science and engineering applications in HPC environments; to discuss topics like opportunities presented by new architectures, existing infrastructure needs, requirements, and gaps, and experiences to foster and enable in situ analysis and visualization; to serve as a "center of gravity" for researchers, practitioners, and users/consumers of in situ methods and infrastructure in the HPC space.

Keynote: Computing Ubiquitous Statistics: Computational Challenges

Break

Scalable In Situ Analysis of Molecular Dynamics Simulations

Preeti Malakar (Argonne National Laboratory)

Analysis of scientific simulation data enables scientists to glean insight from simulations. In situ analysis, which can be simultaneously executed with the simulation, mitigates I/O bottlenecks and can accelerate discovery of new phenomena. However, in typical modes of operation, this requires either stalling simulation during analysis phase or transferring data for analysis. We study the scalability challenges of time- and space-shared modes of analyzing large-scale molecular dynamics simulations. We also propose topology-aware mapping for simulation and analysis. We demonstrate the benefits of our approach using LAMMPS code on two supercomputers.

In Situ Visualization of Radiation Transport Geometry

Mark Kim (Oak Ridge National Laboratory)

The ultimate goal for radiation transport is to perform full-core reactor modeling and simulation. Advances in computational simulation bring this goal close to reality and the newest Monte Carlo transport codes have begun to shift to using accelerators that have become a stalwart in the supercomputing and HPC space.

Within the reactor modeling and simulation community, Monte Carlo transport simulations are considered the gold standard for simulation. Through the use of "combinatorial geometry" (constructive solid geometry), complex models can be used with fewer approximation compromises while at the same time scale to run on some of the largest supercomputers in the world.

Unfortunately, the state-of-the-art for "combinatorial geometry" visualization is to decompose the geometry into a mesh. This approach could require a significant amount of memory which is antithetical to in situ visualization. To address this issue, we introduce a ray caster for visualizing combinatorial geometry in radiation transport code. By only using the accelerators for the radiation transport code and leaving the CPU cores idle, there is an opportunity to conduct on node in situ visualization with the idle CPU cores, something domain experts have up to this point been unable to do. By utilizing VTK-m, the visualization can be run on the CPU as this particular application demands, but also run on any architecture that is supported by VTK-m, enabling future re-use across different platforms.

Cosmological Particle Data Compression in Practice

James Ahrens (Los Alamos National Laboratory)

In cosmological simulations, trillions of particles are handled and several terabytes of particle data are generated in each time step. Transferring this data directly from memory to disk in an uncompressed way results in a massive load on I/O and storage systems. Hence, one goal of domain scientists is to compress the data before storing it to disk while minimizing the loss of information. In this in situ scenario, the available time for the compression of one time step is limited. Therefore, the evaluation of compression techniques has shifted from only focusing on compression rates to including throughput and scalability. This study aims to evaluate and compare state-of-the-art compression techniques applied to particle data. For the investigated compression techniques,

quantitative performance indicators such as compression rates, throughput, scalability, and reconstruction errors are measured. Based on these factors, this study offers a comprehensive analysis of the individual techniques and discusses their applicability for in situ compression. Based on this study, future challenges and directions in the compression of cosmological particle data are identified.

Flexible In Situ Visualization of LAMMPS Simulations

Will Usher (University of Utah)

As simulations grow in scale, the need for in situ analysis methods to handle the large data produced grows correspondingly. Often, however, these methods are not interactive and are run as batch computations alongside the simulation, hampering exploratory analysis. In this work, we discuss a prototype integration of an interactive OSPRay based particle viewer with LAMMPS, via our communication library libIS, to implement an exploratory in situ analysis tool. The flexible design allows for running the viewer and simulation across the spectrum of in situ proximity configurations transparently. We use the SENSEI in situ interface to communicate data from LAMMPS to libIS and demonstrate our prototype system in different configurations on the Theta and Stampede 2 supercomputers using LAMMPS benchmarks.

realfast@VLA

Martin Pokorny (National Radio Astronomy Observatory)

We describe a system being deployed at the National Radio Astronomy Observatory's Karl G. Jansky Very Large Array (VLA) to commensally identify and record millisecond timescale astrophysical transient events in real time. This system distributes a high time resolution data stream to a dedicated fast transient detection system while allowing processing of a primary observation to continue with the typical (lower) time resolution data. This form of dual time resolution, commensal observing is enabled by the vys protocol, implemented with existing VLA computing infrastructure. The fast transient detection system performs real-time analysis in situ to detect events of interest and record relatively short duration data "cut-outs" of those events. By selectively recording high time resolution data, provided by vys at rates of up to 1.4 GB/s, realfast will reduce the recorded data volume by an estimated factor of up to 1000. This makes it possible to search for transients commensally in a high data rate stream over the thousands of hours needed to find the rarest events.

In Situ Visualization with Task-Based Parallelism

Alan Heirich (Stanford University, SLAC National Accelerator Laboratory)

This short paper describes an experimental prototype of in situ visualization in a task-based parallel programming framework. A set of reusable visualization tasks were composed with an existing simulation. The visualization tasks include a local OpenGL renderer, a parallel image compositor, and a display task. These tasks were added to an existing fluid-particle-radiation simulation, and weak scaling tests were run on up to 512 nodes of the Piz Daint supercomputer. Benchmarks showed that the visualization components scaled and did not reduce the simulation throughput. The compositor latency increased logarithmically with increasing node count.

In Situ Workflows at Exascale: System Software to the Rescue

Implementing an in situ workflow involves several challenges related to data placement, task scheduling, efficient communications, scalability, and reliability. Most of the current implementations provide reasonably performant solutions to these issues by focusing on high-performance communications and low-overhead execution models at the cost of reliability and flexibility. One of the key design choices in such infrastructures is between providing a single-program, integrated environment or a multiple-program, connected environment, both solutions having their own strengths and weaknesses. While these approaches might be appropriate for current production systems, the expected characteristics of exascale machines will shift current priorities. After a survey of the trade-offs and challenges of integrated and connected in situ workflow solutions available today, we discuss in this paper how exascale systems will impact those designs. In particular, we identify missing features of current system-level software required for the evolution of in situ workflows toward exascale and how system software innovations from the Argo Exascale Computing Project can help address those challenges.

A Novel Shard-Based Approach for Asynchronous Many-Task Models for In Situ Analysis

Philippe P. Pébay (Sandia National Laboratories)

We present the current status of our work towards a scalable, asynchronous many-task, in situ statistical analysis engine using the Legion runtime system, expanding upon earlier work, that was limited to a prototype implementation with a proxy mini-application as a surrogate for a full-scale scientific simulation code.

In contrast, we have more recently integrated our in situ analysis engines with S3D, a full-size scientific application, and conducted numerical tests on the largest computational platform currently available for DOE science applications.

The goal of this presentation is thus to describe the SPMD-Legion methodology that we have devised in this context, and to compare the data aggregation technique deployed herein to the approach taken within our previous work.

Lunch Break

Faodail: Enabling In Situ Analytics for Next-Generation Systems

Patrick M. Widener (Sandia National Laboratories)

Existing approaches for in situ analysis and visualization (ISAV) assume that scientific simulations are written in a bulk synchronous parallel (BSP) execution model. However, because of the projected increase in heterogeneity on future systems, alternative execution models, e.g., asynchronous many-task (AMT), have been proposed. In an AMT environment, application data is no longer fixed to a particular location which makes in situ processing more difficult. One solution to this problem is to enhance the data management services used by AMT runtimes to make their data accessible to both AMT applications and non-AMT ISAV tools. Decoupling these data management services from the AMT runtime provides an opportunity to support ISAV tools that can interact with either AMT or BSP applications.

In this paper, we introduce Faodail, a new data management layer that can support the diverse needs of multiple communities on modern platforms. While Faodail is designed to serve as a native data management service for Sandia's DARMA AMT framework, it also provides a flexible means of integrating applications with ISAV tools. We explore this idea with an example that connects ISAV tools to a particle-in-cell plasma simulation.

Parallel Streaming for In Transit Analysis with Heterogeneous Data Layout

Thomas Marrinan (University of St. Thomas)

Performing analysis or generating visualizations concurrently with high performance simulations can yield great benefits compared to post-processing data. Writing and reading large volumes of data can be reduced or eliminated, thereby producing an I/O cost savings. One such method for concurrent simulation and analysis is in transit - streaming data from the resource running the simulation to a separate resource running the analysis. In transit analysis can be beneficial since computational resources may not have certain resources needed for visualization and analysis (e.g. GPUs) and to reduce the impact of performing analysis tasks to the run time of the simulation. When sending and receiving data in transit, data redistribution mechanisms are needed in order to support heterogeneous data layouts that may be required by the simulation and analysis applications. The work described in this paper compares two mechanisms for on-the-fly data redistribution when streaming data in parallel between two distributed memory applications. Our results show that it is often advantageous to stream data in the same layout as the sender and redistribute data amongst processes on the receiving end than to stream data in the final layout needed by the receiver.

In Situ Summarization with VTK-m

Andrew Bauer (Kitware Inc)

Summarization and compression at current and future scales requires a framework for developing and benchmarking algorithms. We present a framework created by integrating existing, production-ready projects and provide timings of two particular algorithms that serve as exemplars for summarization: a wavelet-based data reduction filter and a generator for creating image-like databases of extracted features (isocontours in this case). Both support browser-based, post-hoc, interactive visualization of the summary for decision-making. A study of their weak-scaling on a distributed multi-GPU system is included.

Performance Impacts of In Situ Wavelet Compression on Scientific Simulations

Shaomeng Li (National Center for Atmospheric Research, University of Oregon)

In situ compression is a compromise between traditional post hoc and emerging in situ visualization and analysis. While the merits and limitations of various compressor options have been well studied, their performance impacts on scientific simulations are less clear, especially on large scale supercomputer systems. This study fills in this gap by performing in situ compression experiments on a leading supercomputer system. More specifically, we measured the computational and I/O impacts of a lossy wavelet compressor and analyzed the results with respect to various in situ processing concerns. We believe this study provides a better understanding of in situ compression as well as new evidence supporting its viability, in particular for wavelets.

The ALPINE In Situ Infrastructure: Ascending from the Ashes of Strawman

Matthew C. Larsen (Lawrence Livermore National Laboratory)

This paper introduces ALPINE, a flyweight in situ infrastructure. The infrastructure is designed for leading-edge supercomputers, and has support for both distributed-memory and shared-memory parallelism. It can take advantage of computing power on both conventional CPU architectures and on many-core architectures such as NVIDIA GPUs or the Intel Xeon Phi. Further, it has a flexible design that supports for integration of new visualization and analysis routines and libraries. The paper describes ALPINE's interface choices and architecture, and also reports on initial experiments performed using the infrastructure.

Data Analysis of Earth System Simulation within an In Situ Infrastructure

This talk covers the background, significance, and challenges of using in-situ analysis for Earth system modeling. It presents a generic procedure to implement a scalable and high performance data analysis framework for large-scale scientific simulation and demonstrates a unique capability for global Earth system simulations using advanced computing technologies (i.e., automated code analysis and instrumentation), in-situ infrastructure (i.e., ADIOS) and big data analysis engines (i.e., SciKit-learn).

Break

Discussion/Open Mic Session/Closing Remarks

Room: 607
9:00 am - 5:30 pm

Innovating the Network for Data Intensive Science (INDIS)

Wide area networks are now an integral and essential part of this data-driven supercomputing ecosystem connecting information sources, data stores, processing, simulation, visualization and user communities together. Networks for data-intensive science have more extreme requirements than general-purpose networks. These requirements not only closely impact the design of processor interconnects in supercomputers and cluster computers, but they also impact campus networks, regional networks and national backbone networks. This workshop brings together the network researchers and innovators to present challenges and novel ideas that stretch network research and SC's own innovative network, SCinet. We invite papers that propose new and novel techniques to present solutions for meeting these networking needs; and developments that are essential in the information systems infrastructure for the scientific discovery process.

Room: 704-706
9:00 am - 5:30 pm

Eighth Annual Workshop for the Energy Efficient HPC Working Group (EE HPC WG)

This annual workshop is organized by the Energy Efficient HPC Working Group (<http://eehpcwg.llnl.gov/>). This workshop closes the gap between facility and IT system with regards to energy efficiency analysis and improvements. For sustainable exascale computing, power and energy are a main concern, which can only be addressed by taking a holistic view combining the HPC facility, HPC system, HPC system software, and the HPC application needs. The EE HPC WG, which is a group with over 700 members from ~25 different countries, provides this cross-sectional perspective.

This workshop is unique in that it provides a forum for sharing power and energy related information and research from supercomputing centers from around the world. Discussion and audience participation is encouraged. There are presentations, panels and discussions. Presenters are mostly from major governmental and academic supercomputing centers. The panels encourage discussion around more controversial topics and include panelists from supercomputing centers, academic institutions as well as the vendor community. Topics include case studies of energy efficient operational lessons learned; the power grid- or "what you need to know about the power grid before adding a 10 MW step-function load generator"; the United States Department of Energy's Path Forward and other Exascale programs; and the software stack's implications for energy efficiency.

Arthur S Buddy Bland from Oak Ridge National Laboratory will be the keynote speaker. Buddy has seen more than 30 years of HPC deployment at ORNL and his keynote will provide insight into operations and energy efficiency for some of the largest supercomputers.

Room: 708
9:00 am - 5:30 pm

Medical Image Analysis and Visualization

Interest in quantitative image analysis and visualization in healthcare, from population-level academic research studies to patient-specific analysis has grown precipitously in the recent past. While applications for these tools to large-scale research have found significant support from the research community, opportunities at the clinical level where most healthcare effort and research is executed have been extremely limited, while their potential value has exploded. Interest and awareness of these tools has never been higher, as physicians and researchers have recognized how HPC and visualization can complement and even drive their own work. Mostly recently, the ever-increasing interest in machine learning, and particularly deep learning, to volumetric image analysis has once again begun to challenge both the HPC and medical communities. Lines of communication between these intersecting fields, however, have been extremely limited, stifling engagement and deployment of the most advanced imaging analysis and technology for healthcare research and clinical practice. Our workshop will bring together these disparate communities to engage on problems and solutions encountered in the practice of image analysis and visualization in healthcare. Thought leaders from across the aisle divide will come together to discuss problems where HPC, imaging analysis, and visualization can dramatically affect patient care. In addition, we will invite high-quality publications in key topic areas in medical image analysis to participate as invited speakers with full-length presentations. Lastly, we will hold practical sessions in which selected problems will be solved interactively with the audience, with code samples and iPython notebooks available for distribution to the audience.

Room: 710
9:00 am - 5:30 pm

Women in HPC: Diversifying the HPC Workforce

Introduction - Women in HPC: Diversifying the HPC Community

The seventh international Women in HPC workshop will be held at SC17, Denver, USA. Following the overwhelming success of the WHPC workshop at SC16, we will once again discuss the following topics: • How to successfully address discrimination in the workplace. • How to identify the roadblocks facing those in underrepresented groups that we may be overlooking. • The benefits of mentoring and professional networks.

We will also provide opportunities aimed at promoting and providing women with the skills to thrive in HPC including: • Poster session including 'lightning' talks by women working in HPC • Speed mentoring • Career Coaching session by Trish Damkroger (LLNL), certified executive coach. • Handling conflicts at workplace and how to respond to discrimination wisely • Short talks on: hints and tips for public speaking, how to take the next step in your career, effective workplace communication.

Self Branding and Advocacy: How to Get Known in Your Organization and Push Your Ideas Forward

Kelly Nolan (Talent Strategy)

Early Career Coaching

Morning Break

Negotiation Skills

Deb Goldfarb (Intel Corporation)

Hints and Tips for Public Speaking

Kelly Gaither (University of Texas)

How to Take the Next Step in Your Career

Elsa Gonsiorowski (Lawrence Livermore National Laboratory)

Effective Workplace Communication

Rebecca Hartman-Baker (Lawrence Berkeley National Laboratory)

Overcoming the Confidence Gap

Fernanda Foertter (Oak Ridge National Laboratory)

How to Find the Help You Need – Identifying Mentors and Those Who Can Help You in Your Career

Misbah Mubarak (Argonne National Laboratory)

Career Panel Discussion: Hints and Tips to Progress Your Career

Toni Collis (University of Edinburgh), Kelly Gaither (University of Texas), Deb Goldfarb (Intel Corporation), Elsa Gonsiorowski (Lawrence Livermore National Laboratory), Rebecca Hartman-Baker (Lawrence Berkeley National Laboratory), Fernanda Foertter (Oak Ridge National Laboratory), Misbah Mubarak (Argonne National Laboratory)

Speed Networking

Lunch (participants on their own)

Early Career Lightning Talks

Virtual Poster Networking and Mixer

Afternoon Break

Embracing Diversity: the Benefits

Toni Collis (University of Edinburgh)

The Benefits of Mentoring: Why and How to Set Up a Program

Scott Callaghan (University of Southern California)

Why Subtle Bias is Often Worse than Blatant Discrimination

Lorna Rivera (Georgia Institute of Technology)

Identifying the Roadblocks Facing Women in your Workforce

Lee Beausoleil (US Department of Defense)

Building a Community: Outreach Strategies for Coordinating a Local WHPC Program

Marisa Brazil (Purdue University)

From Outreach to Education to Researcher: Innovative Ways of Expanding the HPC Community

Nick Brown (University of Edinburgh)

Panel Discussion: Diversifying the HPC workforce

Kelly Gaither (University of Texas), Toni Collis (University of Edinburgh), Scott Callaghan (University of Southern California), Lorna Rivera (Georgia Institute of Technology), Lee Beausoleil (US Department of Defense), Marisa Brazil (Purdue University), Nick Brown (University of Edinburgh)

Workshop Outcomes and Closing

Toni Collis (University of Edinburgh)

Room: 712

9:00 am - 5:30 pm

Workshop on Extreme-Scale Programming Tools (ESPT)

The path to exascale computing will challenge HPC application developers in their quest to achieve the maximum potential that the machines have to offer. Factors such as limited power budgets, clock frequency variability, heterogeneous load imbalance, hierarchical memories, and shrinking I/O bandwidths will make it increasingly difficult to create high-performance applications. Tools for debugging, performance measurement and analysis, and tuning will be needed to overcome the architectural, system, and programming complexities envisioned in exascale environments. At the same time, research and development progress for HPC tools faces equally difficult challenges from exascale factors. Increased emphasis on autotuning, dynamic monitoring and adaptation, heterogeneous analysis, and so on will require new methodologies, techniques, and engagement with application teams. This workshop will serve as a forum for HPC application developers, system designers, and tools researchers to discuss the requirements for exascale-enabled tools and the roadblocks that need to be addressed. The workshop is the fifth in a series of SC conference workshops organized by the Virtual Institute - High Productivity Supercomputing (VI-HPS), an international initiative of HPC researchers and developers focused on parallel programming and performance tools for large-scale systems. The workshop includes a keynote address, peer-reviewed technical papers, and a lively panel session.

Room: 501

2:00 pm - 5:30 pm

The 2017 International Workshop on Software Engineering for High Performance Computing in Computational and Data-Enabled Science and Engineering (SE-CoDeSE 2017)

Introduction - The 2017 International Workshop on Software Engineering for High Performance Computing in Computational and Data-Enabled Science and Engineering (SE-CoDeSE 2017)

Researchers are increasingly using high performance computing (HPC), including GPGPUs and clusters, for computational and data-enabled science and engineering (CoDeSE) applications. Unfortunately, HPC software developers must solve reliability, availability, and maintainability problems at extreme scales, consider reproducibility, understand domain specific constraints, deal with uncertainties inherent in scientific exploration, and develop algorithms that use computing resources efficiently. Software engineering (SE) researchers have developed tools and practices to support development tasks, including: requirements management, design, validation and verification, testing (unit and system), continuous integration, and maintenance. HPC CoDeSE software requires appropriately tailored SE tools/methods. The SE-CoDeSE workshop addresses this need by bringing together members of the SE and HPC communities to share perspectives and present findings from research and practice (both successes and failures), and to generate an agenda to improve tools and practices for developing HPC software. This workshop builds on the success of 2013-2016 editions of similar workshops.

Software Engineering for Computational Science and Engineering: What Can Work and What Will Not

Michael A. Heroux (Sandia National Laboratories, St. John's University)

Computational Science and Engineering (CSE) exhibits a broad spectrum of software development and usage models and can benefit substantially from the knowledge and experience of the broader software engineering community. Even so, the nature of scientists, scientific research, and the role that software plays in scientific and engineering processes requires careful adaptation and adoption of mainstream concepts and practices.

In this talk, we highlight some of the key characteristics that shape the development and use of software in the CSE community. We then discuss some specific concepts and practices used in the broader software community that have strong appeal and promise for our community, and argue that some approaches used in the past, and still being used today, are not a good fit.

Supporting Software Engineering Practices in the Development of Data-Intensive HPC Applications with the JuML Framework

Markus Götz (Research Center Juelich), Matthias Book (University of Iceland), Christian Bodenstein (Research Center Juelich), Morris Riedel (Research Center Juelich)

Break

Position Paper: Experiences on Clustering High-Dimensional Data Using pbdR

Sadika Amreen (University of Tennessee)

Motivation: Software engineering for HPC environments, in general, and for big data, in particular, faces a set of unique challenges including the high complexity of middleware and of computing environments. Tools that make it easier for scientists to use HPC are, therefore, of paramount importance. We provide an experience report of using one of such highly effective middleware pbdR that

allow the scientist to use R programming language without, at least nominally, having to master many layers of HPC infrastructure, such as OpenMPI and ScaLAPACK.

Objective: To evaluate the extent to which middleware helps improve scientist productivity, we use pbdR to solve a real problem that we, as scientists, are investigating. Our big data comes from the commits on GitHub and other project hosting sites, and we are trying to cluster developers based on the text of these commit messages.

Context: We need to be able to identify the developer for every commit and to identify commits for a single developer. Developer identifiers in the commits, such as login, email, and name are often spelled in multiple ways since that information may come from different version control systems and may depend on which computer is used.

Method: We train a Doc2Vec model where existing credentials are used as a document identifier and then use the resulting 200-dimensional vectors for the 2.3M identifiers to cluster these identifiers so that each cluster represents a specific individual.

Position Paper: Software Engineering for Efficient Development of Flexible Numerical Software

Nathan Wukie (University of Cincinnati)

Computational physics and numerical modeling require substantial levels of software development. Recently, there have been community efforts to incorporate good software engineering practices within the development of scientific computing applications. This paper further advocates for the use of good software engineering practices in science and engineering and presents some software engineering solutions, challenges, and lessons learned from recent development efforts on a framework for computational fluid dynamics.

pFlogger: The Parallel Fortran Logging Framework for HPC Applications

Thomas Clune (NASA Goddard Space Flight Center)

In the context of HPC, software investments in support of text-based diagnostics, which monitor a running application, are typically limited compared to those for other types of IO. Examples of such diagnostics include reiteration of configuration parameters, progress indicators, simple metrics (e.g., mass conservation, convergence of solvers, etc.), and timers. To some degree, this difference in priority is justifiable as other forms of output are the primary products of a scientific model, and, due to their large data volume, much more likely to be a significant performance concern. In contrast, text-based diagnostic content is generally not shared beyond the individual or group running an application and is most often used to troubleshoot when something goes wrong.

We suggest that a more systematic approach enabled by a logging facility (or "logger") similar to those routinely used by many communities would provide significant value to complex scientific applications. In the context of high-performance computing, an appropriate logger would provide specialized support for distributed and shared-memory parallelism and have low performance overhead. In this presentation, we present our prototype implementation of pFlogger -- a parallel Fortran-based logging framework, and assess its suitability for use in a complex scientific application.

Proposal for a Scientific Software Lifecycle Model

Anshu Dubey (Argonne National Laboratory, University of Chicago)

Improvements in computational capabilities have lead to rising complexity in scientific modeling, simulation, and analytics and thus the software implementing them. In addition, a paradigm shift in platform architectures has added another dimension to complexity, to the point where software productivity (or the time, effort, and cost for software development, maintenance, and support) has emerged as a growing concern for computational science and engineering. Clearly communicating about the lifecycle of scientific software provides a foundation for community dialogue about processes and practices for various lifecycle phases that can improve developer productivity and software sustainability---key aspects of overall scientific productivity. While the mainstream software engineering community have produced lifecycle models that meet the needs of software projects in business and industry, none of the available models adequately describes the lifecycle of scientific computing software. In particular, software for end-to-end computations for obtaining scientific results has no formalized development model. Examining development approaches employed by teams implementing large multicomponent codes reveals a great deal of similarity in their strategies. In earlier work, we organized related approaches into workflow schematics, with loose coupling between submodels for development of scientific capabilities and reusable infrastructure. Here we consider an orthogonal approach, formulating models that capture the workflow of software development in slightly different scenarios, and we propose a scientific software lifecycle model based on agile principles.

Introduction to Small Group Discussion Exercise

Small Group Discussions

Small Group Reports and Workshop Wrapup

2:00 pm - 5:30 pm

Fourth SC Workshop on Best Practices for HPC Training

We propose to conduct a fourth annual half-day workshop on HPC training during the SC17 Conference.

Community interest in HPC training continues to grow as the organizers work to promote this critically important topic internationally. The efforts of the SC16 Conference, the Technical Program Committee, and the HPC Training workshop organizers to actively recruit participants resulted in more than 100 people attending the SC16 workshop. There continues to be strong international community participation via submissions, presentations and attendance during the annual SC Conference and through the year-long engagement efforts..

As a result of strong community need and interest in HPC training, and the significant level of community engagement at the previous SC Conference workshops on HPC training, the International HPC Training Consortium continues to flourish and grow. The Consortium conducts the planning and implementation of the SC HPC Training workshops, and the Consortium members engage in year-long efforts to sustain the interests identified during the SC workshops. Membership in the Consortium has grown from 70 individuals at SC16 to over 100 people from 18 countries.

The SC17 workshop will be used to highlight the results of collaborative efforts during 2017 to develop and deploy HPC training, to identify new challenges and opportunities, and to foster new, enhanced and expanded collaborations to pursue during 2018.

Room: 702

2:00 pm - 5:30 pm

Computational Reproducibility at Exascale 2017 (CRE2017)

Reproducibility is an important concern in all areas of computation. As such, computational reproducibility is receiving increasing interest from a variety of parties who are concerned with different aspects of computational reproducibility. Computational reproducibility encompasses several concerns including the sharing of code and data, as well as reproducible numerical results which may depend on operating system, tools, levels of parallelism, and numerical effects. In addition, the publication of reproducible computational results motivates a host of computational reproducibility concerns that arise from the fundamental notion of reproducibility of scientific results that has normally been restricted to experimental science. This workshop combines the Numerical Reproducibility at Exascale Workshops (conducted in 2015 and 2016 at SC) and the panel on Reproducibility held at SC16 (originally a BOF at SC15) to address several different issues in reproducibility that arise when computing at exascale. The workshop will include issues of numerical reproducibility as well as approaches and best practices to sharing and running code.

Monday, November 13th

Room: 501

9:00 am - 5:30 pm

WORKS 2017 (12th Workshop on Workflows in Support of Large-Scale Science)

Data Intensive Workflows (a.k.a. scientific workflows) are routinely used in most scientific disciplines today, especially in the context of e-Science. Workflows provide a systematic way of describing a scientific computation and rely on workflow management systems to execute on a variety of parallel and/or distributed resources. This workshop focuses on the many facets of scientific workflows and workflow management systems, ranging from (parallel) job execution to service orchestration and the coordination of data, service and job dependencies. The workshop therefore covers a broad range of issues in the scientific workflow lifecycle that include (but are not limited to): data intensive workflows representation and enactment; designing workflow composition interfaces; workflow mapping techniques to optimize the execution of the workflow; workflow enactment engines that need to deal with failures in the application and execution environment; and a number of computer science problems related to scientific workflows such as semantic technologies, compiler methods, fault detection and tolerance, performance modeling, scheduling, etc.

9:00 am - 5:30 pm

Introduction - WORKS 2017 (12th Workshop on Workflows in Support of Large-Scale Science)

Introduction - WORKS 2017 (12th Workshop on Workflows in Support of Large-Scale Science)

Data Intensive Workflows (a.k.a. scientific workflows) are routinely used in most scientific disciplines today, especially in the context of parallel and distributed computing. Workflows provide a systematic way of describing the analysis and rely on workflow management systems to execute the complex analyses on a variety of distributed resources. This workshop focuses on the many facets of data-intensive workflow management systems, ranging from job execution to service management and the coordination of data, service and job dependencies. The workshop therefore covers a broad range of issues in the scientific workflow lifecycle that include: data intensive workflows representation and enactment; designing workflow composition interfaces; workflow mapping techniques that may optimize the execution of the workflow; workflow enactment engines that need to deal with failures in the application and execution environment; and a number of computer science problems related to scientific workflows such as semantic technologies, compiler methods, fault detection and tolerance.

Extreme Scale Data Management for In-Situ Scientific Workflows

Manish Parashar (Rutgers University)

Data staging and in-situ/in-transit data processing are emerging as attractive approaches for supporting extreme scale scientific workflows. These approaches can improve end-to-end performance by enabling efficient data sharing between coupled simulations and data analytics components of an in-situ workflow. However, complex and dynamic data access/exchange patterns coupled with architectural trends toward smaller memory per core and deeper memory hierarchies threaten to impact the effectiveness of this approach. In this talk, I will explore a policy-based autonomic data management approach that can adaptively respond at runtime to dynamic data management requirements. Specifically, I will formulate the autonomic data management approach and present the design and implementation of autonomic policies as well as cross layer mechanisms, and will experimentally demonstrate how these autonomic adaptations can tune the application behaviors and resource allocations at runtime while meeting the data management requirements and constraints. This research is part of the DataSpaces project at the Rutgers Discovery Informatics Institute.

Morning Break

A Compiler Transformation-Based Approach to Scientific Workflow Enactment

Matthias Janetschek (University of Innsbruck)

We investigate the application of compiler transformations to workflow applications using the Manycore Workflow Runtime Environment (MWRE), a compiler-based workflow environment for modern manycore computing architectures. MWRE translates scientific workflows into equivalent C++ programs and efficiently executes them using a novel callback mechanism for dependency resolution and data transfers, with explicit support for full-ahead scheduling. We evaluate four different classes of compiler transformations, analyse their advantages and possible solutions to overcome their limitations, and present experimental results for improving the performance of a combination of real-world and synthetic workflows through compiler transformations. Our experiments were able to improve the workflow enactment by a factor of two and to reduce the memory usage of the engine by up to 33%. We achieved a speedup of up to 1.7 by eliminating unnecessary activity invocations, an improved parallel throughput up to 2.8 times by transforming the workflow structure, and a better performance of the HEFT scheduling algorithm by up to 36%.

Supporting Task-level Fault-Tolerance in HPC Workflows by Launching MPI Jobs inside MPI Jobs

Matthieu Dorier (Argonne National Laboratory)

While the use of workflows for HPC is growing, MPI interoperability remains a challenge for workflow management systems. The MPI standard and/or its implementations provide a number of ways to build multiple-programs-multiple-data (MPMD) applications. These methods present limitations related to fault tolerance, and are not easy to use. In this paper, we advocate for a novel MPI_Comm_launch function acting as the parallel counterpart of a system(3) call. MPI_Comm_launch allows a child MPI application to be launched inside the resources originally held by processes of a parent MPI application. Two important aspects of MPI_Comm_launch is that it pauses the calling process, and runs the child processes on the parent's CPU cores, but in an isolated manner with respect to memory. This function makes it easier to build MPMD applications with well-decoupled subtasks. We show how this feature can provide better flexibility and better fault tolerance in ensemble simulations and HPC workflows. We report results showing 2x throughput improvement for application workflows with faults, and scaling results for challenging workloads up to 256 nodes.

On the Use of Burst Buffers for Accelerating Data-Intensive Scientific Workflows

Rafael Ferreira da Silva (University of Southern California)

Science applications frequently produce and consume large volumes of data, but delivering this data to and from compute resources can be challenging, as parallel file system performance is not keeping up with compute and memory performance. To mitigate this I/O bottleneck, some systems have deployed burst buffers, but their impact on performance for real-world workflow applications is not always clear. In this paper, we examine the impact of burst buffers through the remote-shared, allocatable burst buffers on the Cori system at NERSC. By running a subset of the SCEC CyberShake workflow, a production seismic hazard analysis workflow, we find that using burst buffers offers read and write improvements of about an order of magnitude, and these improvements lead to increased job performance, even for long-running CPU-bound jobs.

rvGAHP – Push-Based Job Submission Using Reverse SSH Connections

Scott Callaghan (University of Southern California)

Computational science researchers running large-scale scientific workflow applications often want to run their workflows on the

largest available compute systems to improve time to solution. Workflow tools used in distributed, heterogeneous, high performance computing environments typically rely on either a push-based or a pull-based approach for resource provisioning from these compute systems. However, many large clusters have moved to two-factor authentication for job submission, making traditional automated push-based job submission impossible. On the other hand, pull-based approaches such as pilot jobs may lead to increased complexity and a reduction in node-hour efficiency. In this paper, we describe a new, efficient approach based on HTCondor-G called reverse GAHP (rvGAHP) that allows us to push jobs using reverse SSH submissions with better efficiency than pull-based methods. We successfully used this approach to perform a large probabilistic seismic hazard analysis study using SCEC's CyberShake workflow in March 2017 on the Titan Cray XK7 hybrid system at Oak Ridge National Laboratory.

Tracking of Online Parameter Fine-Tuning in Scientific Workflows

Marta Mattoso (Federal University of Rio de Janeiro)

In long-lasting large-scale workflow executions, computational scientists need to adapt the workflow by fine-tuning several parameters of complex computational models. These specific tunings may significantly reduce overall execution time. In executions that last for weeks, for instance, they can easily lose track of what has been tuned at previous simulation stages if the adaptations are not registered properly. In this work, we propose a solution for tracking parameter fine-tunings at runtime. With support of sophisticated online data analysis, scientists get a detailed view of the execution, providing insights to determine when and how to tune parameters. We developed DfAdapter*, a tool that collects human adaptations in the dataflow, while the workflow runs with or without a Scientific Workflow Management System. It controls and stores specific parameter-tunings in a provenance database, relating the human adaptation actions with data for: domain, dataflow provenance, execution, and performance. An extended PROV-compliant data diagram records the adaptation data. We evaluate DfAdapter by plugging it into a high performance workflow built with the libMesh library. The experiments with real data, from the Oil and Gas domain, showed that tunings significantly reduced the simulation time. * DfAdapter Repository. Available at: <https://github.com/hpcdb/DfAdapter>

Lightweight Container Integration into Workflow Systems: A First Look at Singularity and Makeflow

Kyle Sweeney (University of Notre Dame)

Traditionally, HPC centers have relied on professional system administrators to install, upgrade, and manage operating systems and application software suitable for a particular site. This relieves end users of the burden of managing these details, but also limits users' freedom to deploy software environments specialized for their task. Container technology addresses this limitation by allowing the end user to create and deploy custom software environments for a task. However, when used at large scale with a workflow management system in the most obvious way, container images can result in considerable waste in both filesystem space and network transfers. To address this, we consider two methods of composing software and data elements into a container – stacked and flat composition – and use these to create an efficient architecture for integrating containers into workflow systems. We demonstrate these techniques using the Singularity lightweight container environment and the Makeflow workflow system, demonstrating that flat composition results in reduced storage consumption and network transfers, on both synthetic and bioinformatics workflow.

Lunch (participants on their own)

E-HPC: A Library for Elastic Resource Management in HPC Environments

Devarshi Ghoshal (Lawrence Berkeley National Laboratory), William Fox (Georgia Institute of Technology; University of California, San Francisco)

Next-generation data-intensive scientific workflows need to support streaming and real-time applications with dynamic resource needs on high performance computing (HPC) platforms. The static resource allocation model on current HPC systems that was designed for monolithic MPI applications is insufficient to support the elastic resource needs of current and future workflows. In this paper, we discuss the design, implementation and evaluation of Elastic-HPC (E-HPC), an elastic framework for managing resources for scientific workflows on current HPC systems. E-HPC considers a resource slot for a workflow as an elastic window that might map to different physical resources over the duration of a workflow. Our framework uses checkpoint-restart as the underlying mechanism to migrate workflow execution across the dynamic window of resources. E-HPC provides the foundation necessary to enable dynamic resource allocation of HPC resources that are needed for streaming and real-time workflows. E-HPC has negligible overhead beyond the cost of checkpointing. Additionally, E-HPC results in decreased turnaround time of workflows compared to traditional model of resource allocation for workflows, where resources are allocated per stage of the workflow. Our evaluation shows that E-HPC improves core hour utilization for common workflow resource use patterns and provides an effective framework for elastic expansion of resources for applications with dynamic resource needs.

Toward Preserving Results Confidentiality in Cloud-Based Scientific Workflows

Daniel de Oliveira (Fluminense Federal University)

Cloud computing has established itself as a solid computational model that allows for scientists to deploy their simulation-based experiments on distributed virtual resources to execute a wide range of scientific experiments. These experiments can be modeled as scientific workflows. Many of these workflows are data-intensive and produce a large volume of data, which is also stored in the cloud using storage services by Scientific Workflow Management Systems (SWfMS). One main issue regarding cloud storage services is confidentiality of stored data, i.e. if unauthorized people access data files they can infer knowledge about the results or even about the workflow structure. Encryption is a possible solution, but it may not be sufficient and a new level of security can be added to preserve data confidentiality: data dispersion. In order to reduce this risk, generated data files cannot be stored in the same bucket, or at least sensitive data files have to be distributed across many cloud storage. In this paper, we present IPConf, an

approach to preserve workflow results confidentiality in cloud storage. IPConf generates a distribution plan for data files generated during a workflow execution. This plan disperses data files in several cloud storage to preserve confidentiality. This distribution plan is then sent to the SWfMS that effectively stores generated data into specific buckets during workflow execution. Experiments performed using real data from SciPhy workflow executions indicate the potential of the proposed approach.

e-Science Central Workflows for Stream Processing: Adding Streaming Support to an Existing Workflow System

Simon Woodman (Newcastle University)

Workflows have proved to be a popular method for processing and analysing scientific data. The graphical programming approach adopted enables non-programmers to rapidly construct and share sophisticated analysis pipelines. While workflows have seen significant adoption for processing files in a batch fashion, their application to streams of data has been less widespread. While there are tools that can apply dataflows to streams of data, these tools generally tend to only operate in the streaming domain.

This paper proposes extensions to the workflow engine provided by the e-Science Central platform that enable it to enact workflows that contain mixtures of streaming and batch operations in a consistent manner while still retaining the provenance capture, auditing and sharing features provided by the underlying platform.

Break

Heuristic Dynamic Workflow Scheduling

Kris Bubendorfer (Victoria University of Wellington)

The advantages of cloud computing including elastic, on demand, and pay per use instances, provide an ideal model for resourcing large scale state-of-the-art scientific analyses. Large scale scientific experiments are typically represented as workflows and are the common model for characterizing escience experiments and data analytics. Hosting and managing scientific applications on the cloud poses new challenges in terms of workflow scheduling which is key in leveraging cloud benefits. Prior research has studied static scheduling when the number of workflows is known in advance and all are submitted at the same time. However, in practice, a scheduler may have to schedule an unpredictable stream of workflows. In this paper, we present a new algorithm, Dynamic Workload Scheduler (DWS). Our algorithm addresses scheduling of multiple workflows with the aim of satisfying the deadline for each workflow in a typical cloud environment in which workflows can be submitted at any time. Our results show that the DWS algorithm achieves an average 10% higher success rate in terms of fulfilling deadlines for different workloads and reduces the overall cost by an average 23% when compared to the most recent comparable algorithm.

Budget-Aware Scheduling Algorithms for Scientific Workflows on IaaS Cloud Platforms

Yves Robert (ENS Lyon, University of Tennessee)

This paper introduces several budget-aware algorithms to deploy scientific workflows on IaaS cloud platforms, where users can request Virtual Machines (VMs) of different types, each with specific cost and speed parameters. We use a realistic application/platform model with stochastic task weights, and VMs communicating through a datacenter. We extend two well-known algorithms, HEFT and MinMin, and make scheduling decisions based upon machine availability and available budget. During the mapping process, the budget-aware algorithms make conservative assumptions to avoid exceeding the initial budget; we further improve our results with refined versions that aim at re-scheduling some tasks onto faster VMs, thereby spending any budget fraction leftover by the first allocation. These refined variants are much more time-consuming than the former algorithms, so there is a trade-off to find in terms of scalability. We report an extensive set of simulations with workflows from the Pegasus benchmark suite. Budget-aware algorithms generally succeed in achieving efficient makespans while enforcing the given budget, and despite the uncertainty in task weights.

A Machine Learning Approach for Modular Workflow Performance Prediction

Alok Singh (San Diego Supercomputer Center)

Scientific workflows provide an opportunity for declarative computational experiment design in an intuitive and efficient way. A distributed workflow is typically executed on a variety of resources and it uses a variety of computational algorithms or tools to achieve the desired outcomes. Such a variety imposes additional complexity in scheduling these workflows on large scale computers. As computation becomes more distributed, insights into expected workload that a workflow presents become critical for effective resource allocation. In this paper, we present a modular framework that leverages Machine Learning for creating precise performance predictions of a workflow. The central idea is to partition a workflow in such a way that makes the task of forecasting each atomic unit manageable and gives us a way to combine the individual predictions efficiently. We recognize a combination of an executable and a specific physical resource as a single module. This gives us a handle to characterize workload and machine power as a single unit of prediction. The modular approach of the presented framework allows it to adapt to highly complex nested workflows and scale to new scenarios. We present performance estimation results of independent workflow modules executed on the XSEDE SDSC Comet cluster using various Machine Learning algorithms. The results provide insights into the behavior and effectiveness of different algorithms in the context of scientific workflow performance prediction.

Processing of Crowd-Sourced Data from an Internet of Floating Things

Raffaele Montella (Parthenope University of Naples)

Sensors incorporated into mobile devices provide unique opportunities to capture detailed environmental information that cannot be readily collected in other ways. We show here how data from networked navigational sensors on leisure vessels can be used to

construct unique new datasets, using the example of underwater topography (bathymetry) to demonstrate the approach. Specifically, we describe an end-to-end workflow that involves the collection of large numbers of timestamped (position, depth) measurements from "internet of floating things" devices on leisure vessels; the communication of data to cloud resources, via a specialized protocol capable of dealing with delayed, intermittent, or even disconnected networks; the integration of measurement data into cloud storage; the efficient correction and interpolation of measurements on a cloud computing platform; and the creation of a continuously updated bathymetric database. Our prototype implementation of this workflow leverages the FACE-IT Galaxy workflow engine to integrate network communication and database components with a CUDA-enabled algorithm running in a virtualized cloud environment.

Promoting Scientific Workflows

Scott Lathrop (Shodor and University of Illinois), Jarek Nabrzyski (University of Notre Dame), Daniel S. Katz (University of Illinois)

Room: 502-503-504

9:00 am - 5:30 pm

Machine Learning in HPC Environments

Introduction - Machine Learning in HPC Environments

The intent of this workshop is to bring together researchers, practitioners, and scientific communities to discuss methods that utilize extreme scale systems for machine learning. This workshop will focus on the greatest challenges in utilizing HPC for machine learning and methods for exploiting data parallelism, model parallelism, ensembles, and parameter search. We invite researchers and practitioners to participate in this workshop to discuss the challenges in using HPC for machine learning and to share the wide range of applications that would benefit from HPC powered machine learning.

Leadership AI - Keynote by Jack Wells - Director of Science - Oak Ridge Leadership Computing Facility

Morning Break

TensorQuant - A Simulation Toolbox for Deep Neural Network Quantization

Dominik Marek Lorocho (Fraunhofer Institute for Industrial Mathematics)

Recent research implies that training and inference of deep neural networks (DNN) can be computed with low precision numerical representations of the training/test data, weights and gradients without a general loss in accuracy. The benefit of such compact representations is twofold: they allow a significant reduction of the communication bottleneck in distributed DNN training and faster neural network implementations on hardware accelerators like FPGAs. Several quantization methods have been proposed to map the original 32-bit floating point problem to low-bit representations. While most related publications validate the proposed approach on a single DNN topology, it appears to be evident, that the optimal choice of the quantization method and number of coding bits is topology dependent. To this end, there is no general theory available, which would allow users to derive the optimal quantization during the design of a DNN topology.

In this paper, we present a quantization tool box for the Tensor Flow framework. TensorQuant allows a transparent quantization simulation of existing DNN topologies during training and inference. TensorQuant supports generic quantization methods and allows to experimentally evaluate the impact of the quantization on single layers as well as the on the full topology. In a first series of experiments with TensorQuant, we show an analysis of fix-point quantizations of popular CNN topologies.

An Efficient Task-Based All-Reduce for Machine Learning Applications

Zhenyu Li (University of Warwick)

All-Reduce is a collective-combine operation frequently utilised in synchronous parameter updates in parallel machine learning algorithms. The performance of this operation - and subsequently of the algorithm itself - is heavily dependent on its implementation, configuration and on the supporting hardware on which it is run. Given the pivotal role of all-reduce, a failure in any of these regards will significantly impact the resulting scientific output.

In this research, we explore the performance of alternative all-reduce algorithms in data-flow graphs and compare these to the commonly used reduce-broadcast approach. We present an architecture and interface for all-reduce in task-based frameworks, and a parallelization scheme for object-serialization and computation. We present a concrete, novel application of a butterfly all-reduce algorithm on the Apache Spark framework on a high-performance compute cluster, and demonstrate the effectiveness of the new butterfly algorithm with a logarithmic speed-up with respect to the vector length compared with the original reduce-broadcast method - a 9x speed-up is observed for vector lengths in the order of 10^8 . This improvement is comprised of both algorithmic changes (65%) and parallel-processing optimization (35%).

The effectiveness of the new butterfly all-reduce is demonstrated using real-world neural network applications with the Spark framework. For the model-update operation we observe significant speed-ups using the new butterfly algorithm compared with the original reduce-broadcast, for both smaller (Cifar and Mnist) and larger (ImageNet) datasets.

Accelerating Deep Neural Network Learning for Speech Recognition on a Cluster of GPUs

Guojing Cong (IBM)

We train deep neural networks to solve the acoustic modeling problem for large-vocabulary continuous speech recognition. We employ distributed processing using a cluster of GPUs. On modern GPUs, the sequential implementation takes over a day to train, and efficient parallelization without losing accuracy is notoriously hard. We show that ASGD methods for parallelization are not efficient for this application. Even with 4 GPUs, the overhead is significant, and the accuracies achieved are poor. We adapt a P-learner K-step model averaging algorithm that with 4 GPUs achieves accuracies comparable to that achieved by the sequential implementation. We further introduce adaptive measures that make our parallel implementation scale to the full cluster of 20 GPUs. Ultimately our parallel implementation achieves better accuracies than the sequential implementation with a 6.1 times speedup.

Optimizing Convolutional Neural Networks for Cloud Detection

Deep convolutional neural networks (CNNs) have become extremely popular and successful at a number of machine learning tasks. One of the great challenges of successfully deploying a CNN is designing the network: specifying the network topology (sequence of layer types) and configuring the network (setting all the internal layer hyper-parameters). There are a number of techniques which are commonly used to design the network. One of the most successful is a simple (but lengthy) random search.

In this paper, we demonstrate how a random search can be dramatically improved by a two-phase search. The first phase is a traditional random search on n network configurations. The second phase exploits a support vector machine to guide a second random search on N network configurations. We apply this technique to a dataset containing satellite imagery and demonstrate that we can, with very high accuracy, identify regions containing clouds which obscure the landscape below.

Lunch (participants on their own)

Production Deep Learning and Scale - Keynote by Mike Houston - Senior Distinguished Engineer - Deep Learning - Nvidia

Training Distributed Deep Recurrent Neural Networks with Mixed Precision on GPU Clusters

Alexey Svyatkovskiy (Princeton University)

In this paper, we evaluate training of deep recurrent neural networks with half-precision floats. We implement a distributed, data-parallel, synchronous training algorithm by integrating TensorFlow and CUDA-aware MPI to enable execution across multiple GPU nodes and making use of high-speed interconnects. We introduce a learning rate schedule facilitating neural network convergence at up to $O(100)$ workers.

Strong scaling tests performed on clusters of NVIDIA Pascal P100 GPUs show linear runtime scaling and logarithmic communication time scaling for both single and mixed precision training modes. Performance is evaluated on a scientific dataset taken from the Joint European Torus (JET) tokamak, containing multi-modal time series of sensory measurements leading up to deleterious events called plasma disruptions. Half-precision significantly reduces memory and network bandwidth, allowing training of state-of-the-art models with over 70 million trainable parameters while achieving a comparable test set performance as single precision.

Afternoon Break

Toward Scalable Parallel Training of Deep Neural Networks

Sam Ade Jacobs (Lawrence Livermore National Laboratory)

We propose a new framework for parallelizing deep neural network training that maximizes the amount of data that is ingested by the training algorithm. Our proposed framework called Livermore Tournament Fast Batch Learning (LTFB) targets large-scale data problems. The LTFB approach creates a set of Deep Neural Network (DNN) models and trains each instance of these models independently and in parallel. Periodically, each model selects another model to pair with, exchanges models, and then run a local tournament against held-out tournament datasets. The winning model continues training on the local training datasets. This new approach maximizes computation and minimizes amount of synchronization required in training deep neural network, a major bottleneck in existing synchronous deep learning algorithms. We evaluate our proposed algorithm on two HPC machines at Lawrence Livermore National Laboratory including an early access IBM Power8+ with NVIDIA Tesla P100 GPUs machine. Experimental evaluations of the LTFB framework on two popular image classification benchmark: CIFAR10 and ImageNet, show significant speed up compared to the sequential baseline.

BlazingText: Scaling and Accelerating Word2Vec using Multiple GPUs

Word2Vec is a popular algorithm used for generating dense vector representations of words in large corpora using unsupervised learning. The resulting vectors have been shown to capture semantic relationships between the corresponding words and are used extensively for many downstream natural language processing (NLP) tasks like sentiment analysis, named entity recognition and machine translation. Most open-source implementations of the algorithm have been parallelized for multi-core CPU architectures including the original C implementation by Mikolov et al. and FastText by Facebook. A few other implementations have attempted to leverage GPU parallelization but at the cost of accuracy and scalability. In this work, we present BlazingText, a highly optimized implementation of Word2Vec in CUDA, that can leverage multiple GPUs for training. BlazingText can achieve a training speed of up to 43M words/sec on 8 GPUs, which is a 9x speedup over 8-threaded CPU implementations, with minimal effect on the quality of the embeddings.

An In-Depth Performance Characterization of CPU- and GPU-Based DNN Training on Modern Architectures

Ammar Ahmad Awan (Ohio State University)

Traditionally, Deep Learning (DL) frameworks like Caffe, TensorFlow, and Cognitive Toolkit exploited GPUs to accelerate the training process. This has been primarily achieved by aggressive improvements in parallel hardware as well as through sophisticated software frameworks like cuDNN and cuBLAS. However, recent enhancements to CPU-based hardware and software has the potential to significantly enhance the performance of CPU-based DL training. In this paper, we provide a complete performance landscape of CPU- and GPU-based DNN training. We characterize performance of DNN training for AlexNet and ResNet-50 for a wide-range of CPU and GPU architectures including the latest Intel Xeon Phi (Knights Landing) processors and NVIDIA Pascal GPUs. We also present multi-node DNN training performance results for AlexNet and ResNet-50 using Intel Machine Learning Scaling (MLSL) Library and Intel-Caffe. In addition, we provide a CPU vs. GPU comparison for multi-node training using OSU-Caffe and Intel-Caffe. To the best of our knowledge, this is the first study that dives deeper into the performance of DNN training in a holistic manner yet provides an in-depth look at layer-wise performance for different DNNs.

We provide multiple key insights: 1) Convolutions account for the majority of time (up to 83% time) consumed in DNN training, 2) GPU-based training continues to deliver excellent performance (up to 18% better than KNL) across generations of GPU hardware and software, and 3) Recent CPU-based optimizations like MKL-DNN and OpenMP-based thread parallelism leads to excellent speed-ups over under-optimized designs (up to 3.2X improvement for AlexNet training).

Designing a Synchronization-Reducing Clustering Method on Manycores: Some Issues and Improvements

Weijian Zheng (Indiana University-Purdue University Indianapolis)

The k-means clustering method is one of the most widely used techniques in big data analytics. In this paper, we explore the ideas of software blocking, asynchronous local optimizations, and heuristics of simulated annealing to improve the performance of k-means clustering. Like most of the machine learning methods, the performance of k-means clustering relies on two main factors: the computing speed (per iteration), and the convergence rate. A straightforward realization of the software-blocking synchronization-reducing clustering algorithm, however, sees sporadic slower convergence rate than the standard k-means algorithm. To tackle the issues, we design an annealing-enhanced algorithm, which introduces the heuristics of stop conditions and annealing steps to provide as good or better performance than the standard k-means algorithm. This new enhanced k-means clustering algorithm is able to offer the same clustering quality as the standard k-means. Experiments with real-world datasets show that the new parallel implementation is faster than the open source HPC library of Parallel K-Means Data Clustering (e.g., 19% faster on relatively large datasets with 32 CPU cores, and 11% faster on a large dataset with 1,024 CPU cores). Moreover, the extent to which the program performance improves is largely determined by the actual convergence rate of applying the algorithm to different datasets.

Evolving Deep Networks Using HPC

While a large number of deep learning networks have been studied and published that produce outstanding results on natural image datasets, these datasets only make up a fraction of those to which deep learning can be applied. These datasets include text data, audio data, and arrays of sensors that have very different characteristics than natural images. As these "best" networks for natural images have been largely discovered through experimentation and cannot be proven optimal on some theoretical basis, there is no reason to believe that they are the optimal network for these drastically different datasets. Thus, hyperparameter search is often a very important process when applying deep learning to a new problem. In this work, we present an evolutionary approach to searching the possible space of network hyperparameters and construction that can scale to 18,000 nodes. This approach is applied to datasets of varying types and characteristics where we demonstrate the ability to rapidly find best hyperparameters in order to reduce enable practitioners to quickly iterate between idea and result.

Room: 505

9:00 am - 5:30 pm

Workshop on Education for High Performance Computing (EduHPC)

Introduction - Workshop on Education for High Performance Computing (EduHPC)

The EduHPC Workshop is devoted to the development and assessment of educational resources for undergraduate and graduate education in High Performance Computing (HPC) and Parallel and Distributed Computing (PDC). This year we are broadening the scope to explicitly include data science curriculum (e.g. for new degree programs and within data science centers) and topics related to Internet of Things. PDC, HPC and data science now permeate the world of computing to a degree that makes it imperative for even entry-level computer professionals to incorporate these computing modalities into their computing toolboxes, no matter what type of computing problems they work on. This workshop focuses on the state of the art in HPC and PDC education via contributed and invited papers from academia, industry, government laboratories and other educational and research institutions. Topics of interest include all topics pertaining to the teaching of PDC and HPC within Computer Science and Engineering, Computational Science, and Domain Science and Engineering curricula. The emphasis of the workshop is undergraduate education, but fundamental issues related to graduate education are also welcome. The target audience will broadly include SC17 attendees from academia, industry, and research laboratories. This includes both researchers and educators, as well as the early adopters of the NSF/TCPP curriculum guidelines on teaching PDC (<http://www.cs.gsu.edu/~tcpp/curriculum/index.php>). The workshop is coordinated by the NSF-supported Center for Parallel and Distributed Computing Curriculum Development and Educational Resources (CDER). EduHPC has been an SC workshop since 2013 with attendance of 90 in 2015 and 75 in 2016.

Vistas in Advanced Computing

Peggy Lindner (University of Houston), Andrea Prosperetti (University of Houston)

This paper describes a summer program, "Vistas in Advanced Computing," in which rising sophomores were instructed on major components of High Performance Computing including C programming, numerical methods, computer architecture, parallel programming and other topics over a period of eight weeks in the summer of 2017. Students spent 6-7 hours a day in a classroom attending lectures, studying, doing homework and coding. The students were very positive about the experience, an evaluation which is supported by the data collected in the course of the program.

A Path from Serial Execution to Hybrid Parallelization for Learning HPC

Parallel and distributed computing are becoming necessary in almost all aspects of computation. Due to this growing demand, curriculum initiatives have been developed for integrating parallel and distributed computing into traditional undergraduate computer science programs. However, adoption has been slow resulting in many students lacking proper training for parallel and distributed computing. Two potential barriers for slow adoption are a deficiency in example programs that step students through the processes of parallelizing serial code, and the inaccessibility of dedicated machines to run highly parallel programs at scale within the confines of a course schedule. We have developed course material using a simple two-dimensional Lattice-Boltzmann Method Computational Fluid Dynamic simulation to walk students through shared memory parallelism, distributed memory parallelism, and hybrid parallel execution. We also created a custom mini-cluster comprised of 16 credit-card sized compute nodes, with a total of 288 cores, as an inexpensive solution for testing the scalability of different parallel models that can be deployed in a classroom setting.

Multidisciplinary Education on Big Data + HPC + Atmospheric Sciences

We present a new initiative to create a training program or graduate-level course (cybertraining.umbc.edu) in big data applied to atmospheric sciences as application area and using high-performance computing as an indispensable tool. The training consists of instruction in all three areas of "Big Data + HPC + Atmospheric Sciences" supported by teaching assistants and followed by faculty-guided project research in a multidisciplinary team of participants from each area. Participating graduate students, post-docs, and junior faculty from around the nation will be exposed to multidisciplinary research and have the opportunity for significant career impact. The paper discusses the challenges, proposed solutions, practical issues of the initiative, and how to integrate high-quality developmental program evaluation into the improvement of the initiative from the start to aid in ongoing development of the program.

Morning Break

Teaching, Learning and Collaborating through Cloud Computing Online Classes

Knowledge of parallel and distributed computing is important for students needing to address big data problems for jobs in either industry or academia; however, many college campuses do not offer courses in these areas due to curriculum limitations, insufficient faculty expertise, and instructional computing resources. Massively Open Online Courses (MOOCs) provide an opportunity to scale learning environments and help institutions advanced curriculum. In this paper, we discuss a Cloud Computing course offered at Indiana University and use it as a model for improving curriculum at institutions, which otherwise wouldn't be exposed to parallel and distributed computing.

Teaching Parallel Computing with Container Virtualization

Joshua Higgins (University of Huddersfield)

Incorporating modules that equip students with parallel programming and high performance computing skills into core computing and engineering courses poses unique technical challenges. A typical PC laboratory environment may not be suitable, and allowing

learners access to a production resource may introduce an unacceptable risk, if it is even possible. For decades, Beowulf clusters have offered an attractive solution where a system can be built using commodity components, often by the students themselves. This can instill in the learner both an understanding of parallel code and how it is orchestrated on the hardware - a strategy used successfully in delivering courses at the University of Huddersfield. However, there are two problems; Firstly, the depth of knowledge required to be learnt in order to successfully build, program and profile a cluster. Secondly, the use of seemingly antiquated tools in order to achieve this. In this paper, we outline some of the experience of integrating parallel and distributed computing into our courses. In addition, we present a novel, cross-platform virtual cluster toolkit developed to address the technical requirements of delivering such courses.

The OLCF GPU Hackathon Series: The Story Behind Advancing Scientific Applications with a Sustained Impact

The OLCF GPU Hackathons are a one-week code-development/learning event to better enable attendees to utilize GPUs. It only took three years to grow from a "Let's give this a try"-event to a repeatedly copied format with several spin-offs that inspired HPC centers around the world. Sticking to a few fundamental principles—work on your own code, learn from your mentors just what you need and when you need it, stay flexible in achieving your goal—the week long hackathon format created at Oak Ridge Leadership Computing Facility (OLCF) has been just the spark needed by many groups of scientists to light the fire of a wider GPU adoption in leading-edge as well as university-scale HPC environments. Most interestingly, the format enabled both ends of the experience spectrum—graduate students vs. postdoc fellows—the same kind of progress and chance of success.

Parallware Trainer: Interactive Tool for Experiential Learning of Parallel Programming Using OpenMP and OpenACC

Manuel Arenaz (University of A Coruña and Appentra Solutions)

STEM education plays a key role in the sustained growth and stability of the US economy and worldwide. There is currently a shortage of a skilled STEM workforce, and that gap is expected to grow widely in the next decade. It is key to widen the audience of STEM people trained in parallel programming targeting parallel architectures like Intel Xeon, IBM Power, NVIDIA GPU and Intel Xeon Phi. In this regard, the standards OpenMP 4.5 and OpenACC 2.5 offer pragma-based parallel programming paradigms that promise performance portability and higher productivity.

This paper presents Parallware Trainer, a new interactive tool for high-productivity STEM education and training in parallel programming using OpenMP 4.5 and OpenACC 2.5. It enables experiential learning by providing an interactive, real-time GUI with editor capabilities to assist in the design, implementation and benchmarking of OpenMP/OpenACC-enabled parallel code. We envision Parallware Trainer as a key enabler for STEM education from PhD down to undergraduate in computer science, maths, physics,.... This paper also describes a success story resulting from a GPU Hackathon organized at the Supercomputing Center of Galicia (CESGA). We present the progress of a 2-people team of the EDANYA group learning how to address the parallelization of a simulation code for prediction of tsunamis used by the National Oceanic and Atmospheric Administration (NOAA).

Lowering Barriers into HPC through Open Education

In this paper, we describe a trilogy of Massive Open Online Courses (MOOCs) that together expose knowledge and skills of fundamental importance to HPC. Linear Algebra: Foundations to Frontiers (LAFF) covers topics found in an introductory undergraduate course on linear algebra. It links abstraction in mathematics to abstraction in programming, with many enrichments that connect to HPC. LAFF- On Programming for Correctness introduces how to systematically derive programs to be correct. Of importance to HPC is that this methodology yields families of algorithms so that the best one for a given situation can be chosen. Programming for HPC (working title) is in the design stage. We envision using a very simple example, matrix-matrix multiplication, to illustrate how to achieve performance on a single core, on multicore and many-core architectures, and on distributed memory computers. These materials lower barriers into HPC by presenting insights, supports, and challenges to novices and HPC experts while scaling access to the world.

A Proposed Model for Teaching Advanced Parallel Computing and Related Topics

The heterogeneous, multi-core and many-core architecture of modern computers offers opportunities to solve critical problems in science and engineering as well as parallel programming challenges to those who wish to use them. This problem includes computer science programs but extends to a wide variety of STEM programs where computation has become an integral part of the research infrastructure. Students from science and engineering majors often have limited programming experience. Instructors from these fields are often not knowledgeable enough about the techniques used for efficient parallel computing to create and run these courses independently. Even in places where there is faculty expertise to teach the courses, they tend to be low in enrollment and thus are less likely to be offered. This leaves students who wish to learn the advanced concepts of parallel and large-scale computing to seek resources elsewhere or stumble on blindly in their field and pick things up as they go. Although some of the required content is available online as self-paced or MOOC style courses, those approaches are not effective in getting the majority of the target students the level of expertise they require. Both the Blue Waters and XSEDE projects have experimented with collaborative online courses that have proved to be an excellent model for resolving some of these problems. In this paper, we define the problems more fully and then describe our proposed collaborative model. We then conclude with a discussion about the challenges of implementing the model more permanently and at scale.

Lunch (participants on their own)

"Peachy Assignments:" A New Edu* Conference Component

David Bunde (Knox College)

Course assignments are integral to student learning HPC and also play an important role in student perceptions of the field. Instructors love to give exciting assignments that highlight important applications while emphasizing important principles and techniques. Unfortunately, creating great assignments is time-consuming and even our best efforts do not always succeed. This talk will introduce "Peachy Assignments", a new session for the Edu workshops. The idea is to showcase great assignments that are readily adoptable by other instructors. Come and learn about what makes a Peachy Assignment!

Keynote: Teaching Sound Principles and Good Practices for Parallel Algorithms.

Kunal Agrawal (Washington University in St. Louis)

Afternoon Break

Panel: Attracting Women and Underrepresented Minorities to HPC and Data Science

Martina Barnas (Indiana University), Trilce Estrada (University of New Mexico), Sally Goldman (Google)

Revisions to NSF/IEEE-TCPP Curriculum on Parallel and Distributed Computing (PDC) for Undergraduate Education - Updates on the Curriculum Revision and Audience Comments

Sushil K. Prasad (Georgia State University, National Science Foundation), Charles Weems (University of Massachusetts)

Room: 506

9:00 am - 5:30 pm

Energy Efficient Supercomputing (E2SC)

With exascale systems on the horizon, we have ushered in an era with power and energy consumption as the primary concerns for scalable computing. To achieve a viable exaflop high performance computing capability, revolutionary methods are required with a stronger integration among hardware features, system software and applications. Equally important are the capabilities for fine-grained spatial and temporal measurement and control to facilitate these layers for energy efficient computing across all layers. Current approaches for energy efficient computing rely heavily on power efficient hardware in isolation. However, it is pivotal for hardware to expose mechanisms for energy efficiency to optimize power and energy consumption for various workloads. At the same time, high fidelity measurement techniques, typically ignored in data-center level measurement, are of high importance for scalable and energy efficient inter-play in different layers of application, system software and hardware.

This workshop seeks to address the important energy efficiency aspects in the HPC community that have not been previously addressed by aspects covered in the data center or cloud computing communities. Emphasis is given to the application's view related to significant energy efficiency improvements and to the required hardware/software stack that must include necessary power and performance measurement and analysis harnesses. The workshop aims to bring together researchers from different communities working on challenging problems in this area for a dynamic exchange of ideas.

Room: 507

9:00 am - 5:30 pm

IA³ 2017 - 7th Workshop on Irregular Applications: Architectures and Algorithms

Introduction - IA³ 2017 - 7th Workshop on Irregular Applications: Architectures and Algorithms

Due to the heterogeneous data sets they process, data intensive applications employ a diverse set of methods and data structures. Consequently, they abound with irregular memory accesses, control flows, and communication patterns. Current supercomputing systems are organized around components optimized for data locality and bulk synchronous computations. Managing any form of irregularity on them demands substantial effort, and often leads to poor performance. Holistic solutions to address these challenges can emerge only by considering the problem from all perspectives: from micro- to system-architectures, from compilers to languages, from libraries to runtimes, from algorithm design to data characteristics. Strong collaborative efforts among researchers with different expertise, including domain experts and end users, could lead to significant breakthroughs. This workshop brings together scientists with these different backgrounds to discuss methods and technologies for efficiently supporting irregular

applications on current and future architectures.

A Taxonomy of HPDA Algorithms

Steve Conway (Hyperion Research)

This talk will be the first public presentation of the rationale, methodology and key findings of a late-2016, government-funded study Hyperion Research (then called IDC) conducted, entitled Developing a Real-World Taxonomy of the Underlying Mathematical, Algorithmic, and Technological Aspects of HPDA Applications. For this study, Hyperion first consulted public- and private-sector experts to develop a taxonomy matrix aimed at matching advanced analytics (HPDA) application types with users' preferred algorithms. Next came in-depth interviews with scientific, academic and industrial users to identify the hardware-software requirements of their applications and the attributes of the algorithms that generate those requirements. The final step was a report designed as a reference tool for HPDA users (including non-HPC specialists) from the broad spectrum of application domains investigated in the study.

Accelerating Energy Games Solvers on Modern Architectures

Andrea Formisano (University of Perugia)

Quantitative games, where quantitative objectives are defined on weighted game arenas, provide natural tools for designing faithful models of embedded controllers. Instances of these games are the so called Energy Games. Starting from a sequential baseline implementation, we investigate the use of massively data computation capabilities supported by modern GPUs to solve the initial credit problem for Energy Games.

We present different parallel implementations on multi-core CPU and GPU systems. Our solution outperforms the baseline implementation by up to 36x speedup and obtains a faster convergence time on real-world graphs.

Morning Break

Overcoming Load Imbalance for Irregular Sparse Matrices

Goran Flegar (Jaume I University)

In this paper we propose a load-balanced GPU kernel for computing the sparse matrix vector (SpMV) product. Making heavy use of the latest GPU programming features, we also enable satisfying performance for irregular and unbalanced matrices. In a performance comparison using 400 test matrices we reveal the new kernel being superior to the most popular SpMV implementations.

Progressive Load Balancing of Asynchronous Algorithms

Justs Zarins (University of Edinburgh)

Synchronisation in the presence of noise and hardware performance variability is a key challenge that prevents applications from scaling to large problems and machines. Using asynchronous or semi-synchronous algorithms can help overcome this issue, but at the cost of reduced stability or convergence rate. In this paper we propose progressive load balancing to manage progress imbalance in asynchronous algorithms dynamically. In our technique the balancing is done over time, not instantaneously.

Using Jacobi iterations as a test case, we show that, with CPU performance variability present, this approach leads to higher iteration rate and lower progress imbalance between parts of the solution space. We also show that under these conditions the balanced asynchronous method outperforms synchronous, semi-synchronous and totally asynchronous implementations in terms of time to solution.

Enabling Work-Efficiency for High Performance Vertex-Centric Graph Analytics on GPUs

Farzad Khorasani (Georgia Institute of Technology)

Massive parallel processing power of GPUs has attracted researchers to develop iterative vertex-centric graph processing frameworks for GPUs. Enabling work-efficiency in these solutions, however, is not straightforward and comes at the cost of SIMD-inefficiency and load imbalance. This paper offers techniques that overcome these challenges when processing the graph on a GPUs. For a SIMD-efficient kernel operation involving gathering of neighbors and performing reduction on them, we employ an effective task expansion strategy that avoids intra-warp thread underutilization. As recording vertex activeness requires additional data structures, to attenuate the graph storage overhead on limited GPU DRAM, we introduce vertex grouping as a technique that enables trade-off between memory consumption and the work efficiency in our solution. Our experiments show that these techniques provide up to 5.46x over the recently proposed WS-VR framework over multiple algorithms and inputs.

Parallel Depth-First Search for Directed Acyclic Graphs

Maxim Naumov (Nvidia Corporation)

Depth-First Search (DFS) is a pervasive algorithm, often used as a building block for topological sort, connectivity and planarity testing, among many other applications. We propose a novel work-efficient parallel algorithm for the DFS traversal of directed acyclic graph (DAG). The algorithm traverses the entire DAG in a BFS-like fashion no more than three times. As a result it finds the DFS pre-order (discovery) and post-order (finish time) as well as the parent relationship associated with every node in a DAG. We analyse the runtime and work complexity of this novel parallel algorithm. Also, we show that our algorithm is easy to implement and

optimize for performance. In particular, we show that its CUDA implementation on the GPU outperforms sequential DFS on the CPU by up to 6x in our experiments.

Optimizing Word2Vec Performance on Multicore Systems

Vasudevan Rengasamy (Pennsylvania State University)

The Skip-gram with negative sampling (SGNS) method of Word2Vec is an unsupervised approach to map words in a text corpus to low dimensional real vectors. The learned vectors capture semantic relationships between co-occurring words and can be used as inputs to many natural language processing and machine learning tasks. There are several high-performance implementations of the Word2Vec SGNS method. In this paper, we introduce a new optimization called context combining to further boost SGNS performance on multicore systems. For processing the One Billion Word benchmark dataset on a 16-core platform, we show that our approach is 3.53X faster than the original multithreaded Word2Vec implementation and 1.28X faster than a recent parallel Word2Vec implementation. We also show that our accuracy on benchmark queries is comparable to state-of-the-art implementations.

Spherical Region Queries on Multicore Architectures

Hao Lu (Oak Ridge National Laboratory)

In this short paper, we report the performance of two thread-parallel algorithms for spherical region queries on multicore architectures motivated by a challenging data analytics application in materials science. Performances of two tree-based algorithms and a naive algorithm are compared to identify the length scales at which these approaches perform optimally. The optimal algorithm is then used to scale the driver materials science application, which is shown to deliver over 17X speedup using 32 OpenMP threads on data sets containing many millions of atoms.

Lunch (participants on their own)

Quantum Computing and Irregular Applications

Fred Chong (University of Chicago)

Quantum computing is on the cusp of a revolution as prototypes with 100 quantum bits will soon appear and larger machines are on the horizon. The challenge will be to develop vertically-integrated systems in which programming languages, compilers, and runtime systems effectively map quantum algorithms to physical machines. In particular, I will discuss the opportunities and challenges for irregular applications on quantum machines. Quantum machines face challenges of communication and control which will be affected by irregularity. Physical connectivity faces scaling challenges and hierarchical structures may be needed. Yet, underneath these similarities with classical machines, quantum computers involve non-local interactions that make them uniquely different.

An Efficient Data Layout Transformation Algorithm for Locality-Aware Parallel Sparse FFT

Sunita Chandrasekaran (University of Delaware)

Fast Fourier Transform (FFT) is one of the most important numerical algorithms widely used in numerous scientific and engineering computations. With the emergence of big data problems, however, it is challenging to acquire, process and store a sufficient amount of data to compute the FFT in the first place. Recently developed sparse FFT (sFFT) algorithm provides a solution to this problem. sFFT computes a compressed Fourier transform by using only a small subset of the input data, thus achieving significant performance improvement.

While the increase in the number of cores and memory bandwidth on modern architectures provide an opportunity to improve the performance through sophisticated parallel algorithm design, sFFT is inherently complex, and numerous challenges need to be addressed. Among all the challenges, sFFT falls into the category of irregular applications in which memory access patterns are indirect and irregular that exhibit poor data locality. In this paper, we explore data layout transformation algorithms to tackle the challenge. Our approach shows that an optimized and locality-aware parallel sFFT can perform 7x faster than the original sequential sFFT library on a multicore platform. This optimized locality-aware parallel sFFT is also approximately 10x faster than the parallel FFTW.

Afternoon Break

A Case for Migrating Execution for Irregular Applications

Peter Kogge (University of Notre Dame)

Modern supercomputers have millions of cores, each capable of executing one or more threads of program execution. In these computers the site of execution for program threads rarely, if ever, changes from the node in which they were born. This paper discusses the advantages that may accrue when thread states migrate freely from node to node, especially when migration is managed by hardware without requiring software intervention. Emphasis is on supporting the growing classes of algorithms where there is significant sparsity, irregularity, and lack of locality in the memory reference patterns. Evidence is drawn from reformulation of several kernels into a migrating thread context approximating that of an emerging architecture with such capabilities.

Pressure-Driven Hardware Managed Thread Concurrency for Irregular Applications

John Leidel (Texas Tech University)

Given the increasing importance of efficient data intensive computing, we find that modern processor designs are not well suited to the irregular memory access patterns found in these algorithms. This research focuses on mapping the compiler's instruction cost scheduling logic to hardware managed concurrency controls in order to minimize pipeline stalls. In this manner, the hardware modules managing the low-latency thread concurrency can be directly understood by modern compilers.

We introduce a thread context switching method that is managed directly via a set of hardware-based mechanisms that are coupled to the compiler instruction scheduler. As individual instructions from a thread execute, their respective cost is accumulated into a control register. Once the register reaches a pre-determined saturation point, the thread is forced to context switch. We evaluate the performance benefits of our approach using a series of 24 benchmarks that exhibit performance acceleration of up to 14.6X.

Evaluation of Knight Landing High Bandwidth Memory for HPC Workloads

Yonghong Yan (University of South Carolina)

The Intel Knight Landing (KNL) manycore chip includes 3D-stacked memory named MCDRAM, also known as High Bandwidth Memory (HBM) for parallel applications that needs to scale to high thread count. In this paper, we provide a quantitative study of the KNL for HPC proxy applications including Lulesh, HPCG, AMG, and Hotspot when using DDR4 and MCDRAM. The results indicate that HBM significantly improves the performance of memory intensive applications for as many as three times better than DDR4 in HPCG, and Lulesh and HPCG for as many as 40% and 200%. For the selected compute intensive applications, the performance advantage of MCDRAM over DDR4 varies from 2% to 28%. We also observed that the cross-points, where MCDRAM starts outperforming DDR4, are around 8 to 16 threads.

IA³ Debate

Ruud Van Der Pas (Oracle), Franz Franchetti (Carnegie Mellon University), Eric Van Hensbergen (ARM Ltd), Richard Vuduc (Georgia Institute of Technology)

The panelists will debate the proposition:

"Specialized, perhaps configurable, hardware and software are necessary to achieve high- performance, scalable data analytics"

Room: 601

9:00 am - 5:30 pm

Joint International Workshop on Parallel Data Storage and Data Intensive Scalable Computing Systems (PDSW-DISCS)

The Joint International Workshop on Parallel Data Storage and Data Intensive Scalable Computing Systems (PDSW-DISCS) is a merger of two successful SC workshop communities, the Parallel Data Storage Workshop (PDSW) and Data Intensive Scalable Computing Systems (DISCS). Our goal is to better promote and stimulate researchers' interactions and to better address some of most critical challenges for scientific data storage, management, devices, and processing infrastructure for both traditional compute intensive simulations as well as data-intensive high performance computing solutions. Novel submitted papers are peer-reviewed in the late summer, about 8-12 selected for presentation at the full day SC workshop, published in either the ACM or IEEE Digital Library and shepherded for proposal to a journal special issue. The workshop web site can be reached through www.pdsw.org/

Room: 603

9:00 am - 5:30 pm

HPC Systems Professionals Workshop

In order to meet the demands of high performance computing (HPC) researchers, large-scale computational and storage machines require many staff members who design, install, and maintain these systems. These HPC systems professionals include system engineers, system administrators, network administrators, storage administrators and operations staff who face problems that are unique to high performance computing systems. While many conferences exist for the HPC field and the system administration field, none exist that focus on the needs of HPC systems professionals. Support resources can be difficult to find to help with the issues encountered in this specialized field. Often times, systems staff turn to the community as a support resource, and opportunities to strengthen and grow those relationships are highly beneficial. This workshop is designed to share solutions to common problems, provide a platform to discuss upcoming technologies, and present state of the practice techniques so that HPC centers will get a better return on their investment, increase performance and reliability of systems, and researchers will be more productive.

Room: 605

9:00 am - 5:30 pm

2nd International Workshop on Post Moore's Era Supercomputing (PMES)

This interdisciplinary workshop is organized to explore the scientific issues, challenges, and opportunities for supercomputing beyond the scaling limits of Moore's Law, with the ultimate goal of keeping supercomputing at the forefront of computing technologies beyond the physical and conceptual limits of current systems. Moore's Law—the doubling the number of transistors in a chip every two years—has so far contributed to every aspect of supercomputing system architectures, including GPU and many-core accelerators, large on-chip caches, integrated special purpose hardware, and increasing memory capacities. However, it is now well accepted that current approaches will reach their limits in next decade due to the confluence of several limitations including both fundamental physics and economics. Although device and manufacturing technologies continue to make progress, most experts predict that CMOS transistor shrinking may stop at around 2025 to 2030 due to these limits. Nevertheless, continuing the progress of supercomputing beyond the scaling limits of Moore's Law is likely to require a comprehensive re-thinking of technologies, ranging from innovative materials and devices, circuits, system architectures, programming systems, system software, and applications. In this regard, the goal of the workshop is to explore the technological directions of supercomputing to prepare for this "Post Moore's Law" era by fostering interdisciplinary dialog across the spectrum of stakeholders: applications, algorithms, software, and hardware. Experts from academia, government, and industry in the fields of computational science, mathematics, engineering, and computer science will participate in the workshop as invited speakers, position papers, and panelists.

Room: 607

9:00 am - 5:30 pm

8th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems

Introduction - 8th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems

Vassil Alexandrov (Barcelona Supercomputing Center)

Novel scalable scientific algorithms are needed in order to enable key science applications to exploit the computational power of large-scale systems. This is especially true for the current tier of leading petascale machines and the road to exascale computing as HPC systems continue to scale up in compute node and processor core count. These extreme-scale systems require novel scientific algorithms to hide network and memory latency, have very high computation/communication overlap, have minimal communication, and have no synchronization points. With the advent of Big Data in the past few years, the need of such scalable mathematical methods and algorithms able to handle data and compute intensive applications at scale becomes even more important.

Scientific algorithms for multi-petaflop and exaflop systems also need to be fault tolerant and fault resilient, since the probability of faults increases with scale. Resilience at the system software and at the algorithmic level is needed as a crosscutting effort. Finally, with the advent of heterogeneous compute nodes that employ standard processors as well as GPGPUs, scientific algorithms need to match these architectures to extract the most performance. This includes different system-specific levels of parallelism as well as co-scheduling of computation. Key science applications require novel mathematics and mathematical models and system software that address the scalability and resilience challenges of current- and future-generation extreme-scale HPC systems.

Keynote - Application Development Framework for Manycore Architectures on Post-Peta/Exascale Systems

Kengo Nakajima (University of Tokyo)

"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 manycore architectures. Source code developed on a PC with a single processor is linked and the parallel code generated is optimized for post-petascale systems with manycore architectures, such as the Oakforest-PACS system. "ppOpen-HPC" is part of a five-year project spawned by the "Development of System Software Technologies for Post-PetaScale 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 technology enables automatic generation of optimized libraries and applications under various types of environments. We release the most updated version of ppOpen-HPC as open source software every year in November at <http://ppopenhpc.cc.u-tokyo.ac.jp/ppopenhpc/>. In 2016, the team of ppOpen-HPC joined ESSEX-II (Equipping Sparse Solvers for Exascale) project, which is funded by JST-CREST and the German DFG priority program 1648 "Software for Exascale Computing" (SPPEXA) under Japan-Germany collaboration. 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 with performance model. In the presentation, various types of achievements of ppOpen-HPC, ESSEX-II, and pK-OpenHPC project, such as applications using HACAPK library for H-matrix computation, and parallel preconditioned iterative solvers will be shown.

Dynamic Task Discovery in ParSEC- A Data-Flow Task-Based Runtime

Reazul Hoque (University of Tennessee)

Successfully exploiting distributed collections of heterogeneous many-core architectures with complex memory hierarchy through a portable programming model is a challenge for application developers. The literature is not short of proposals addressing this problem, including many evolutionary solutions that seek to extend the capabilities of current message passing paradigms with

intranode features (MPI+X). A different, more revolutionary, solution explores data-flow task-based runtime systems as a substitute to both local and distributed data dependencies management. The solution explored in this paper, PaRSEC, is based on such a programming paradigm, supported by a highly efficient task-based runtime. This paper compares two programming paradigms present in PaRSEC, Parameterized Task Graph (PTG) and Dynamic Task Discovery (DTD) in terms of capabilities, overhead, and potential benefits.

Break

Keynote - Breakthrough Science at the Exascale

Katherine Yelick (Lawrence Berkeley National Laboratory)

In the next few years, exascale computing systems will become available to the scientific community. They will require new levels of parallelization, new models of memory and storage, and a variety of node architectures for processors and accelerators. They will also provide simulation capabilities with unprecedented scale and complexity across many fields of science, helping to answer fundamental science questions in cosmology, astrophysics, and nuclear science; understand and address issues in climate change, the environment and seismic safety; and aid in the design of advanced materials, manufacturing, energy systems, and pharmaceuticals. These systems may also offer exascale data analysis capabilities, allowing genomics, images, and sensor data to be processed, analyzed, and modeled using comparisons with simulations, deep learning algorithms or other machine learning methods. In this talk, I will present some of the exciting science opportunities, the need for advanced in algorithms and mathematics to advance along with the system performance, and how the variety of workloads will stress the different aspects of exascale hardware and software systems.

Flexible Batched Sparse Matrix-Vector Product on GPUs

Hartwig Anzt (University of Tennessee, Karlsruhe Institute of Technology)

We propose a variety of batched routines for concurrently processing a large collection of small-size, independent sparse matrix vector products (SpMV) on graphics processing units (GPUs). These batched SpMV kernels are designed to be flexible in order to handle a batch of matrices which differ in size, nonzero count, and nonzero distribution. Furthermore, they support three most commonly used sparse storage formats: CSR, COO and ELL. Our experimental results on a state-of-the-art GPU reveal performance improvements of up to 25× compared to non-batched SpMV routines.

Snowpack: Efficient Parameter Choice for GPU Kernels via Static Analysis and Statistical Prediction

Ignacio Laguna (Lawrence Livermore National Laboratory)

The running time of GPU kernels depends on an invocation parameter, the number of threads in each thread block. Sometime the dependence is quite strong leading to 50-100% change in execution time for long-running kernels. Until now, it has been an art form to decide on the optimal setting for this parameter. Nvidia provides a tool for CUDA kernels, called OCC, that guides a developer toward this goal. In this paper, we show that OCC maximizes occupancy of GPU cores but does not meet the performance goal in a wide class of applications. We develop a solution called Snowpack that uses static features in a statistical learning framework to choose the optimal block size parameter. It does this without needing to execute the kernel multiple times, as a possible alternate solution Autotuner does. We evaluate our solution, Snowpack, on 89 kernels of 10 applications.

Leveraging NVLINK and Asynchronous Data Transfer to Scale Beyond the Memory Capacity of GPUs

David Appelhans (IBM)

In this paper we demonstrate the utility of fast GPU to CPU interconnects to weak scale on hierarchical nodes without being limited to problem sizes that fit only in the GPU memory capacity. We show the speedup possible for a new regime of algorithms which traditionally have not benefited from being ported to GPUs because of an insufficient amount of computational work relative to bytes of data that must be transferred (offload intensity). This new capability is demonstrated with an example of our hierarchical GPU port of UMT, the 51K line CORAL benchmark application for Lawrence Livermore National Lab's radiation transport code. By overlapping data transfers and using the NVLINK connection between IBM POWER 8 CPUs and NVIDIA P100 GPUs, we demonstrate a speedup that continues even when scaling the problem size well beyond the memory capacity of the GPUs.

Application of a Communication-Avoiding Generalized Minimal Residual Method to a Gyrokinetic Five Dimensional Eulerian Code on ManyCore Platforms

Yasuhiro Idomura (Japan Atomic Energy Agency)

A communication-avoiding generalized minimal residual (CA-GMRES) method is applied to the gyrokinetic toroidal five dimensional Eulerian code GT5D, and its performance is compared against the original code with a generalized conjugate residual (GCR) method on the JAEA ICEX (Haswell), the Plasma Simulator (FX100), and the Oakforest-PACS (KNL). Although the CA-GMRES method dramatically reduces the number of data reduction communications, computation is largely increased compared with the GCR method. To resolve this issue, 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. The modified CA-GMRES method has ~3.8x higher arithmetic intensity than the GCR method, and thus, is suitable for future Exa-scale architectures with limited memory and network bandwidths. The CA-GMRES solver is implemented using a hybrid CA approach, in which we apply CA to data reduction communications and use communication overlap for halo data communications, and is highly optimized for distributed caches on KNL. It is shown that compared with the GCR solver, its computing kernels are accelerated by 1.47x ~ 2.39x, and the cost of data reduction communication is reduced from 5% ~ 13% to ~1% of the total cost at 1,280 nodes.

Parallel Jaccard and Related Graph Clustering Techniques

Alexandre Fender (Nvidia Corporation, University of Versailles)

We propose to generalize Jaccard and related measures, often used as similarity coefficients between two sets. We define Jaccard, Dice-Sorensen and Tversky edge weights on a graph and generalize them to account for vertex weights. We develop an efficient parallel algorithm for computing Jaccard edge and PageRank vertex weights. We highlight that the weights computation can obtain more than 10x speedup on the GPU versus CPU on large realistic data sets. Also, we show that finding a minimum balanced cut for modified weights can be related to minimizing the sum of ratios of the intersection and union of nodes on the boundary of clusters. Finally, we show that the novel weights can improve the quality of the graph clustering by about 15% and 80% for multi-level and spectral graph partitioning and clustering schemes, respectively.

Lunch (participants on their own)

Keynote - An Overview of High Performance Computing and Challenges for the Future

Jack Dongarra (University of Tennessee)

In this talk, we will look at the current state of high performance computing and look to the future toward exascale. In addition, we will examine some issues that can help in reducing the power consumption for linear algebra computations.

Investigating Half-Precision Arithmetic to Accelerate Dense Linear System Solvers

Azzam Haidar (University of Tennessee)

The use of low-precision arithmetic in mixed-precision computing methods has been a powerful tool to accelerate numerous scientific computing applications. Artificial intelligence (AI), in particular, has pushed this to current extremes, making use of half-precision floating-point arithmetic (FP16) in approaches based on neural networks. The appeal of FP16 is in the high performance that can be achieved using it on today's powerful manycore GPU accelerators, e.g., like the NVIDIA V100, that can provide 120 TeraFLOPS alone in FP16. We present an investigation showing that other HPC applications can harness this power too, and, in particular, the general HPC problem of solving $(A x = b)$, where A is a large dense matrix, and the solution is needed in FP32 or FP64 accuracy. Our approach is based on the mixed-precision iterative refinement technique -- we generalize and extend prior advances into a framework, for which we develop architecture-specific algorithms and highly-tuned implementations that resolve the main computational challenges of efficiently parallelizing, scaling, and using FP16 arithmetic in the approach on high-end GPUs. Subsequently, we show for the first time how the use of FP16 arithmetic can significantly accelerate, as well as make more energy efficient, FP32 or FP64-precision $(A x = b)$ solvers. Our results are reproducible and the developments will be made available through the MAGMA library. We quantify in practice the performance and limitations of the approach.

Break

Keynote - A Holistic Approach to Advancing Science and Engineering through Extreme-Scale Computing

Michael A. Heroux (Sandia National Laboratories)

Extreme-scale computing must ultimately provide a means to an end: enabling scientific discovery, and advancing engineering design. The broadest and deepest impact of extreme-scale computing requires an understanding of its role in this larger context. In this talk, we provide a description of the ecosystem to which extreme-scale computing belongs, and the current challenges of multi-physics and multi-scale simulations, the importance of resilience, the possible role of machine learning, the increased need for developer productivity and software sustainability, and the overarching concerns of trustworthiness and reproducibility. From this holistic perspective, we present the specific challenges and opportunities for algorithms research and development.

A Highly Scalable, Algorithm-Based Fault-Tolerant Solver for Gyrokinetic Plasma Simulations

Michael Obersteiner (Technical University Munich)

With future exascale computers expected to have millions of compute units distributed among thousands of nodes, system faults are predicted to become more frequent. Fault tolerance will thus play a key role in HPC at this scale. In this presentation, we focus on solving the 5-dimensional gyrokinetic Vlasov-Maxwell equations using the application code GENE as it represents a high-dimensional and resource-intensive problem which is a natural candidate for exascale computing. We discuss the Fault-Tolerant Combination Technique, a resilient version of the Combination Technique, a method to increase the discretization resolution of existing PDE solvers. For the first time, we present an efficient, scalable and fault-tolerant implementation of this algorithm for plasma physics simulations based on a manager-worker model and test it under very realistic and pessimistic environments with simulated faults. We show that the Fault-Tolerant Combination Technique -- an algorithm-based forward recovery method -- can tolerate a large number of faults with a low overhead and at an acceptable loss in accuracy. Our parallel experiments with up to 32k cores show good scalability at a relative parallel efficiency of 93.61%. We conclude that algorithm-based solutions to fault tolerance are attractive for this type of problem.

Analyzing the Criticality of Transient Faults-Induced SDCs on GPU Applications

Paolo Rech (Federal University of Rio Grande do Sul)

In this paper, we compare the soft-error sensitivity of parallel applications on modern GPUs obtained through architectural-level fault injections and high-energy particle beam radiation experiments. Fault-injection and beam experiments provide different information and use different transient-fault sensitivity metrics, which are hard to combine. In this presentation, we show how correlating beam and fault-injection data can provide a deeper understanding of the behavior of GPUs in the occurrence of transient faults. In particular, we demonstrate that commonly used architecture-level fault models (and fast injection tools) can be used to identify critical kernels and to associate some experimentally observed output errors with their causes. Additionally, we show how register file and instruction-level injections can be used to evaluate ECC efficiency in reducing the radiation-induced error rate.

Dynamic Load Balancing of Massively Parallel Unstructured Meshes

Gerrett Diamond (Rensselaer Polytechnic Institute)

Simulating systems with evolving relational structures on massively parallel computers require the computational work to be evenly distributed across the processing resources throughout the simulation. Adaptive, unstructured, mesh-based finite element and finite volume tools best exemplify this need. We present EnGPar and its diffusive partition improvement method that accounts for multiple application specified criteria. EnGPar's performance is compared against its predecessor, ParMA. Specifically, partition improvement results are provided on up to 512K processes of the Argonne Leadership Computing Facility's Mira BlueGene/Q system.

Invited Talk - On Improved Monte Carlo Hybrid Methods for Preconditioner Computations

Vassil Alexandrov (Barcelona Supercomputing Center)

In this talk, we present certain improvements of the Markov Chain Monte Carlo Matrix Inversion and show their impact on performance and scalability of the method. The method is used for the computation of preconditioners for iterative methods, such as generalized minimal residuals (GMRES) or bi-conjugate gradient stabilized (BICGstab), for the solution of linear systems of equations. The problem of communication overhead is addressed via a modification of the method and via optimizations in the communication patterns of a chosen implementation of the method. Numerical experiments are carried out to highlight the benefits and deficiencies of both approaches and to assess their overall usefulness in light of scalability of the method.

Room: 702

9:00 am - 5:30 pm

PAW 2017: The 2nd Annual PGAS Applications Workshop

Introduction - PAW 2017: The 2nd Annual PGAS Applications Workshop

In order to manage the complexity of emerging architectures and technologies, parallel programmers require effective approaches for developing large-scale applications that fully exploit these transformative features. Ideally, such approaches should elevate the programmer's expression of the computation from mechanism-specific abstractions (messages, kernels) to higher-level abstractions, similar to how Fortran and C permitted programmers to think in terms of variables rather than registers. The Partitioned Global Address Space (PGAS) family of programming models provides such an approach. It simplifies the complexity of parallel programming across distinct memories by making it more similar to shared-memory computation, while enhancing scalability by exposing data locality through its semantics. This workshop will bring together applications experts who will present concrete practical examples to illustrate the benefits of exploiting PGAS languages including Fortran, Unified Parallel C (UPC), X10, and Chapel as well as libraries such as Unified Parallel C++ (UPC++), Coarray C++, OpenSHMEM, MPI-3, and Global Arrays.

Keynote: Shared Memory HPC Programming: Past, Present and Future

Break

Cosmological Particle-Mesh Simulations in Chapel

Nikhil Padmanabhan (Yale University)

This presentation will describe the implementation and performance of a gravitational N-body particle-mesh code in Chapel. Our goal here is to go beyond simple benchmarks and kernels, and present a case-study on the readiness of Chapel to be used for end-to-end computations.

Graph500 on OpenSHMEM: Using a Practical Survey of Past Work to Motivate Novel Algorithmic Developments

Max Grossman (Rice University)

Graph500 is an open specification of a graph-based benchmark for high-performance computing (HPC). The core computational kernel of Graph500 is a breadth-first search of an undirected graph. Unlike many other HPC benchmarks, Graph500 is therefore characterized by heavily irregular and fine-grain computation, memory accesses, and network communication. Therefore, it can serve as a more realistic stress test of modern HPC hardware, software, and algorithmic techniques than other benchmarking

efforts.

On the other hand, OpenSHMEM is an open, PGAS, and SPMD specification of a communication model for communicating across large numbers of processing elements. OpenSHMEM explicitly focuses on applications characterized by fine-grain communication, of which Graph500 is one example.

Therefore, there is a natural synergy between the communication patterns of Graph500 and the capabilities of OpenSHMEM. In this work we explore that synergy by developing several novel implementations of Graph500 on various OpenSHMEM implementations. We contribute a review of the state-of-the-art in distributed Graph500 implementations, as well as a performance and programmability comparison between the state-of-the-art and our own OpenSHMEM-based implementations. Our results demonstrate improved scaling of Graph500's BFS kernel out to 1,024 nodes of the Edison supercomputer, achieving 2.5x performance improvement relative to the highest performing reference implementation at that scale.

Preliminary Performance Evaluation of Coarray-based Implementation of Fiber Miniapp Suite Using XcalableMP PGAS Language

Hitoshi Murai (RIKEN)

XcalableMP (XMP) is a PGAS language that is defined by the XMP Specification Working Group of the PC Cluster Consortium. This paper provides the implementation and evaluation of the Fiber miniapp suite, which is maintained mainly by RIKEN AICS, on the basis of the local-view parallelization model using the coarray feature of XMP. In many cases, a coarray-based implementation can be obtained by replacing original MPI functions with coarray assignment statements. For irregular applications, we show a method to rewrite it into coarray-based one in this paper. The evaluation on the K computer using the Omni XcalableMP compiler we are developing showed that some of the XMP implementations are comparable to the original ones but there are performance degradations with the others, which is due to a large overhead of allocating dynamic coarrays at runtime.

Performance Portability of an Intermediate-Complexity Atmospheric Research Model in Coarray Fortran

Damian Rouson (Sourcery Institute)

We present results on the scalability and performance of an open-source, Coarray Fortran (CAF) mini-application (mini-app) that solves several parallel, numerical algorithms known to dominate the execution of the Intermediate Complexity Atmospheric Research (ICAR) model developed at the National Center for Atmospheric Research (NCAR). The solver employs standard Fortran 2008 features and includes several Fortran 2008 implementations of the collective subroutines that are defined in the Committee Draft the upcoming Fortran 2015 standard. The ability of CAF to run atop various communication layers and the increasing compiler support for CAF facilitated initial evaluations of several compiler/runtime/hardware combinations. Results are presented for the GNU, Intel, and Cray compilers, each of which offers different parallel runtime libraries employing one or more communication layers, including MPI, OpenSHMEM, and proprietary alternatives. We studied the performance on both multi- and many-core processors running on distributed-memory systems. The results of our initial investigations suggest promising scaling behavior across a range of hardware, compiler, and runtime choices on platforms ranging up to 100,000 cores.

Lunch (participants on their own)

Keynote: How Does PGAS Collaborate with MPI+X?

Mitsuhsa Sato (RIKEN)

While MPI is commonly used as a programming model between nodes for large-scale distributed memory systems, new programming models such as PGAS are emerging for many core and new communication models. PGAS is expected to be one of the programming models for exascale computing due to light-weight one-sided communication and low overhead synchronization semantics. While it is reported that PGAS can take advantage of its programming model in some applications, in reality, PGAS programming systems have several problems in performance and run-time design as a general-purpose communication layer. And, only one-side communication by PGAS is not sufficient to implement whole parallel applications so that it needs to cooperate with high-level global operations such as MPI collective communications. We are developing a PGAS programming language XcalableMP including coarray extension. In this talk, I will discuss the present status of PGAS and its future direction with respect to current dominant programming model MPI+X, and our experience of PGAS application development will be also addressed.

Break

PGAS Applications Workshop Panel

Katherine Yelick (Lawrence Berkeley National Laboratory), William Carlson (Institute for Defense Analyses), Bradford Chamberlain (Cray Inc), Damian Rouson (Sourcery Institute), Mitsuhsa Sato (RIKEN), Anton Shterenlikht (University of Bristol), Lauren Smith (US Department of Defense)

This panel, moderated by Katherine Yelick, will feature the keynote speakers from the PGAS Applications Workshop and other proponents of Partitioned Global Address Space languages and programming models.

Free Discussion

Room: 704-706
9:00 am - 5:30 pm

The 8th International Workshop on Performance Modeling, Benchmarking, and Simulation of High Performance Computer Systems (PMBS17)

Introduction - The 8th International Workshop on Performance Modeling, Benchmarking, and Simulation of High Performance Computer Systems (PMBS17)

Simon D. Hammond (Sandia National Laboratories)

The PMBS17 workshop is concerned with the comparison of high-performance computing systems through performance modeling, benchmarking, or through the use of tools such as simulators. We are particularly interested in research which reports the ability to measure and make tradeoffs in software/hardware co-design to improve sustained application performance. We are also keen to capture the assessment of future systems, for example, through work that ensures continued application scalability through peta- and exa-scale systems.

The aim of this workshop is to bring together researchers, from industry and academia, concerned with the qualitative and quantitative evaluation and modeling of high-performance computing systems. Authors are invited to submit novel research in all areas of performance modeling, benchmarking and simulation, and we welcome research that brings together current theory and practice. We recognize that the coverage of the term 'performance' has broadened to include power consumption and reliability, and that performance modeling is practiced through analytical methods and approaches based on software tools and simulators.

Performance modeling, benchmarking, and simulation will underpin software and hardware design choices as we advance toward the exascale era. This workshop continues to attract high quality input from industry, government and academia.

Evaluating On-Node GPU Interconnects for Deep Learning Workloads

Nathan Tallent (Pacific Northwest National Laboratory)

Scaling deep learning workloads across multiple GPUs on a single node has become increasingly important in data analytics. A key question is how well a PCIe-based GPU interconnect can perform relative to a custom high-performance interconnect such as NVIDIA's NVLink. This paper evaluates two such on-node interconnects for eight NVIDIA Pascal P100 GPUs: (a) the NVIDIA DGX-1's NVLink 1.0 'hybrid cube mesh'; and (b) the Cirrascale GX8's two-level PCIe tree using dual SR3615 switch risers. To show the effects of a range of neural network workloads, we define a parameterized version of the popular ResNet architecture. We define a workload intensity metric that characterizes the expected computation/communication ratio; we also locate AlexNet and GoogLeNet within that space. As expected, the DGX-1 typically has superior performance. However, the GX8 is very competitive on all ResNet workloads. With 8 GPUs, the GX8 can outperform the DGX-1 on all-to-all reductions by 10% for medium-sized payloads; and in rare cases, the GX8 slightly outperforms on ResNet.

Modeling UGAL on the Dragonfly Topology

Md Atiqul Mollah (Florida State University)

The Dragonfly topology has been proposed and deployed as the interconnection network topology for next generation supercomputers. Practical routing algorithms developed for Dragonfly are based on a routing scheme called Universal Globally Adaptive Load-balanced routing with Global information (UGAL-G). While UGAL-G and UGAL-based practical routing schemes have been extensively studied, all existing results are based on simulation or measurement. There is no theoretical understanding of how the UGAL-based routing schemes achieve their performance on a particular network configuration as well as what the routing schemes optimize for. In this work, we develop and validate throughput models for UGAL-G on the Dragonfly topology, and identify a robust model that is both accurate and efficient across many Dragonfly variations. Given a traffic pattern, the proposed models estimate the aggregate throughput for the pattern accurately and effectively. Our results not only provide a mechanism to predict the communication performance for large scale Dragonfly networks but also reveal the inner working of UGAL-G, which furthers our understanding of the UGAL-based routing on Dragonfly.

Morning Break

A Performance Study of Quantum ESPRESSO's PWscf Code on Multi-Core and GPU Systems

Joshua Romero (Nvidia Corporation)

We describe the porting of PWscf (Plane-Wave Self Consistent Field), a key component of the Quantum ESPRESSO open-source suite of codes for materials modeling, to GPU systems using CUDA Fortran. Kernel loop directives (CUF kernels) have been extensively used in order to have a single source code for both CPU and GPU implementations. The results of the GPU version have been carefully validated and the performance of the code on several GPU systems (both x86 and POWER8 based) has been compared with traditional Intel multi-core (CPU only) systems. This current GPU version can reduce the time-to-solution by an average factor of 2-3 running two different input cases widely used as benchmarks on small and large high performance computing systems.

Comparison of Parallelization Approaches, Languages, and Compilers for Unstructured Mesh Algorithms on GPUs *Gabor Daniel Balogh (Pazmany Peter Catholic University)*

Efficiently exploiting GPUs is increasingly essential in scientific computing, as many current and upcoming supercomputers are built using them. To facilitate this, there are a number of programming approaches, such as CUDA, OpenACC and OpenMP 4, supporting different programming languages (mainly C/C++ and Fortran). There are also several compiler suites (clang, nvcc, PGI, XL) each supporting different combinations of languages. In this study, we take a detailed look at some of the currently available options, and carry out a comprehensive analysis and comparison using computational loops and applications from the domain of unstructured mesh computations. Beyond runtimes and performance metrics (GB/s), we explore factors that influence performance such as register counts, occupancy, usage of different memory types, instruction counts, and algorithmic differences. Results of this work show how clang's CUDA compiler frequently outperform NVIDIA's nvcc, performance issues with directive-based approaches on complex kernels, and OpenMP 4 support maturing in clang and XL; currently around 10% slower than CUDA.

Path-Synchronous Performance Monitoring in HPC Interconnection Networks with Source-Code Attribution *Adarsh Yoga (Rutgers University, Hewlett Packard Enterprise)*

Performance anomalies involving interconnection networks have largely remained a “black box” for developers relying on traditional CPU profilers. Network-side profilers collect aggregate statistics and lack source-code attribution. We have incorporated an effective protocol extension in the Gen-Z communication protocol for tagging network packets in an interconnection network; additionally, we have backed the protocol extension with hardware and software enhancements that allow tracking the flow of a network transaction through every hop in the interconnection network and associate it back to the application source code. The result is a first-of-its-kind hardware-assisted telemetry of disparate, autonomous interconnection networking components with application source code association that offers better developer insights. Our scheme works on a sampling basis to ensure low runtime overhead and generates modest volumes of data. Simulation of our methods in the open-source Structural Simulation Toolkit (SST/Macro) shows its effectiveness - deep insights into the underlying network details to the developer at minimal overheads.

Performance and Energy Usage of Workloads on KNL and Haswell Architectures *Tyler Allen (Clemson University)*

Manycore architectures are an energy-efficient step towards exascale computing within a constrained power budget. The Intel Knights Landing (KNL) manycore chip is a specific example of this and has seen early adoption by a number of HPC facilities. It is therefore important to understand the performance and energy usage characteristics of KNL. In this paper, we evaluate the performance and energy efficiency of KNL in contrast to the Xeon (Haswell) architecture for applications representative of the workload of users at NERSC. We consider the optimal MPI/OpenMP configuration of each application and use the results to characterize KNL in contrast to Haswell. As well as traditional DDR memory, KNL contains MCDRAM and we also evaluate its efficacy. Our results show that, averaged over our benchmarks, KNL is 1.84x more energy efficient than Haswell and has 1.27x greater performance.

A Survey of Application Memory Usage on a National Supercomputer: An Analysis of Memory Requirements on ARCHER *Andrew Turner (University of Edinburgh), Simon McIntosh-Smith (University of Bristol)*

In this presentation, we set out to provide a set of modern data on the actual memory per core and memory per node requirements of the most heavily used applications on a contemporary, national-scale supercomputer. This report is based on data from all jobs run on the UK national supercomputing service, ARCHER, a 118,000 core Cray XC30, in the 1 year period from 1st July 2016 to 30th June 2017 inclusive. Our analysis shows that 80% of all usage on ARCHER has a maximum memory use of 1 GiB/core or less (24 GiB/node or less) and that there is a trend to larger memory use as job size increases. Analysis of memory use by software application type reveals differences in memory use between periodic electronic structure, atomistic N-body, grid-based climate modelling, and grid-based CFD applications. We present an analysis of these differences, and suggest further analysis and work in this area. Finally, we discuss the implications of these results for the design of future HPC systems, in particular the applicability of high bandwidth memory type technologies.

Lunch (participants on their own)

Resilient N-Body Tree Computations with Algorithm-Based Focused Recovery: Model and Performance Analysis *Aurelien Cavelan (University of Basel)*

This presentation presents a model and performance study for Algorithm-Based Focused Recovery (ABFR) applied to N-body computations, subject to latent errors. We make a detailed comparison with the classical Checkpoint/Restart (CR) approach. While the model applies to general frameworks, the performance study is limited to perfect binary trees, due to the inherent difficulty of the analysis. With ABFR, the crucial parameter is the detection interval, which bounds the error latency. We show that the detection interval has a dramatic impact on the overhead, and that optimally choosing its value leads to significant gains over the CR approach.

Multi-Fidelity Surrogate Modeling for Application/Architecture Co-Design *Aravind Neelakantan (University of Florida)*

The HPC community has been using abstract, representative applications and architecture models to enable faster co-design cycles. While developers often qualitatively verify the correlation of the app abstractions to the parent application, it is equally

important to quantify this correlation to understand how the co-design results translate to the parent application. In this paper, we propose a multi-fidelity surrogate (MFS) approach which combines data samples of low-fidelity (LF) models (representative apps and architecture simulation) with a few samples of a high-fidelity (HF) model (parent app). The application of MFS is demonstrated using a multi-physics simulation application and its proxy-app, skeleton-app, and simulation models. Our results show that RMSE between predictions of MFS and the baseline HF models was 4%, which is significantly better than using either LF or HF data alone, demonstrating that MFS is a promising approach for predicting the parent application performance while staying within a computational budget.

Break

Modeling Large Compute Nodes with Heterogeneous Memories with the Cache-Aware Roofline Model

Nicolas Denoyelle (French Institute for Research in Computer Science and Automation (INRIA))

In order to fulfill modern applications needs, computing systems become more powerful, heterogeneous, and complex. NUMA platforms and emerging high bandwidth memories offer new opportunities for performance improvements. However, they also increase hardware and software complexity, thus making application performance analysis and optimization an even harder task. The Cache-Aware Roofline Model (CARM) is an insightful, yet simple model designed to address this issue. It provides feedback on potential applications bottlenecks and shows how far is the application performance from the achievable hardware upper-bounds. However, it does not encompass NUMA systems and next generation processors with heterogeneous memories. Yet, some application bottlenecks belong to those memory subsystems and would benefit from the CARM insights. In this presentation, we fill the missing requirements to scope recent large shared memory systems with the CARM. We provide the methodology to instantiate, and validate the model on a NUMA system as well as on the latest Xeon Phi processor equipped with configurable hybrid memory. Finally, we show the model ability to exhibits several bottlenecks of such systems, which were not supported by CARM.

A Scalable Analytical Memory Model for CPU Performance Prediction

Gopinath Chennupati (Los Alamos National Laboratory)

As the US Department of Energy (DOE) invests in exascale computing, performance modeling of physics codes on CPUs remain a challenge in computational co-design due to the complex design of processors including memory hierarchies, instruction pipelining, and speculative execution. We present Analytical Memory Model (AMM), a model of cache memory hierarchy, embedded in the Performance Prediction Toolkit (PPT) -- a suite of discrete-event-simulation-based co-design hardware and software models. AMM enables PPT to significantly improve the quality of its runtime predictions of scientific codes.

AMM uses a computationally efficient, stochastic method to predict the reuse distance profiles of codes, where reuse distance is a hardware architecture-independent measure of the patterns of virtual memory accesses. AMM relies on a stochastic, static basic block-level analysis of reuse profiles measured from the memory traces of applications on small instances. The analytical reuse distribution is useful to estimate the effective latency and throughput of memory access, which in turn are used to predict the overall runtime of a scientific application.

Our experimental results demonstrate the scalability of AMM, where the predicted and actual runtimes of three scientific mini-applications are similar.

A Slurm Simulator: Implementation and Parametric Analysis

Nikolay A. Simakov (University at Buffalo)

Slurm is an open-source resource manager for HPC that provides high configurability for inhomogeneous resources and job scheduling. Various Slurm parametric settings can significantly influence HPC resource utilization and job wait time, however in many cases it is hard to judge how these options will affect the overall HPC resource performance. The Slurm simulator can be a very helpful tool to aid parameter selection for a particular HPC resource. Here, we report our implementation of a Slurm simulator and the impact of parameter choice on HPC resource performance. The simulator is based on a real Slurm instance with modifications to allow simulation of historical jobs and to improve the simulation speed. The simulator speed heavily depends on job composition, HPC resource size and Slurm configuration. For an 8000 cores heterogeneous cluster, we achieve about 100 times acceleration, e.g. 20 days can be simulated in 5 hours.

Periodic I/O Scheduling for Supercomputers

Guillaume Aupy (French Institute for Research in Computer Science and Automation (INRIA))

With the ever-growing need of data in HPC applications, the congestion at the I/O level becomes critical in supercomputers. Architectural enhancement such as burst-buffers and pre-fetching are added to machines, but are not sufficient to prevent congestion. Recent online I/O scheduling strategies have been put in place, but they add an additional congestion point and overhead in the computation of applications.

In this work, we show how to take advantage of the periodic nature of HPC applications in order to develop efficient periodic scheduling strategies for their I/O transfers. Our strategy computes once during the job scheduling phase a pattern where it defines the I/O behavior for each application, after which the applications run independently, transferring their I/O at the specified times. Our strategy limits the amount of I/O congestion at the I/O node level and can be easily integrated into current job schedulers. We validate this model through extensive simulations and experiments by comparing it to state-of-the-art online solutions.

Specifically, we show that not only our scheduler has the advantage of being de-centralized, thus overcoming the overhead of online schedulers, but we also show that on Mira one can expect an average dilation improvement of 22% with an average throughput improvement of 32%!

Finally, we show that one can expect those improvements to get better in the next generation of platforms where the compute - I/O bandwidth imbalance increases.

Room: 708
9:00 am - 5:30 pm

ATIP Workshop on International Exascale and Next-Generation Computing Programs

Welcome and Introduction

David Kahaner (Asian Technology Information Program)

Speakers at this workshop include HPC leaders in the international community who develop and implement major national programs. Presentations will include a variety of related topics ranging from infrastructure development, facilities, technologies, funding, applications, human resources, industry development, impediments, etc. In addition to senior developers and policy makers, younger scientists from each of the designated countries will showcase their work using HPC in brief five-minute presentations, followed by a panel-style Q&A session that will give the audience an opportunity to get more details about their work. The aim is to foster potential global and collaborative research. Such a Workshop has never occurred before in the US.

China's New HPC Key Project

Depei Qian (Beihang University, Sun Yat-Sen University)

After a brief review on HPC research and development under China's high-tech R&D program in the past 5 years, this talk introduces the new HPC key project in the 13th 5-year plan. The motivation, goals, major activities and the current status of the new key project are briefly presented. The major challenges in developing the exascale system and the major research topics are discussed, including architecture, memory, interconnect, and system software, etc. The need for establishing the HPC application eco-system for the HPC systems based on home-grown processors is emphasized especially.

Redesigning CAM-SE for Peta-Flops Performance on Sunway TaihuLight

Haohan Fu (Tsinghua University; National Supercomputing Center, Wuxi)

With radical architectural changes in both the computing architecture and the memory hierarchy for recent leadership supercomputers, it is becoming more and more difficult for well-established numerical codes, such as the millions lines of code in the climate domain, to gain performance benefits. In this talk, we will report our efforts on achieving an efficient utilization of the Sunway TaihuLight for climate-kind applications, such as CAM-SE. We refactored and optimized the complete code using OpenACC directives at the first stage. A more aggressive and finer-grained redesign is then applied on the CAM, to achieve finer memory control and usage, more efficient vectorization and compute and communication overlapping. We further improve the CAM performance of a 260-core Sunway processor to the range of 28 to 184 Intel CPU cores, and achieve a sustainable double-precision performance of 3.3 PFlops for a 750 m global simulation when using 10,075,000 cores. CAM on Sunway achieves the simulation speed of 3.4 and 21.5 simulation-year-per-day (SYPD) for global 25-km and 100-km resolution respectively; and enables us to perform, to our knowledge, the first simulation of the complete lifecycle of hurricane Katrina, and achieve close-to-observation simulation results for both track and intensity.

Scalable Distributed Infrastructure for Data Intensive Science

David Abramson (University of Queensland)

The rise of big data science has created new demands for modern computer systems. While floating performance has driven computer architecture and system design for the past few decades, there is renewed interest in the speed at which data can be ingested and processed. Early exemplars such as Gordon, the NSF funded system at the San Diego Supercomputing Centre, shifted the focus from pure floating-point performance to memory and IO rates.

At the University of Queensland we have continued this trend with the design of FlashLite, a parallel cluster equipped with large amounts of main memory, flash disk, and a distributed shared memory system (ScaleMP's vSMP). This allows applications to place data "close" to the processor, enhancing processing speeds. Further, we have built a geographically distributed multi-tier hierarchical data fabric called MeDiCI, which provides an abstraction of very large data stores across the metropolitan area. MeDiCI leverages industry solutions such as IBM's Spectrum Scale and SGI's DMF platforms.

Caching underpins both FlashLite and MeDiCI. In this I will describe the design decisions and illustrate some early application studies that benefit from the approach. I will also highlight some of the challenges that need to be solved for this approach to become mainstream.

Morning Coffee Break

TSUBAME3.0: A Green, Accelerated, Big-Data Supercomputer *Toshio Endo (Tokyo Institute of Technology)*

Global Scientific Information and Computing center (GSIC) at Tokyo Institute of Technology has started the operation of the brand-new supercomputer, called TSUBAME3.0 in August 2017. The proposition submitted by SGI Japan has been adopted in January; since then, GSIC, SGI Japan and related vendors have cooperatively promoted preparations towards the operation start. The target applications area of TSUBAME3.0 is not limited to typical HPC area, but includes big-data and artificial intelligence (AI) area. The main components of the system are 540 compute nodes, which are customized version of SGI ICE XA. The total numbers of CPUs and GPUs are 1,080 and 2,160, respectively. The total performance is 12.15PFlops in double precision and 47.2PFlops in half precision (or more), which is an expression of floating point values in 16bit, which is expected to be useful in big-data and AI area. TSUBAME3.0 has excellent power efficiency, largely owing to GPU accelerators of the latest generation. The power-performance ratio, 14.11 GFlops, won world No.1 in the Green500 List in June 2017.

HPC Initiatives in India *Rupesh Nasre (Indian Institute of Science)*

In this talk we present the High Performance Computing (HPC) scenario in India and (new) initiatives taken. It will also cover the National Supercomputing Mission (NSM), an initiative by Government of India, jointly funded by the Department of Science and Technology and Ministry of Electronics and Information Technology. The objectives of this mission are manifold: create a number of state-of-the-art supercomputing centers, connected by a powerful network, to meet the HPC demands of researchers in the national academic and R&D institutions; develop a number of supercomputing application packages and advancement in several areas of science and engineering; create a large pool of highly skilled manpower in supercomputing; and initiate R&D activities for next generation HPC Systems. In addition, the talk will also cover the plans and initiatives of other national organizations such as the Ministry of Earth Sciences and Council of Scientific and Industrial Research.

KAUST's HiCMA Library: 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. HiCMA's target is their intersection on accelerators, where the hierarchy needs flattening. 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.

National Strategic Computing Initiative *William Harrod (US Department of Energy)*

The National Strategic Computing Initiative (NSCI) is a US national initiative that focuses the efforts of key US federal agencies into a unified strategy for advancing high performance computing (HPC) in an era of uncertain future technology. As the reliable progress in computer performance afforded by Moore's Law approaches its end, this initiative must investigate technologies and approaches for providing new sustainable pathways of both increased capability and capacity. The speaker will describe plans and progress towards achieving NSCI's five key objectives, and will discuss the Exascale Computing Project (ECP) and beyond Moore's Law projects.

Five-minute presentations by young researchers from around the world - part 1 *Yulong Ao (Chinese Academy of Sciences), Hailong Yang (Beihang University), Jiajia Li (Georgia Institute of Technology), Hong Wang (Stony Brook University), Lihao Zhang (Stony Brook University), Minh Dinh (University of Queensland), Hoang Nguyen (University of Queensland), Kartik Lakhota (University of Southern California)*

Young Researcher Panel Q&A Group 1

Lunch (participants on their own)

French HPC Ecosystem and Strategy and the Role of CEA *Jean-Phillippe Nomine (French Alternative Energies and Atomic Energy Commission)*

The initialization for decadal predictions using a coupled model includes a long-term cycle of data assimilation in the coupling

framework, which is an essential step of decadal prediction. The skill of decadal prediction closely relies on the performance of the assimilation scheme used in the initialization. Currently, most decadal prediction experiments conducted in the Coupled Model Intercomparison Project Phase 5 (CMIP5) simply used the nudging approach, and very few applied the variational or ensemble methods that are very expensive in computing. Therefore, it is necessary to develop an assimilation scheme that can well use the observations in an economical way. In this paper, we propose a coupled assimilation scheme that can incorporate 4-dimensional ocean observations (say temperature and salinity) into the IC of ocean component model and in the meantime can be timesaving. Numerical experiments using a coupled climate system model are conducted to test the performance of the new assimilation scheme.

European Initiative on HPC

Mateo Valero (Barcelona)

On March 23rd 2017, at the occasion of the 60th year celebration of the Rome European treaty, ministers from major European countries and the European Commission announced the EuroHPC initiative. The goal is to procure and deploy exascale supercomputers based on European technology in the global top 3 ranking by 2022. The initiative was signed by 7 European countries. Several other countries have joined the initiative since. Considerable funding has been committed by the EU and a large number of industrial and scientific partners have gathered together and recently submitted a proposal to develop an European low power HPC processor, as a first step to implement a full European domestic HPC stack. BSC is leading the research effort among some of the most prestigious research teams in Europe. A fundamental choice will be the move towards open standards in the definition of the CPU accelerator architecture in close collaboration with the industrial partners involved. For commercial long-term sustainability, it will be important to address a wider market than just HPC and for the time being the automotive embedded market has been retained. A considerable effort will be dedicated to complement the hardware design with the necessary software development and programming environments to allow the co-design of mission critical applications, which can greatly benefit from exascale computing capacity. The talk will briefly review the goals and plans to implement the EuroHPC initiative in Europe.

The AlgoWiki Project: an Algorithmic Pillar of Exascale Computing

Vladimir Voevodin (Lomonosov Moscow State University)

Computers are developing very quickly, and each new platform requires software developers to repeatedly analyze algorithms to answer the same two questions. Does an algorithm possess the necessary properties to meet the architectural requirements? How can the algorithm be transformed so that the necessary properties can be easily reflected in parallel programs? This analysis has to be performed again and again when a program is ported from a one type of computers to another, largely repeating the work that has been done previously. Is it possible to perform the analysis “once and for all”, describing properties of algorithms so that all of the necessary information can be gleaned from this description any time a new architecture appears? Yes, it is. This is the primary goal of the AlgoWiki project. Changes in computer architecture do not change algorithms – this is the main principle that underlies AlgoWiki and makes the project successful. “Once and for all” mentioned above is realized through comprehensive unified descriptions of algorithms included in AlgoWiki with a special emphasis on their parallel features and properties.

Thoughts on the Path Toward Exascale from a JSC Perspective

Dirk Pleiter (Forschungszentrum Juelich, University of Regensburg)

In this talk, we will discuss different aspects of the path towards exascale from the perspective of the Juelich Supercomputing Centre (JSC) in Germany. The purpose for targeting next levels of performance is the enablement of new science. We therefore understand exascale as the ability to provide 2 orders more performance for scientific applications compared to today. This does not necessarily translates into a corresponding scale-up of the throughput of floating-point operations. We will look into the application domains that play a key role for JSC in future, analyze selected requirements and confront this with current technology roadmaps.

Afternoon Coffee Break

Swiss National Programs

Thomas Schulthess (Swiss National Supercomputing Centre)

ABCI - AI Bridging Cloud Infrastructure for Everyone

Satoshi Sekiguchi (National Institute of Advanced Industrial Science and Technology)

Japan has a number of strengths with regard to AI research. One is its technological foundation in manufacturing. It is likely that AI will revolutionize the manufacturing workplaces which are the pride of Japan. Japan also offers top-class global services in the fields of healthcare and caregiving. If the huge amounts of previously accumulated data and expertise can be shared using the latest technology, including big data analytics, it will become possible to offer advanced services in even more locations. This will result in possibilities to move towards healthcare plans that are increasingly personalized for the individual.

The recent AI boom is supported by advancing Algorithms including Deep Learning, more volume of Big Data accessible, and Computing Power. For the use of AI in the real life there is a huge demand of computing resources to train networks with big data, which may take over a year with a standard server. However, manufacturing industry and/or healthcare service providers may not wish to install such larger computing machines on premise. ABCI – AI Bridging Cloud Infrastructure is designed to offer shared computing service for such business to accelerate of pervade AI. ABCI is an open innovation platform with computing resources of

more than hundred PetaFlops for world-class AI R&D in particular to accelerate Deep Learning type of computation.

In this short presentation, the ABCI project will be overviewed.

Five-minute presentations by young researchers from around the world - part 2

Teodor Nikolov (Marie Skłodowska Curie Initial Training Networks), Youenn Lebras (University of Versailles), Ilya Afanasyev (Lomonosov Moscow State University), Canstantino Gomez (Barcelona Supercomputing Center), Daniel Ruiz (Barcelona Supercomputing Center), Amani Alonazi (King Abdullah University of Science and Technology), Dalal Sukkari (King Abdullah University of Science and Technology), Saurabh Hukerikar (Oak Ridge National Laboratory), Rupesh Nasre (Indian Institutes of Technology.)

Young Researcher Panel Q&A Group 2

Open Discussion Session

Room: 710

9:00 am - 5:30 pm

LLVM-HPC2017: Fourth Workshop on the LLVM Compiler Infrastructure in HPC

Introduction - LLVM-HPC2017: Fourth Workshop on the LLVM Compiler Infrastructure in HPC

LLVM, winner of the 2012 ACM Software System Award, has become an integral part of the software-development ecosystem for optimizing compilers, dynamic-language execution engines, source-code analysis and transformation tools, debuggers and linkers, and a whole host of programming-language and toolchain-related components. Now heavily used in both academia and industry, where it allows for rapid development of production-quality tools, LLVM is increasingly used in work targeted at high-performance computing. Research in and implementation of programming-language analysis, compilation, execution and profiling has clearly benefited from the availability of a high-quality, freely-available infrastructure on which to build. This fourth annual workshop will focus on recent developments, from both academia and industry, that build on LLVM to advance the state of the art in high-performance computing.

Heterogeneous Parallel Virtual Machine and Parallelism in LLVM

Vikram Adve (University of Illinois)

Today's mainline LLVM IR, optimizers and code generators have no explicit knowledge of parallelism available in target programs, except for SIMD vector parallelism. In this talk, I will briefly describe two closely related efforts. The first, HPVM, is an ongoing research project at the University of Illinois to enable optimization, code generation, and (virtual) object code portability for diverse heterogeneous hardware, including GPUs, vector ISAs, FPGAs and domain-specific hardware accelerators. The second is a collaborative effort with several other research groups to develop mechanisms for explicitly parallel IRs to integrate cleanly with the LLVM IR, and to design a specific language-neutral parallel IR for homogeneous and heterogeneous parallel systems implemented using these mechanisms.

Break

How The Flang Frontend Works - Introduction to the Interior of the Open-Source Fortran Frontend for LLVM

Paul Osmialowski (ARM Ltd)

In May 2017, PGI publicized Flang, an Open-Source Fortran frontend for LLVM along with a complementary runtime library. The ultimate goal set for Flang is to make it part of the whole LLVM ecosystem with level of support and attention equal to that experienced by the Clang frontend. To come closer to this goal, it is important to make Flang widely known and more visible. A good introduction to the frontend interior could serve such a purpose and the intention of this presentation is to describe how Flang works and how its source code is structured.

Benchmarking and Evaluating Unified Memory for OpenMP GPU Offloading

Lingda Li (Brookhaven National Laboratory)

The latest OpenMP standard offers automatic device offloading capabilities which facilitate GPU programming. Despite this, there remain many challenges. One of these is the unified memory feature introduced in recent GPUs. GPUs in current and future HPC systems have enhanced support for unified memory space. In such systems, CPU and GPU can access each other's memory transparently, that is, the data movement is managed automatically by the underlying system software and hardware. Memory

oversubscription is also possible in these systems. However, there is a significant lack of knowledge about how this mechanism will perform, and how programmers should use it. In this paper, we aim to study and improve the performance of unified memory for automatic GPU offloading via the OpenMP API and runtime, and leveraging the Rodinia benchmark suite. We also modify the LLVM compiler to allow OpenMP to use unified memory. Then we conduct our evaluation on these benchmarks. The results reveal that while the performance of unified memory is comparable with that of normal GPU offloading for benchmarks with little data reuse, it suffers from significant overhead when GPU memory is oversubscribed for benchmarks with large amount of data reuse. Based on these results, we provide several guidelines for programmers to achieve better performance with unified memory.

PACXXv2 + RV -- An LLVM-Based Portable High-Performance Programming Model

Michael Haidl (University of Munster)

To achieve high performance on today's HPC systems, multiple programming models have to be used. An example for this burden to the developer is OpenCL: the OpenCL's SPMD programming model must be used together with a host programming model, commonly C or C++. Different programming models require different compilers for code generation, which introduce challenges for the software developer, e.g, different compilers must be convinced to agree on basic properties like type layouts to avoid subtle bugs. Moreover, the resulting performance highly depends on the features of the used compilers and may vary unpredictably.

We present PACXXv2 -- an LLVM based, single-source, single-compiler programming model which integrates explicitly parallel SPMD programming into C++. Our novel CPU back-end provides portable and predictable performance on various state-of-the-art CPU architectures comprising Intel x86 architectures, IBM Power8 and ARM Cortex CPUs. We efficiently integrate the Region Vectorizer (RV) into our back-end and exploit its whole function vectorization capabilities for our kernels. PACXXv2 utilizes C++ generalized attributes to transparently propagate information about memory allocations to the PACXX back-ends to enable additional optimizations.

We demonstrate the high-performance capabilities of PACXXv2 together with RV on benchmarks from well-known benchmark suites and compare the performance of the generated code to Intel's OpenCL driver and POCL -- the portable OpenCL project based on LLVM.

OpenMPIR

George Stelle (Los Alamos National Laboratory)

Optimizing compilers for task-level parallelism are still in their infancy. This work explores a compiler front end that translates OpenMP tasking semantics to Tapir, an extension to LLVM IR that represents fork-join parallelism. This enables analyses and optimizations that were previously inaccessible to OpenMP codes, as well as the ability to target additional runtimes at code generation. Using a Cilk runtime back end, we compare results to existing OpenMP implementations. Initial performance results for the Barcelona OpenMP task suite show performance improvements over existing implementations.

Lunch (participants on their own)

An LLVM Instrumentation Plug-In for Score-P

Ronny Tschueter (Technical University Dresden)

Reducing application runtime, scaling parallel applications to higher numbers of processes/threads, and porting applications to new hardware architectures are tasks necessary in the software development process. Therefore, developers have to investigate and understand application runtime behavior. Tools such as monitoring infrastructures that capture performance relevant data during application execution assist in this task. The measured data forms the basis for identifying bottlenecks and optimizing the code. Monitoring infrastructures need mechanisms to record application activities in order to conduct measurements. Automatic instrumentation of the source code is the preferred method in most application scenarios. We introduce a plug-in for the LLVM infrastructure that enables automatic source code instrumentation at compile-time. In contrast to available instrumentation mechanisms in LLVM/Clang, our plug-in can selectively include/exclude individual application functions. This enables developers to fine-tune the measurement to the required level of detail while avoiding large runtime overheads due to excessive instrumentation.

QUARC: An Optimized DSL Framework Using LLVM

Diptorup Deb (University of North Carolina)

We describe aspects of the implementation of QUARC, a framework layered on C++ used for a domain specific language for Lattice Quantum Chromodynamics. It is built on top of Clang/LLVM to leverage long term support and performance portability. QUARC implements a general array extension to C++ with implicit data parallelism. A notable innovation is the method for using templates to capture and encode the high-level abstractions and to communicate these abstractions transparently to LLVM through an unmodified Clang. Another notable feature is a general array transformation mechanism used to improve memory hierarchy performance and maximize opportunities for vectorization. This reshapes and transposes arrays of structures containing nested complex arrays into arrays of structures of arrays. We discuss an example for which QUARC generated code has performance competitive with the very best hand-optimized libraries.

Break

Implementing Implicit OpenMP Data Sharing on GPUs

Gheorghe-Teodor Bercea (IBM)

OpenMP is a shared memory programming model which supports the offloading of target regions to accelerators such as Nvidia GPUs. The implementation in Clang/LLVM aims to deliver a generic GPU compilation toolchain that supports both the native CUDA C/C++ and the OpenMP device offloading models. There are situations where the semantics of OpenMP and those of CUDA diverge. One such example is the policy for implicitly handling local variables. In CUDA, local variables are implicitly mapped to thread local memory and thus become private to a CUDA thread. In OpenMP, due to semantics that allow the nesting of regions executed by different numbers of threads, variables need to be implicitly shared among the threads of a contention group.

In this paper we introduce a re-design of the OpenMP device data sharing infrastructure that is responsible for the implicit sharing of local variables in the Clang/LLVM toolchain. We introduce a new data sharing infrastructure that lowers implicitly shared variables to the shared memory of the GPU.

We measure the amount of shared memory used by our scheme in cases that involve scalar variables and statically allocated arrays. The evaluation is carried out by offloading to K40 and P100 NVIDIA GPUs. For scalar variables the pressure on shared memory is relatively low, under 26% of shared memory utilization for the K40, and does not negatively impact occupancy. The data sharing scheme offers the users a simple memory model for controlling the implicit allocation of device shared memory.

Applying Temporal Blocking with a Directive-Based Approach

Toshio Endo (Tokyo Institute of Technology)

Stencil kernels are important, iterative computation patterns heavily used in scientific simulations and other operations such as image processing. The performance of stencil kernels is usually bound by memory bandwidth, and the common method of overcoming this is to apply Temporal Blocking (TB) as a form of bandwidth reducing algorithm. However, applying TB to existing code incurs high programming cost due to real-life codes embodying complex loop structures, and moreover, multitudes of parameters and blocking schemes involved in TB complicating the tuning process. We propose an automated, directive-based compiler approach for TB by extending the polyhedral compilation in the Polly/LLVM framework, significantly reducing programming cost as well as being easily subject to auto-tuning. Evaluation of the performance of our generated stencil codes on Core i7 and Xeon Phi show that the auto-generated stencil kernels achieve performance that is close to and often on par with hand TB-converted and optimized codes.

LLVM Compiler Implementation for Explicit Parallelization and SIMD Vectorization

Xinmin Tian (Intel Corporation)

With advances of modern multi-core processors and accelerators, many modern applications are increasingly turning to compiler-assisted parallel and vector programming models such as OpenMP, OpenCL, Halide, Python and Tensor-Flow. It is crucial to ensure that LLVM-based compilers can optimize parallel and vector code as effectively as possible. In this paper, we first present a set of updated LLVM IR extensions for explicitly parallel, vector, and offloading program constructs in the context of C/C++/OpenCL; Secondly, we describe our LLVM design and implementation for advanced features in OpenMP such as parallel loop scheduling, task and taskloop, SIMD loop and functions, and we discuss the impact of our updated implementation on existing LLVM optimization passes. Finally, we present a re-use case of our infrastructure to enable explicit parallelization and vectorization extensions in our OpenCL compiler to achieve ~35x performance speedup for a well-known autonomous driving workload on a multi-core platform configured with Intel Xeon Scalable Processors.

OpenMP 4.5 Validation and Verification Suite

Sunita Chandrasekaran (University of Delaware)

As directive based programming APIs like OpenMP introduce directives for accelerator devices and people are starting to use them in production codes, it is critical to have a mechanism that checks for an implementation's conformance to the standard to make sure they work correctly across architectures. This process at the same time uncovers possible ambiguities in the standard for implementors and users. We try to fill this gap through a validation and verification test-suite. This ongoing work focus first on the offload directives available in OpenMP 4.5. Our tests focus on functionality as well as use-cases from kernels extracted from applications. We have tested our tests with LLVM OpenMP compiler and runtime implementation amongst other compilers (GNU, IBM XL, Cray CCE), and we document some interesting test-cases that have uncovered implementation bugs in LLVM as well as ambiguities in the standard. In this paper, we share our methodology and experiences toward a comprehensive validation and verification test suite.

Developing an OpenMP Runtime for UVM-Capable GPUs

Hashim Sharif (University of Illinois)

With the emergence of new hardware architectures, programming models such as OpenMP must consider new design choices. In the GPU programming space, Unified Virtual Memory (UVM) is one of the new technologies that warrant such considerations. In particular, the new UVM capabilities supported by the Nvidia Pascal architecture introduce new optimization opportunities. While on-demand paging offered by a UVM-based system simplifies kernel programming, it can hamper performance due to excessive page faults at runtime. Accordingly, we have developed an OpenMP runtime that optimizes the data communication and kernel execution for UVM-capable systems. The runtime evaluates the different design choices by leveraging cost models that determine the communication and computation cost, given the application and hardware characteristics. Specifically, we employ static and dynamic analysis to identify application data access patterns that feed into the performance cost models. Our preliminary results demonstrate that the developed optimizations can provide significantly improved performance for OpenMP applications.

Improved Loop Distribution in LLVM Using Polyhedral Dependences

Tobias Grosser (ETH Zurich)

We propose a framework that can be used for improving loop-optimizations in LLVM using the Polyhedral framework of Polly. In our framework, we use the precise polyhedral dependences from Polly (provided by PolyhedralInfo), to construct a dependence graph, and perform loop transformations. As the first transformation case study of such a framework, we implemented loop distribution targeting improvement of inner-loop vectorization. Our loop distribution pass shows promising results on the TSVC benchmark; it is able to distribute 11 loops, while the earlier distribution pass is unable to distribute at all. We also have preliminary performance numbers from SPEC 2006. We believe that our work is the first step towards scalable and pre-defined loop-transformations in LLVM using exact dependences from Polly.

Implementation of a Cache Miss Calculator in LLVM/Polly

Keyur Joshi (University of Illinois)

We propose an LLVM pass to mathematically measure cache misses for Static Control Parts (SCoPs) of programs. Our implementation builds on top of the Polly infrastructure and has support for features such as LRU associativity, unknown array base addresses, and (some) approximation. We describe our preliminary results and limitations by using this pass on a selection of SCoPs. Finally we list directions for expanding and improving this work.

Concluding Remarks – LLVM-HPC2017 *Hal Finkel (Argonne National Laboratory)*

Room: 712

9:00 am - 5:30 pm

Fourth Workshop on Accelerator Programming Using Directives (WACCPD),

Introduction - Fourth Workshop on Accelerator Programming Using Directives (WACCPD)

Guido Juckeland (Helmholtz-Zentrum Dresden-Rossendorf)

In the current pre-Exascale era domain and computational scientists still struggle with adapting large applications or prototyping new ideas on the plethora of novel hardware architecture with diverse memory subsystems or cores with different ISAs or accelerators of varied types. The HPC community is in constant need for sophisticated software tools and techniques to port legacy code to these emerging platforms.

Given the complexity in hardware, maintaining a single code base yet achieving performance portable solution continues to pose a daunting task. Directive-based programming models such as OpenACC, OpenMP have been tackling this issue by offering scientists a high-level approach to accelerate scientific applications and develop solutions that are portable and yet do not compromise on performance or accuracy. Such programming paradigm has facilitated complex heterogeneous systems to be classified as first-class citizens for HPC

This workshop aims to solicit papers that explore innovative language features and their implementations, stories and lessons learnt while using directives to migrate scientific legacy code to parallel processors, state-of-the-art compilation and runtime scheduling techniques, performance optimization and analysis on state-of-the-art hardware and so on.

WACCPD has been one of the major forums for bringing together users, developers, software and tools community together to share knowledge and experiences to program emerging complex parallel computing systems.

Using Accelerator Directives to Adapt Science Applications for State-of-the-Art HPC Architectures *John Stone (University of Illinois)*

The tremendous early success of accelerated HPC systems has been due to their ability to reach computational key performance targets with high energy efficiency, and without placing undue burden on the developers of HPC applications software. Upcoming accelerated systems are bringing tremendous performance levels and technical capabilities, but fully exploiting these advances requires science applications to incorporate fine-grained parallelism pervasively, even deep within application code that historically was not considered performance-critical.

Accelerator directives provide a critical high-productivity path to resolving this requirement, even within applications that already make extensive use of hand-coded accelerator kernels. This talk will explore the role of accelerator directives and their complementarity to both hand-coded kernels and new hardware features to maintain high development productivity and performance for production scientific applications. The talk will also consider some potential challenges posed by both both mundane and disruptive hardware evolution in future HPC systems and how they might relate to directive-based accelerator programming approaches.

Morning Break

An Example of Porting PETSc Applications to Heterogeneous Platforms with OpenACC

Pi-Yueh Chuang (George Washington University)

In this paper, we document the workflow of our practice to port a PETSc application with OpenACC to a supercomputer, Titan, at Oak Ridge National Laboratory. Our experience shows a few lines of code modifications with OpenACC directives can give us a speedup of 1.34x in a PETSc-based Poisson solver (conjugate gradient method with algebraic multigrid preconditioner). This demonstrates the feasibility of enabling GPU capability in PETSc with OpenACC. We hope our work can serve as a reference to those who are interested in porting their legacy PETSc applications to modern heterogeneous platforms.

Hybrid Fortran: High Productivity GPU Porting Framework Applied to Japanese Weather Prediction Model

Michel Mueller (Tokyo Institute of Technology)

In this work we use the GPU porting task for the operative Japanese weather prediction model "ASUCA" as an opportunity to examine productivity issues with OpenACC when applied to structured grid problems. We then propose "Hybrid Fortran", an approach that combines the advantages of directive based methods (no rewrite of existing code necessary) with that of stencil DSLs (memory layout is abstracted). This gives the ability to define multiple parallelizations with different granularities in the same code. Without compromising on performance, this approach enables a major reduction in the code changes required to achieve a hybrid GPU/CPU parallelization - as demonstrated with our ASUCA implementation using Hybrid Fortran.

Implicit Low-Order Unstructured Finite-Element Multiple Simulation Enhanced by Dense Computation Using OpenACC

Takuma Yamaguchi (University of Tokyo), Kohei Fujita (University of Tokyo, RIKEN)

In this paper, we develop a low-order three-dimensional finite-element solver for fast multiple-case crust deformation analysis on GPU-based systems. Based on a high-performance solver designed for massively parallel CPU based systems, we modify the algorithm to reduce random data access, and then insert OpenACC directives. The developed solver on ten Reedbush-Hnodes (20 P100 GPUs) attained speedup of 14.2 times from 20 K computer nodes, which is high considering the peak memory bandwidth ratio of 11.4 between the two systems. On the newest Volta generation V100 GPUs, the solver attained a further 2.45 times speedup from P100 GPUs. As a demonstrative example, we computed 368 cases of crustal deformation analyses of northeast Japan with 400 million degrees of freedom. The total procedure of algorithm modification and porting implementation took only two weeks; we can see that high performance improvement was achieved with low development cost. With the developed solver, we can expect improvement in reliability of crust-deformation analyses by many-case analyses on a wide range of GPU-based systems.

Lunch

The Challenges Faced by OpenACC Compilers *Randy Allen (Mentor Graphics)*

It is easy to assume that a directive-based system such as OpenACC solves all the problems faced by compiler-developers for high performance architectures. After all, the hard part of optimizing compilers for high performance architectures is figuring out what parts of the application to do in parallel, and OpenACC directives serve up that information to the compiler up on a platter. What more is left for the compiler writer to do but to lay back, collect paychecks, and watch application developers do all of his/her work?

Unfortunately (or fortunately if you are employed as a compiler developer), the world is not so simple. While OpenACC directives definitely simplify development of some parts of optimizing compilers, they introduce a new set of challenges. Just because a user has said a loop can be done in parallel does not mean that the compiler automatically knows how to schedule it in parallel. For that matter, is the user actually applying intelligence when inserting directives, or is she/he instead randomly inserting directives to see what runs faster? The latter occurs far more often than many people suspect.

This talk briefly overviews the challenges faced by a compiler that targets high performance architectures and the assistance in meeting those challenges provided by a directive-based system like OpenACC. It will then present the challenges introduced to the compiler by directive-based systems, with the intent of steering application developers toward an understanding of how to assist compilers in overcoming the challenges and obtaining supercomputer performance on applications without investing superhuman effort.

The Design and Implementation of OpenMP 4.5 and OpenACC Backends for the RAJA C++ Performance Portability Layer

William Killian (Millersville University of Pennsylvania, University of Delaware)

Portability abstraction layers such as RAJA enable users to quickly change how a loop nest is executed with minimal modifications to high-level source code. Directive-based programming models such as OpenMP and OpenACC provide easy-to-use annotations on for-loops and regions which change the execution pattern of user code. Directive-based language backends for RAJA have previously been limited to few options due to multiplicative clauses creating version explosion. In this work, we introduce an updated implementation of two directive-based backends which helps mitigate the aforementioned version explosion problem by leveraging the C++ type system and template meta-programming concepts. We implement partial OpenMP 4.5 and OpenACC backends for the RAJA portability layer which can apply loop transformations and specify how loops should be executed. We evaluate our approach by analyzing compilation and runtime overhead for both backends using PGI 17.7 and IBM clang (OpenMP 4.5) on a collection of computation kernels.

Enabling GPU Support for the COMPSs-Mobile Framework

Francesc-Josep Lordan Gomis (Barcelona Supercomputing Center, Polytechnic University of Catalonia)

Using the GPUs embedded on mobile devices allows for increasing the performance of the applications running on them while reducing the energy consumption of their execution. This article presents a task-based solution for adaptative, collaborative heterogeneous computing on mobile cloud environments. To implement our proposal, we extend the COMPSs-Mobile framework — an implementation of the COMPSs programming model for building mobile applications that offload part of the computation to the Cloud — to support offloading computation to GPUs through OpenCL. To evaluate the behavior of our solution, we subject the prototype to three benchmark applications representing different application patterns.

Concurrent parallel processing on Graphics and Multicore Processors with OpenACC and OpenMP *Christopher Stone (Computational Science and Engineering LLC)*

Hierarchical parallel computing is rapidly becoming ubiquitous in high performance computing (HPC) systems. Programming models used commonly in turbomachinery and other engineering simulation codes have traditionally relied upon distributed memory parallelism with MPI and have ignored thread and data parallelism. This paper presents methods for programming multi-block codes for concurrent computational on host multicore CPUs and many-core accelerators such as graphics processing units. Portable and standardized language directives are used to expose data and thread parallelism within the hybrid shared- and distributed-memory simulation system. A single-source, multiple-object strategy is used to simplify code management and allow for heterogeneous computing. Automated load balancing is implemented to determine what portions of the domain are computed by the multicore CPUs and GPUs. Benchmark results show that significant parallel speed-up is attainable on multicore CPUs and many-core devices such as the Intel Xeon Phi Knights Landing using OpenMP SIMD and thread parallel directives. Modest speed-up, relative to a CPU core, was achieved with OpenACC offloading to NVIDIA GPUs. Combining both GPU offloading with multicore host parallelism improved the single-device performance by 30% but further speed-up was not realized when more heterogeneous CPU-GPU device pairs were included.

Afternoon Break

Exploration of Supervised Machine Learning Techniques for Runtime Selection of CPU vs GPU Execution in Java Programs

Akihiro Hayashi (Rice University)

While multi-core CPUs and many-core GPUs are both viable platforms for parallel computing, programming models for them can impose large burdens upon programmers due to their complex and low-level APIs. Since managed languages like Java are designed to be run on multiple platforms, parallel language constructs and APIs such as Java 8 Parallel Stream APIs can enable high-level parallel programming with the promise of performance portability for mainstream (“non-ninja”) programmers. To achieve this goal, it is important for the selection of the hardware device to be automated rather than be specified by the programmer, as is done in current programming models. Due to a variety of factors affecting performance, predicting a preferable device for faster performance of individual kernels remains a difficult problem. While a prior approach makes use of machine learning algorithms to address this challenge, there is no comparable study on good supervised machine learning algorithms and good program features to track. In this paper, we explore 1) program features to be extracted by a compiler and 2) various machine learning techniques that improve accuracy in prediction, thereby improving performance. The results show that an appropriate selection of program features and machine learning algorithms can further improve accuracy. In particular, support vector machines (SVMs), logistic regression, and J48 decision tree are found to be reliable techniques for building accurate prediction models from just two, three, or four program features, achieving accuracies of 99.66%, 98.63%, and 98.28% respectively from 5-fold-cross-validation.

Automatic Testing of OpenACC Applications

Khalid Ahmad (University of Utah)

PCAST (PGI Compiler-Assisted Software Testing) is a feature in our compiler and runtime to help users automate testing high performance numerical programs. PCAST normally works by running a known working version of a program and saving intermediate results to a reference file, then running a test version of a program and comparing the intermediate results against the reference file. Here, we describe the special case of using PCAST on OpenACC programs running on a GPU. Instead of saving and comparing against a saved reference file, the compiler generates code to run each compute region on both the host CPU and the GPU. The values computed on the host and GPU are then compared, using OpenACC data directives and clauses to decide what data to compare.

Evaluation of Asynchronous Offloading Capabilities of Accelerator Programming Models for Multiple Devices

Christian Terboven (RWTH Aachen University)

Accelerator devices are increasingly used to build large supercomputers and current installations usually include more than one accelerator per system node. To keep all devices busy, kernels have to be executed concurrently which can be achieved via asynchronous kernel launches. This work compares the performance for an implementation of the Conjugate Gradient method with CUDA, OpenCL, and OpenACC on NVIDIA Pascal GPUs. Furthermore, it takes a look at Intel Xeon Phi coprocessors when programmed with OpenCL and OpenMP. In doing so, it tries to answer the question whether the higher abstraction level of directive based models is inferior to lower level paradigms in terms of performance.

Friday, November 17th

Room: 301
8:30 am - 12:00 pm

Fourth International Workshop on Visual Performance Analysis – VPA 2017

Introduction - 4th International Workshop on Visual Performance Analytics – VPA 2017

The last few decades have witnessed an unprecedented growth in the computational power of supercomputers. Exploiting the full capabilities of these machines is becoming exponentially more difficult with each new generation of hardware. To help understand and optimize the behavior of parallel applications running on such hardware, the performance analysis community has created a wide range of tools and APIs to collect performance data, such as hardware counters, communication traces, and network traffic data. This success has created a new challenge: the resulting data is far too large and too complex to be analyzed in a straightforward manner. Therefore, new automatic analysis and visualization approaches must be developed to enable application developers to intuitively understand the multiple, interdependent effects that their algorithmic choices have on application performance. This workshop intends to bring together researchers from the fields of performance analysis and visualization for a continued exchange of new approaches to leverage both areas to analyze and optimize large-scale parallel applications.

VPA Keynote: Visual Performance Analysis for Extremely Heterogeneous Systems

Lucy Nowell (US Department of Energy)

Extreme heterogeneity is the result of using multiple types of processors, accelerators, memory and storage in a single computing platform or environment that must support an expanding variety of application workflows to meet the needs of increasingly heterogeneous users. Extremely heterogeneous supercomputers are likely to be acquired by the ASCR-supported supercomputing facilities as we reach the end of Moore's Law while still facing rapidly increasing computational and data intensive requirements. The exponential increase in system complexity will make it essential for system administrators and software developers to have new tools that help them understand the behavior of extremely heterogeneous supercomputing environments and the applications that run in them. The vast bandwidth of visual perception makes the combination of visualization and performance analysis essential.

Toward Aggregated Grain Graphs

Ananya Muddukrishna (Norwegian University of Science and Technology)

Grain graphs simplify OpenMP performance analysis by visualizing performance problems from a fork-join perspective that is familiar to programmers. However, it is tedious to navigate and diagnose problems in large grain graphs with thousands of task and parallel for-loop chunk instances. We present an aggregation method that matches recurring patterns in grain graphs and groups related nodes together, reducing graphs of any size to one root group. The aggregated grain graph is then navigated by progressively uncovering groups and analyzing only those groups that have problems. This enhances productivity by enabling programmers to understand program structure and problems in large grain graphs with less effort than before.

Break

Panel Discussion: Challenges and the Future of HPC Performance Visualization

This panel will discuss the challenges and future of HPC performance visualization. In particular, its focus will be on how we collect and evaluate the performance of HPC systems, the available tools we have and their limitations, and how best we can utilize visualization to improve upon the state-of-the-art. The panel will combine perspectives from users, developers, and experts in visualization and performance analysis, to help stimulate discussions from all sides.

Projecting Performance Data Over Simulation Geometry Using SOSflow and Alpine

The performance of HPC simulation codes is often tied to their simulated domains; e.g., properties of the input decks, boundaries of the underlying meshes, and parallel decomposition of the simulation space. A variety of research efforts have demonstrated the utility of projecting performance data onto the simulation geometry to enable analysis of these kinds of performance problems. However, current methods to do so are largely ad-hoc and limited in terms of extensibility and scalability. Furthermore, few methods enable this projection online, resulting in large storage and processing requirements for offline analysis. We present a general, extensible, and scalable solution for in-situ (online) visualization of performance data projected onto the underlying geometry of simulation codes. Our solution employs the scalable observation system SOSflow with the in-situ visualization framework ALPINE to automatically extract simulation geometry and stream aggregated performance metrics to respective locations within the geometry at runtime. Our system decouples the resources and mechanisms to collect, aggregate, project, and visualize the resulting data, thus mitigating overhead and enabling online analysis at large scales. Furthermore, our method requires minimal user input and modification of existing code, enabling general and widespread adoption.

Visualizing, Measuring, and Tuning Adaptive MPI Parameters

Mattias Diener (University of Illinois)

Adaptive MPI (AMPI) is an advanced MPI runtime environment that offers several features over traditional MPI runtimes, which can lead to a better utilization of the underlying hardware platform and therefore higher performance. These features are overdecomposition through virtualization, and load balancing via rank migration. Choosing which of these features to use, and finding the optimal parameters for them is a challenging task however, since different applications and systems may require different options. Furthermore, there is a lack of information about the impact of each option. In this paper, we present a new visualization of AMPI in its companion Projections tool, which depicts the operation of an MPI application and details the impact of the different AMPI features on its resource usage. We show how these visualizations can help to improve the efficiency and execution time of an MPI application. Applying optimizations indicated by the performance analysis to two MPI-based applications results in performance improvements of up to 18% from overdecomposition and load balancing.

Room: 302-303
8:30 am - 12:00 pm

The 2nd International Workshop on Data Reduction for Big Scientific Data (DRBSD-2)

Introduction - The 2nd International Workshop on Data Reduction for Big Scientific Data (DRBSD-2)

A growing disparity between simulation speeds and I/O rates makes it increasingly infeasible for high-performance applications to save all results for offline analysis. This trend has spurred interest in high-performance online data analysis and reduction methods, motivated by a desire to conserve I/O bandwidth, storage, and/or power; increase accuracy of data analysis results; and/or make optimal use of parallel platforms, among other factors. This requires our community to understand a clear yet complex relationships between application design, data analysis and reduction methods, programming models, system software, hardware, and other elements of a next-generation High Performance Computer, particularly given constraints such as applicability, fidelity, performance portability, and power efficiency.

There are at least three important topics that our community is striving to answer: (1) whether several orders of magnitude of data reduction is possible for exascale sciences; (2) understanding the performance and accuracy trade-off of data reduction; and (3) solutions to effectively reduce data while preserving the information hidden in large scientific data. Tackling these challenges requires expertise from computer science, mathematics, and application domains to study the problem holistically, and develop solutions and hardened software tools that can be used by production applications.

The goal of this workshop is to provide a focused venue for researchers in all aspects of data reduction and analysis to present their research results, exchange ideas, identify new research directions, and foster new collaborations within the community.

SKA: The Data Domino Enabled by DALiuGE

Andreas Wicenec (University of Western Australia, International Centre for Radio Astronomy Research)

The Square Kilometre Array (SKA) will pose interesting new challenges on the way scientific computing is carried out. The processing will require to connect the antenna arrays in South Africa and Australia to dedicated 200PF scale HPC centres over some 700km WAN connections. The first part of the processing will be carried out on a sub-second cadence on data streams of about 1TB/s. The further processing will collect the data of a 6-12 hour long observation and perform an iterative image reconstruction. With current algorithms the bottleneck seems to be in memory bandwidth, and the level of data parallelism and inherent concurrency reaches levels of several tens of millions of tasks and data items to be scheduled and managed during a single processing run. The design of the SKA processing system includes an execution framework detailing the concepts of an architecture enabling the processing at SKA scale. Along with working on the design of this execution framework, we have also implemented a prototype to prove the viability of the proposed design decisions and extract the detailed requirements. The result of the prototyping work is called DALiuGE, which stands for 'Data Activated Flow Graph Engine'. DALiuGE implements most of the concepts required to perform the various radio astronomy workflows, while completely avoiding any unnecessary features. DALiuGE is completely generic and can be adopted to any kind of similar workflow problems. This talk will highlight the key concepts and solutions of DALiuGE and also present the results of test runs at scale.

An Efficient Approach to Lossy Compression with Pointwise Relative Error Bound

Sheng Di (Argonne National Laboratory)

An effective data compressor is becoming increasingly critical because of the extremely large volumes of data produced by scientific simulations. Many lossy compressors have been developed in the context of absolute error bounds (e.g., minimizing the maximum absolute error or root-mean-squared error). In order to achieve a multiresolution effect, however, scientific applications may prefer to compress the data with a pointwise relative error bound (i.e., the larger the data value, the larger the compression error to tolerate). In this paper, we explore a lossy compression strategy based on the requirement of pointwise relative error bound, under a state-of-the-art compression framework (prediction + quantization + entropy encoding). Specifically, we split the data set into many small blocks and compute a real absolute error bound for each block based on the pointwise relative error bound set by users. We implement this solution and evaluate it using a real-world scientific data set. Experiments show that the compression ratio of our solution is higher than that of other state-of-the-art compressors by 17.2–618%, with comparable compression/decompression times, the same relative error bound, and similar peak signal-to-noise ratio (PSNR).

Data Reduction and Partitioning in an Extreme Scale GPU-Based Clustering Algorithm

Benjamin Welton (University of Wisconsin)

The emergence of leadership-class systems with GPU-equipped nodes has the potential to vastly increase the performance of existing distributed applications. An increasing number of applications that are converted to run on these systems are reliant on algorithms that perform computations on spatial data. Algorithms that operate on spatial data, such as density-based clustering algorithms, present unique challenges in data partitioning and result aggregation when porting to extreme scale environments. These algorithms require that spatially-near data points are partitioned to the same node and that the output from individual nodes needs to be aggregated to ensure that relationships between partition boundaries are discovered. In the development of an extreme scale density-based clustering algorithm, called Mr. Scan, we leveraged the increased computational power provided by GPUs to overcome these challenges. Three main techniques allowed Mr. Scan to cluster 6.5 billion points from the Twitter dataset using 8,192 GPU nodes on Cray Titan in 7.5 minutes: (1) a data partitioner scheme that ensures correctness by using shadow regions to selectively duplicate computation between nodes to detect clusters that lie in-between partition boundaries, (2) a dense box algorithm that reduces the computational complexity of clustering dense spatial regions, and (3) a new tree based method for merging results with spatial dependencies that requires only a fraction of the intermediate results to produce an accurate final result. The pure computational power of the GPUs allowed us to effectively summarize the initial data, making the scalable use of these techniques possible.

MGARD: A Multilevel Technique for Compression of Floating-Point Data

Mark Ainsworth (Brown University, Oak Ridge National Laboratory)

MGARD (MultiGrid Adaptive Reduction of Data) is a recently developed technique for multilevel lossy compression and reduction of scientific data. The technique is based on the theory of multigrid methods, which have found widespread application in the computational solution of partial differential equations. MGARD allows the user to specify either a lossiness tolerance level or a size constraint and produces a quasioptimal reduction. Moreover, the data is stored in a hierarchical decomposition well-suited to heterogeneous storage systems.

In this study, we outline algorithms implementing MGARD and perform computational experiments demonstrating its effectiveness. We apply MGARD and state of the art compression tools to two datasets, one generated to prescribed smoothness and another obtained from power grid monitoring micro-phasor measurement units installed at Lawrence Berkeley National Laboratory. The results indicate that in its current preliminary state MGARD offers a competitive alternative to existing state of the art compression tools.

Break

Facing the Big Data Challenge in the Fusion Code XGC

Choongseok Chang (Princeton University)

Boundary plasma of a magnetic fusion reactor is far from a thermodynamic equilibrium, with the physics dominated by nonlinear multiscale multiphysics interactions in a complicated geometry, and requires extreme-scale computing for first-principles based understanding. The scalable particle-in-cell code XGC has been developed for this purpose, in partnership with the computer science and applied mathematics communities over the last decade. XGC's extreme-scale capability has been recognized by being awarded several hundred million hours of computing time on all US three leadership class computers, and by being selected into all three pre-exascale/exascale programs: CAAR, NESAP, and AURORA ESP. The physics data size produced from a 1-day XGC run of ITER plasma on the present ~20PF computer is ~100PB, which is much above the limit imposed by the present technology. We are thus losing most of the valuable physics data in order to keep the data flow within the limits imposed by the I/O rate and the file-system size. Since the problem size will increase in proportion to the parallel computer capability, the challenge will grow at least 100-fold as the exascale computers arrive. Reduction of the data size by several orders of magnitude is required that can still preserve the accuracy for enabling various levels of scientific discoveries. On-the-fly, in-memory data analysis and visualization must occur at the same time. These issues, as well as the necessity to collaborate tightly with the applied mathematics and computer science communities, will be discussed from the application driver point of view.

Adaptive Compression to Improve I/O Performance for Climate Simulations

Swati Singhal (University of Maryland)

We present an adaptive compression tool for scientific applications that automatically determines and adapts the best among a set of well-known effective compression schemes to each data variable and enables optimizing both compression ratio and compression overhead. Our adaptive compression library ACOMPS integrates several lossless compression algorithms and also reorganizes data variables in a preprocessing step to enable the compression schemes to work well for various types of data, including floating point values. Our library can be tuned to pick the best compression schemes based on compression ratio, compression speed or a combination of both. For ease of use, we also provide our library as a plugin for the widely used ADIOS middleware I/O system. In our experiments with a climate simulation application, we show that ACOMPS performs well compared to other lossless compression methods and has low overhead.

Lean Visualization of Large Scale Tree-Based AMR Meshes

Philippe P. Pébaÿ (NexGen Analytics)

We present a novel adaptive method for the visualization and analysis of large-scale, parallel, tree-based adaptive mesh refinement

(AMR) scientific simulations. When compared to those obtained with standard approaches for the visualization of such datasets, our results indicate a gain of at least 80% in terms of memory footprint with a better rendering, while retaining similar execution speed.

Furthermore, our adaptive approach allows for further acceleration of the rendering of 2-dimensional AMR grids, hereby solving the problem posed by the loss of interactivity that occurs when dealing with large and/or deeply refined meshes. This was achieved by taking advantage of the intrinsic structure of tree-based AMR grids, allowing for on-the-fly data reduction at visualization time with guaranteed rendering accuracy.

Contextual Compression of Large-Scale Wind Turbine Array Simulations

Kenny Gruchalla (National Renewable Energy Laboratory)

Data sizes are becoming a critical issue particularly for HPC applications. We have developed a user-driven lossy wavelet-based storage model to facilitate the analysis and visualization of large-scale wind turbine array simulations. The model stores data as heterogeneous blocks of wavelet coefficients, providing high-fidelity access to user-defined data regions believed the most salient, while providing lower-fidelity access to less salient regions on a block-by-block basis. In practice, by retaining the wavelet coefficients as a function of feature saliency, we have seen data reductions in excess of 94%, while retaining lossless information in the turbine-wake regions most critical to analysis and providing enough (low-fidelity) contextual information in the upper atmosphere to track incoming coherent turbulent structures. Our contextual wavelet compression approach has allowed us to deliver interactive visual analysis while providing the user control over where data loss, and thus reduction in accuracy, in the analysis occurs. We argue this reduced but contextualized representation is a valid approach and encourages contextual data management.

Room: 402-403-404
8:30 am - 12:00 pm

H2RC: Third International Workshop on Heterogeneous Computing with Reconfigurable Logic

Introduction - H2RC: Third International Workshop on Heterogeneous Computing with Reconfigurable Logic

For the third time in a row, this workshop will bring together application experts, software developers, and hardware engineers to share experiences and best practices leveraging reconfigurable logic in HPC and “Big Data” applications. In particular, the workshop will focus on sharing experiences and techniques for accelerating applications and/or improving energy efficiency with FPGAs using OpenCL, OpenMP, OpenACC, C, and C++ based design flows, which enable and improve cross-platform functional and performance portability. Particular emphasis is given to cross-platform comparisons that foster a better understanding within the industry and research community on what are the best mappings of applications to a diverse range of hardware architectures that are available today (e.g., FPGA vs. GPU vs. Many-cores and hybrid devices), and on how to most effectively achieve cross-platform compatibility.

Keynote: FPGAs in AWS and First Use Cases (joint talk by AWS, NGcodec, and Xilinx)

Mark Duffield, Kees Vissers, Oliver Gunasekara - (Amazon Web Services, Xilinx Inc, NGCodec)

Amazon focuses on the development and deployment of cloud-scale FPGA-accelerated applications, showing you how to get started building high performance solutions on AWS. The talk also provides an update on AWS FPGA partners, and the AWS Marketplace. With the availability of FPGAs on AWS, developers are creating custom hardware accelerated solutions to solve complex problems in areas like big data processing, healthcare and life sciences, security, image and video processing, financial, and applied research. Amazon EC2 F1 provides developers with a cloud development framework, ability to rapidly scale their accelerations, and enabling access to FPGA technologies to millions of AWS customers via AWS Marketplace.

NGcodec will show the development and deployment of an H.265/HEVC encoder for AWS F1. The design for the total system includes algorithm and implementation trade-offs that will be shown. The resulting implementation is a significant improvement over any implementation that leverages CPUs only. The combination of FPGAs and CPUs in a cloud environment offers novel technical and business opportunities.

Xilinx will illustrate both the Silicon and Software environment that is available in the AWS EC F1 instance. This environment has the details to enable efficient hardware implementations and also has abstractions to enable software developers.

Case Study: Usage of High Level Synthesis in HPC Networking

Petr Kaštovský (Netcope Technologies)

Taking an inspiration from Software Defined Networking and Network Functions Virtualization, also the HPC world has the opportunity to switch to much more flexible network programming paradigm. The P4 language was originally designed for a description of networking switches' dataplane. However, we examine using it in a related domain: SmartNICs. We see the P4 as an interface for application designers to formally express their requirements for the functionality of the network. This achievement can be further extended by embedding hardware accelerators (such as modules for pattern matching, cryptography, etc.) into the FPGA based SmartNIC. Network-related parts of the compute load can be then moved into the networking hardware, effectively blurring the edge between the HPC network and compute components.

The work of FPGA firmware developers is by no means replaced by the P4 to FPGA compiler. Rather, firmware developers can spend less time over small changes and other mundane tasks and will be able to focus on more complex tasks, such as effective

architectures and algorithms for network-compute accelerators. In order to demonstrate the feasibility of P4 for high-speed packet processing in FPGA, we use two existing and mature FPGA-based projects: Netcope Packet Capture is an FPGA firmware capable of 100 Gbps line rate traffic filtering and forwarding to output network ports or host RAM via PCI Express.

RE-HASE: Regular-Expressions Hardware Synthesis Engine *Mohamed El-Hadedy (University of Illinois), Xinfei Guo (University of Virginia), Xiaoping Huang (Northwestern Polytechnical University), Martin Margala (University of Massachusetts, Lowell)*

Adopting OpenCAPI for High Bandwidth Database Accelerators

A FPGA-Pipelined Approach for Accelerated Discrete-Event Simulation of HPC Systems

OpenCL for FPGAs/HPC: Case Study in 3D FFT

Break

Accelerating Deep Neural Networks at Datacenter Scale with the BrainWave Architecture

Kalin Ovtcharov (Microsoft)

In the last several years, advances in deep neural networks (DNN) have led to many breakthroughs in machine learning, spanning diverse fields such as computer vision, speech recognition, and natural language processing. Since then, the size and complexity of DNNs have significantly outpaced the growth of CPU performance, leading to an explosion of DNN accelerators.

Recently, major cloud vendors have turned to ASICs and FPGAs to accelerate DNN serving in latency-sensitive online services. While significant attention has been devoted to deep convolutional neural networks (CNN), much less so has been given toward the more difficult problems of memory-intensive Multilayer Perceptions (MLP) and Recurrent Neural Networks (RNN).

Improving the performance of memory-intensive DNNs (e.g., LSTM) can be achieved using a variety of methods. Batching is one popular approach that can be used to drive up device utilization, but can be harmful to tail latencies (e.g., 99.9th) in the context of online services. Increasing off-chip bandwidth is another option, but incurs high cost and power, and still may not be sufficient to saturate an accelerator with tens or even hundreds of TFLOPS of raw performance.

In this work, we present BrainWave, a scalable, distributed DNN acceleration architecture that enables inferencing of MLPs and RNNs at near-peak device efficiency without the need for input batching. The BrainWave architecture is enabled and powered by FPGAs and is able to reduce latency 10-100X relative to well-tuned software.

LESS: Loop Nest Execution Strategies for Spatial Architectures

Heterogeneous Multi-Processing in Software-Defined Cloud Storage Nodes

Snowflake: Efficient Accelerator for Deep Neural Networks

Porting a GAMESS Computational Chemistry Kernel to FPGAs

FPGAs for Supercomputing? Progress and Challenges

Hal Finkel (Argonne National Laboratory)

Field-programmable gate arrays (FPGAs) have become keystones of our modern electronics ecosystem. Despite higher-end components possessing immense processing power, however, FPGAs are rarely employed as general-purpose computational accelerators. In this talk, I'll discuss why FPGAs seem attractive as HPC accelerators and our experience experimenting with different kinds of algorithms on FPGAs. Finally, I'll share some thoughts on how common HPC programming models and environments can be adapted to support FPGAs, and moreover, how partial reconfigurability and combined CPU/FPGA packages can be leveraged to make this easier.

Room: 405-406-407
8:30 am - 12:00 pm

MTAGS17: 10th Workshop on Many-Task Computing on Clouds, Grids, and Supercomputers

Introduction - MTAGS17: 10th Workshop on Many-Task Computing on Clouds, Grids, and Supercomputers

Ke Wang (Microsoft), Vani Mandava (Microsoft)

The 10th workshop on Many-Task Computing on Clouds, Grids, and Supercomputers (MTAGS17) will provide both the scientific and industrial communities a dedicated forum for presenting new research, development, and deployment efforts of algorithms, frameworks, and systems for many-task computing (MTC), machine learning and big data applications on large scale clusters, clouds, grids, and supercomputers. The theme of the workshop encompasses loosely-coupled applications driven by big data. The applications are generally composed of many tasks (e.g., millions to billions) to achieve some larger application goal. This workshop will cover challenges that can hamper efficiency and utilization in running applications on extreme-scale systems, such as local resource manager's scalability and granularity, data-aware scheduling, efficient utilization of intra-node parallelism, parallel file-system contention and scalability, data locality, I/O management, reliability at scale, and application scalability. We welcome paper submissions in theoretical, simulations, and real systems topics with special consideration to papers addressing the intersection of petascale/exascale challenges with large-scale cloud computing and machine learning. Papers will be peer-reviewed, and accepted papers will be published in the workshop proceedings as part of the ACM SIGHPC. The workshop will be held in conjunction with SC17---The International Conference on High Performance Computing, Networking, Storage and Analysis---in Denver, Colorado, USA.

<https://www.cse.unr.edu/mtags17/>

Keynote: Cloud based systems and challenges for data rich research workloads

Vani Mandava (Microsoft)

In this talk we show a glimpse of typical research topologies across hundreds of research projects over the past few years that leveraged platform services and data systems in the cloud. We then look at a few case studies that dive into specific techniques that researchers have used to distribute and efficiently process different types of data using diverse data architecture and storage mechanisms, in domains such as genomics, smart cities, healthcare, and education among others. Finally, we explore the hard challenges in this space, and some recent and upcoming innovations.

Automated Cluster Provisioning And Workflow Management for Parallel Scientific Applications in the Cloud.

Break

Two stage cluster for resource optimization with Apache Mesos

Scylla: A Mesos Framework for Container Based MPI Jobs

Syndesis: Bridging the Semantic Gap between Object-based and File-based Data Models

Closing Remarks

Room: 605
8:30 am - 12:00 pm

Computational Approaches for Cancer

New computational opportunities and challenges have emerged within the cancer research and clinical application areas as the size, number, variety and complexity of cancer datasets have grown in recent years. Simultaneously, advances in computational capabilities have grown and are expected to continue to reach unprecedented scales. Such opportunities to impact cancer computationally are underscored in the 2016 Cancer Moonshot Blue Ribbon Panel Report. The workshop focuses on bringing together interested individuals ranging from clinicians, mathematicians, data scientists, computational scientists, hardware experts, engineers, developers, leaders and others with an interest in advancing the use of computation at all levels to better understand, diagnose, treat and prevent cancer. With an interdisciplinary focus, the workshop provides opportunities for participants to learn how computation is employed across multiple areas including imaging, genomics, analytics, modeling, pathology and drug discovery with a focus on impacting cancer. As an interdisciplinary workshop, the cross-disciplinary sharing of insight and challenges fosters collaborations and future innovations to accelerate the progress in computationally and data driven cancer research and clinical applications. The forward focus of the workshop looks at challenges and opportunities for large scale HPC, including exascale applications involving cancer.

Room: 710-712
8:30 am - 12:00 pm

WOLFHPC: Workshop on Domain-Specific Languages and High-Level Frameworks for High-Performance Computing

Multi-level heterogeneous parallelism and deep memory hierarchies in current and emerging computer systems make their programming for high-performance applications very difficult. The task is made more challenging by the changing landscape of system architecture mandated by power and micro-architecture considerations. Domain-specific languages (DSLs) and high-level frameworks (HLFs) provide convenient abstractions, shielding application developers from much of the complexity and variability of explicitly programming in standard programming languages like C/C++/Fortran. Effective design of such abstractions for the high-performance computing context requires close interaction between researchers developing such languages and frameworks and domain experts with a deep understanding of the problem to be solved.

While a number of other venues exist that address domain-specific languages and high-level programming abstractions, none of them focus on the challenges to performance optimization and implementation on parallel systems. This workshop seeks to bring together developers and users of DSLs and HLFs to identify challenges and discuss solution approaches for their effective implementation and use on a variety of platforms, including massively parallel systems.

We solicit submissions on all aspects relating to domain-specific languages and high-level frameworks for the HPC context, including, but not limited to, the design, implementation, evaluation, or use of:

Application frameworks Domain-specific languages and libraries High-level data structures High-level aspects of programming languages Directive-based optimization approaches

Acknowledgements Section

SC17 Conference Management

Conference Chair

Bernd Mohr, Juelich Supercomputing Center

Vice Chair

William D. Gropp, University of Illinois

Deputy Chair

Ralph A. McEldowney, US Department of Defense HPC Modernization Program

Executive Director

Philip C. Roth, Oak Ridge National Laboratory

Special Assistant to the Chair

Patty Gropp, Independent

Executive

Conference Chair

Bernd Mohr, Juelich Supercomputing Center

Vice Chair

William D. Gropp, University of Illinois

Deputy Chair

Ralph A. McEldowney, US Department of Defense HPC Modernization Program

Executive Director

Philip C. Roth, Oak Ridge National Laboratory

Special Assistant to the Chair

Patty Gropp, Independent

Inclusivity Chair

Toni Collis, University of Edinburgh

Technical Program Chair

Padma Raghavan, Vanderbilt University

Students @ SC Chair

Alan Sussman, University of Maryland

Infrastructure Chair

Tim Yeager, SC Committee

SCinet Chair

Jeffrey R. Schwab, Purdue University

Finance Chair

Michela Taufer, University of Delaware

Local Arrangements Chair

Cherri M. Pancake, Oregon State University

Exhibits Chair

Bronis R. de Supinski, Lawrence Livermore National Laboratory

Communications Chair

Lauren Rotman, Lawrence Berkeley National Laboratory

ACM

Ashley Cozzi, Association for Computing Machinery

IEEE-CS

June Leach-Barnaby, IEEE Computer Society

Inclusivity

Inclusivity Chair

Toni Collis, University of Edinburgh

Inclusivity Vice Chair

Trish Damkroger, Intel Corporation

Inclusivity Member

Christy Adkinson, Cray Inc

Neil Chue Hong, University of Edinburgh

Kelly Gaither, University of Texas

Christine E. Harvey, MITRE Corporation

Mary Hester, SURFnet

Jackie Kern, University of Illinois

Dustin Leverman, Oak Ridge National Laboratory

Bruce Loftis, Independent

Kimberly A. McMahon, McMahon Consulting

LeAnne Reed, McMahon Consulting

Lorna Rivera, Georgia Institute of Technology

Jamie Van Randwyk, Lawrence Livermore National Laboratory

Verónica Vergara L., Oak Ridge National Laboratory
John West, University of Texas

Keynote

Keynote Chair

Rick Stevens, Argonne National Laboratory

Keynote Member

Mike Bernhardt, Oak Ridge National Laboratory

Sponsoring Societies

ACM

Ashley Cozzi, Association for Computing Machinery

IEEE-CS

June Leach-Barnaby, IEEE Computer Society

Communications

Communications Chair

Lauren Rotman, Lawrence Berkeley National Laboratory

Communications Vice Chair

Jon Bashor, Lawrence Berkeley National Laboratory

Communications Deputy Chair

Ian MacConnell, Battelle

Photography Contractor

Jo Ramsey, Jo Julia Photography

Graphics & Website Contractor

Carlton Bruett, Carlton Bruett Design

Media Relations Contractor

Brian Ban, Ban Communications

Web Content

Sabrina Eisenreich, Leibniz Supercomputing Centre

Social Media

Sabrina Eisenreich, Leibniz Supercomputing Centre

Christine E. Harvey, MITRE Corporation

SCinet Communications Liaison

Amber Rasche, North Dakota State University

Wayfinding

Christy Adkinson, Cray Inc

Press & Media

Jill Dunbar, NASA Ames Research Center

Jeremy Tristram Thomas, Lawrence Livermore National Laboratory

Jamie Van Randwyk, Lawrence Livermore National Laboratory

Newsletter

Kathy Kincade, Lawrence Berkeley National Laboratory

Program Agenda

Christy Adkinson, Cray Inc

Bruce Loftis, Independent

Writer

Jon Bashor, Lawrence Berkeley National Laboratory

Plenary Production Contractor

Curt Apanovich, Independent

Ryan Raitt, Rezin8

Jayne Samuels, Rezin8

Exhibits

Exhibits Chair

Bronis R. de Supinski, Lawrence Livermore National Laboratory

Exhibits Deputy Chair

Christy Adkinson, Cray Inc

Exhibitor Forum

Jeanine Cook, Sandia National Laboratories

Exhibitor Forum Vice Chair

Courtenay T. Vaughan, Sandia National Laboratories

Exhibits Management Contractor

Paul Graller, Hall-Erickson Inc

Mike Weil, Hall-Erickson Inc

Event Services Contractor

Stephen E. Hagstette Jr, Freeman Company
Regina Martin, Freeman Company
Darryl Monahan, Freeman Company

HPC Impact Showcase Chair

Dimitrios Nikolopoulos, Queen's University Belfast

HPC Impact Showcase Member

Yutaka Ishikawa, RIKEN

Mark Parsons, University of Edinburgh

Wilfred Pinfold, Concurrent Systems LLC

Infrastructure

Infrastructure Chair

Tim Yeager, SC Committee

Infrastructure Vice Chair

Trey Breckenridge, Mississippi State University

AV/PC Contractor

Danielle Martinez, AV Concepts

Thomas Nichols, AV Concepts

Brian C. Walker, AV Concepts

Committee Office

Janie Bushn, University of Texas

Carolyn Peters, Argonne National Laboratory

Tracy Rafferty, University of Tennessee

Infrastructure Contractor

Leah Glick, Linklings LLC

Luke Montague, Linklings LLC

Mark Montague, Linklings LLC

Infrastructure Deputy Chair

Dustin Leverman, Oak Ridge National Laboratory

Electrical

Gary New, National Center for Atmospheric Research

Signage

Eric Sills, North Carolina State University

AV/PC

Christine E. Cuicchi, SC Committee

Security

Security Chair

Clea Marples, Lawrence Livermore National Laboratory

Security Deputy Chair

Dustin Leverman, Oak Ridge National Laboratory

Security Member

Lori Diachin, Lawrence Livermore National Laboratory

Jeff Graham, Air Force Research Laboratory

Jackie Kern, University of Illinois

Kate P. Mace, Lawrence Berkeley National Laboratory

Jamie Van Randwyk, Lawrence Livermore National Laboratory

Security Contractor

Peter Alexan, RA Consulting

Yuri Banuelos, RA Consulting

John Gaiosa, RA Consulting

Bev Oren, RA Consulting

Sam Ridley, RA Consulting

Security SCinet Contractor

Guy Julian, RA Consulting

Aaron Parker, RA Consulting

Space

Space Chair

Jay Cliburn, Parsons

Space Deputy Chair

Verónica Vergara L., Oak Ridge National Laboratory

Financial Management

Finance Chair

Michela Taufer, University of Delaware

Finance Deputy Chair

Trey Breckenridge, Mississippi State University

Students@SC Finance Liaison

Barbara Horner-Miller, BHM Consulting LLC

Finance Contractor

Danielle Bertrand, PlanB Assurance Services LLP

Rob Moody, PlanB Assurance Services LLP

Brad Rafish, PlanB Consultancy Inc

Registration/Store/Merchandise Chair

Hai Ah Nam, Los Alamos National Laboratory

Registration/Store/Merchandise Deputy Chair

Michele Bianchini-Gunn, Lawrence Livermore National Laboratory

Registration/Store/Merchandise Contractor

Katie Adair, J. Spargo and Associates

Linda Sheeran Kay, Linda Sheeran Promotions, Geiger

Karen Shipe, J. Spargo and Associates

Local Arrangements

Local Arrangements Chair

Cherri M. Pancake, Oregon State University

Special Events/Busing

Becky Verastegui, Oak Ridge National Laboratory

Catering

Janet M. McCord, University of Texas

Local Arrangements Contractor

Paul Henning, Conference and Logistics Consultants

Lois Sarfo-Mensah, Conference and Logistics Consultants

Local Arrangements Deputy Chair

Becky Verastegui, Oak Ridge National Laboratory

Housing

Housing Chair

Ewa Deelman, University of Southern California

Housing Deputy Chair

Eric Sills, North Carolina State University

Housing Contractor

Anila Singh, onPeak

SCinet

SCinet Chair

Jeffrey R. Schwab, Purdue University

SCinet Vice Chair

Linda Winkler, Argonne National Laboratory

SCinet Deputy Chair

Jason Zurawski, Energy Sciences Network

SCinet Executive Director

Conan Moore, University of Colorado, Boulder

Architecture

Architecture Team Co-Chair

Marc Lyonnais, Ciena Corporation

Jim Stewart, Utah Education Network

Architecture Team Member

Jim Hao Chen, Northwestern University

Jane Combs, University of Cincinnati

Gonzalo P. Rodrigo, Lawrence Berkeley National Laboratory

Jerry Sobieski, NorduNet

Rod Wilson, Ciena Corporation

Communications

Communications Team Co-Chair

Amber Rasche, North Dakota State University

Communications Team Member

Paul Bethge, University of Heidelberg

Rich Finlinson, Utah Education Network

Maria Kalyvaki, South Dakota State University

DevOps

DevOps Team Co-Chair

Benjamin Grover, Lawrence Livermore National Laboratory

Ken Miller, Pennsylvania State University

DevOps Team Member

Omar Aaziz, New Mexico State University

Peter Boers, SURFnetBV
Lucas Brasilino, Indiana University
Andrew Burks, University of Illinois, Chicago
Scott Chevalier, Indiana University
David Ediger, Georgia Institute of Technology
Mark Feit, Internet2
Trevor Goodyear, Georgia Institute of Technology
Julian Hammer, University of Erlangen-Nuremberg
Ezra Kissel, Indiana University
Kristine Lee, University of Illinois, Chicago
Jon Miller, Franklin and Marshall College
Jeremy Musser, Indiana University
George Robb, University of Missouri
Dori Sajdak, University at Buffalo
Greg Veldman, Purdue University

DevOps Vendor

Paul Anderson, Puppet
Neil H. McKee, InMon Corporation

Edge Network

Edge Network Team Co-Chair

Ryan Braithwaite, Los Alamos National Laboratory
Kevin Hayden, Argonne National Laboratory
Benny Sparks, Oak Ridge National Laboratory

Edge Network Team Member

Matt Adler, National Center for Atmospheric Research
Angie Asmus, Colorado State University
Christine Baissac, Easy English 4 All
Jason Charcalla, University of Tennessee
Paul Dial, National Center for Atmospheric Research
Parks Fields, Los Alamos National Laboratory
Peter Finnoff, Independent
Dave Modl, Los Alamos National Laboratory
Megan Sorensen, Idaho State University
Michael V. Sparks, CodeTank Labs

Edge Network Vendor

Matt Chrosniak, Cisco Systems
Rey Crespo, Smart City Networks
Zach Duncan, Leverage Information Systems
Chris Pendleton, Smart City Networks
Steven Phillips, Cisco Systems

Fiber

Fiber Team Co-Chair

Rebecca Hutchinson, SC Committee

Annette E. Kitajima, Sandia National Laboratories

Fiber Team Member

Armando Cisneros, National Center for Atmospheric Research

Anja Gerbes, Goethe University Frankfurt

Nathan Hanford, University of California, Davis

Felix Kaiser, University of Heidelberg

Julia Locke, Los Alamos National Laboratory

Jerome Martinez, National Center for Atmospheric Research

Miles McGee, US Air Force Research Laboratory

Brad Miller, Utah Education Network

Martin Wenzel, University of Heidelberg

Felix Zahn, University of Heidelberg

Fiber Vendor

Dan Schmedtje, Viavi Solutions

Regina Zientek, Viavi Solutions

Help Desk

Help Desk Team Co-Chair

Hervey Allen, Network Startup Resource Center, University of Oregon

Virginia Bedford, US Army Engineer Research and Development Center

Karmen Goddard, Independent

Help Desk Team Member

Xinchen Guo, Purdue University

Kelly L. Rowland, University of California, Berkeley

INDIS Workshop

INDIS Workshop Team Chair

Cees de Laat, University of Amsterdam

Paola Grosso, University of Amsterdam

INDIS Workshop Team Member

Mary Hester, SURFnet

Prof. Malathi Veeraraghavan, University of Virginia

Interconnect

Interconnect Team Co-Chair

Lance V. Hutchinson, Sandia National Laboratories

Scott Richmond, Energy Sciences Network

Logistics/Equipment

Logistics/Equipment Team Co-Chair

Ken Brice, US Army Research Laboratory

Frank Indiviglio, National Oceanic and Atmospheric Administration

Logistics/Equipment Team Member

Austin Jones, US Air Force Research Laboratory

Nathaniel Mendoza, Texas Advanced Computing Center, University of Texas

Ronda Walsh Schwab, Independent

Network Research Exhibition

Network Research Exhibition Team Co-Chair

Corby Schmitz, Argonne National Laboratory

Caroline Carver Weilhamer, Indiana University

Network Research Exhibition Vendor

Yatish Kumar, Corsa Technology

Carolyn Raab, Corsa Technology

David Whittaker, Corsa Technology

Network Security

Network Security Team Co-Chair

Jeff W. Boote, Sandia National Laboratories

Alan Commike, Reservoir Labs Inc

Network Security Team Member

Timothy W. Dunaway, US Army Engineer Research and Development Center

David Fritz, Sandia National Laboratories

Clark Gaylord, Virginia Tech

Greg Goddard, RedSeal Inc

Tanner B. Juedeman, Sandia National Laboratories

Thomas Kroeger, Sandia National Laboratories

Hans Kuhn, Network Startup Resource Center, University of Oregon; University of Oregon

Karl Newell, Internet2

Daniela Novac, University of Erlangen-Nuremberg

Sagar Shankar Singam, Regis University

Patrick Storm, University of Texas

Soledad Toledano, Lawrence Berkeley National Laboratory
Vince Urias, Sandia National Laboratories
Brian Wright, Sandia National Laboratories

Network Security Vendor

Neal Allen, Gigamon
Carlo Beronio, Attivo Networks
Kevin Bivens, Oak Ridge National Laboratory
steve brant, Splunk Corporation
Traci Dawson, Splunk Corporation
Simon Drabble, BluVector Inc
Thomas S. Henretty, Reservoir Labs Inc
Jordi Ros-Giralt, Reservoir Labs Inc
Joseph Salazar, Attivo Networks
Jeff Slabaugh, BluVector Inc
Noam Syrkin, Gigamon
Peter Varga, BluVector Inc

Power

Power Team Co-Chair

Steve Bogol, University of Notre Dame
Jay Harris, Clemson University
Jim Pepin, Clemson University

Routing

Routing Team Co-Chair

Michael Smitasin, Lawrence Berkeley National Laboratory
JP Velders, University of Amsterdam

Routing Team Member

Renuka Arya, University of Chicago
Aaron Botello, University of Colorado, Boulder
Mark Brochu, Internet2
Pieter de Boer, SURFnet
Corey Eichelberger, National Center for Supercomputing Applications, University of Illinois
Debbie Fligor, University of Illinois
Ryan Harden, University of Chicago
Zach Harlan, Energy Sciences Network
Thomas Hutton, University of California, San Diego
Neilda J Franklin, University of Houston
Tania Jareen, Purdue University
Amy Liebowitz, University of Michigan
Nathan Miller, Indiana University

Jessica Shaffer, Georgia Institute of Technology
Alan Verlo, Independent
Fei Yeh, Northwestern University

Routing Vendor

William Brockelsby, Juniper Networks
Dan DeBacker, Brocade Communications Systems
Erik DeHaas, Juniper Networks
Chris Hocker, Cisco Systems
JJ Jamison, Juniper Networks
Mike McDonough, Arista Networks
Ravi Pandey, Ciena Corporation
Dan Rivera, Juniper Networks
Wilbur Smith, Brocade Communications Systems
Robert Tellez, Ixia
Randy Wiggins, Juniper Networks
Michal Wojtczak, Cisco Systems

Volunteer Services

Volunteer Services Team Chair

Sarah M. Neuwirth, University of Heidelberg
Laura Pettit, Indiana University

WAN Transport

WAN Transport Team Co-Chair

Scott Baily, Colorado State University
Matt Smith, National Oceanic and Atmospheric Administration
Matthew J. Zekauskas, Internet2

WAN Transport Team Member

Hans Addleman, Indiana University
Will Black, CENIC
Michael Blodget, University of Wisconsin
Chad Burnham, University of Denver
Scot Colburn, National Center for Atmospheric Research
Ryan Duitman, University of Arizona
Brian A. Eschen, Morgridge Institute for Research, University of Washington
Parinaz Gharagozlou, CENIC
john hess, CENIC
Bill Jensen, University of Wisconsin
Akbar Kara, Lonestar Education and Research Network
Jehandad Khan, Virginia Tech
Scott Kohlert, Ciena Corporation

E. Paul Love, Internet Consulting of Vermont
Kali McLennan, University of Oklahoma
Brenna Meade, University of Denver
DANIEL PARENTEAU, University of Wisconsin
Kevin Quire, Utah Education Network
Kate Robinson, Western State Colorado University
Carlos Rojas-Torres, National Center for Atmospheric Research
Chris Tracy, Energy Sciences Network
Gi Vania, University of Texas, Dallas

WAN Transport Vendor

Thomas Browne, ADVA Optical Networking
Chad Dennis, Infinera Corporation
Kay Henze, Zayo Group
Chris Hunt, Nokia Corporation
Jason King, Cisco Systems
Camille Magno, JDS Uniphase Corporation
Todd McWhirter, Zayo Group
Eric Petrosky, Infinera Corporation
John Reynolds, ADVA Optical Networking
Eric Rhoades, Zayo Group

WINS

WINS Chair

Kate P. Mace, Lawrence Berkeley National Laboratory

Students @ SC

Students @ SC Chair

Alan Sussman, University of Maryland

Students @ SC Vice Chair

Scott Michael, Indiana University

Students @ SC Deputy Chair

Dorian C. Arnold, Emory University

Student Job Fair

Douglas Fuller, Red Hat Inc

Student Outreach

Diana Dugas, New Mexico State University
Christine E. Harvey, MITRE Corporation
Bruce Loftis, Independent

Mentor-Protégé Program

Christine Sweeney, Los Alamos National Laboratory

Student Volunteers

Student Volunteers Chair

Sally Ellingson, University of Kentucky

Student Volunteers Vice Chair

Christine E. Harvey, MITRE Corporation

Student Volunteer Lead

Angel Beltre, Binghamton University

Kristen C. Brown, Colorado State University

William Fox, Georgia Institute of Technology

Ami Gaspar, University of Maine System, Advanced Computing Group

Stephen Herbein, University of Delaware, Lawrence Livermore National Laboratory

Alexander Matz, University of Heidelberg

Thananon Patinyasakdikul, University of Tennessee

Melissa Romanus, Rutgers University

Robert Searles, University of Delaware

Maya Smith, Winston-Salem State University

Student Volunteers Member

Jon Bashor, Lawrence Berkeley National Laboratory

Elizabeth Bautista, Lawrence Berkeley National Laboratory

Tony Baylis, Lawrence Livermore National Laboratory

Jeanine Cook, Sandia National Laboratories

Sally Ellingson, University of Kentucky

Erika Fuentes Parsons, University of Washington, Bothell

Douglas Fuller, Red Hat Inc

Vikram Gazula, University of Kentucky

Jennifer Green, Los Alamos National Laboratory

Stephen Lien Harrell, Purdue University

Christine E. Harvey, MITRE Corporation

Will Killian, Millersville University of Pennsylvania, Lawrence Livermore National Laboratory

Benjamin Klenk, University of Heidelberg

Elizabeth Leake, STEM-Trek

Mary Leung, Sustainable Horizons Institute

Bruce Loftis, Independent

Yashema Mack, University of Tennessee

Ralph A. McEldowney, US Department of Defense HPC Modernization Program

Sarah M. Neuwirth, University of Heidelberg

Kyle Niemeyer, Oregon State University

Laura Pettit, Indiana University
Jacques Pienaar, Google
Ather Sharif, Comcast
Kathryn Traxler, Louisiana State University
Verónica Vergara L., Oak Ridge National Laboratory

Student Cluster Competition

Student Cluster Competition Chair
Stephen Lien Harrell, Purdue University

Student Cluster Competition Deputy Chair
John Cazes, University of Texas

Student Cluster Competition Member
Christopher Bross, University of Erlangen-Nuremberg
Sam Coleman, Independent
Rebecca Hartman-Baker, National Energy Research Scientific Computing Center
Andrew Howard, Cycle Computing
Si Liu, University of Texas
Yashema Mack, University of Tennessee
Ramin Nosrat, IBM
Paul Peltz Jr, Los Alamos National Laboratory
Xuanhua Shi, Huazhong University of Science and Technology
Kathleen Shoga, Lawrence Livermore National Laboratory
Doug Smith, University of Colorado, Boulder
Patrick Storm, University of Texas
Jenett Tillotson, Indiana University
Xiao Zhu, Purdue University

Student Cluster Competition Reproducibility Chair
Charles Garrett, Los Alamos National Laboratory

Student Cluster Competition Reproducibility Committee Member
Christopher Bross, University of Erlangen-Nuremberg
Carola Ellinger, Los Alamos National Laboratory
Stephen Lien Harrell, Purdue University
Michael A. Heroux, Sandia National Laboratories, St. John's University
Hai Ah Nam, Los Alamos National Laboratory
Michela Taufer, University of Delaware
Yuliana Zamora, University of Chicago

Experiencing HPC for Undergraduates

Experiencing HPC for Undergraduates Chair

Erik Saule, University of North Carolina, Charlotte

Experiencing HPC for Undergraduates Member

David P. Bunde, Knox College

Scott Michael, Indiana University

Erik Saule, University of North Carolina, Charlotte

Alan Sussman, University of Maryland

Technical Program

Technical Program Chair

Padma Raghavan, Vanderbilt University

Technical Program Vice Chair

Guillaume Aupy, French Institute for Research in Computer Science and Automation (INRIA)

Technical Program Deputy Chair

David E. Keyes, King Abdullah University of Science and Technology

Invited Speakers

Invited Speakers Chair

Esmond G. Ng, Lawrence Berkeley National Laboratory

Invited Speakers Vice Chair

Anne Trefethen, University of Oxford

Invited Speakers Member

Edmond Chow, Georgia Institute of Technology

Yutong Lu, National Supercomputer Center, Guangzhou

Robert Lucas, Livermore Software Technology Corporation

Mitsuhisa Sato, RIKEN

Frédéric Vivien, French Institute for Research in Computer Science and Automation (INRIA)

Awards

Awards Chair

Jeffrey K. Hollingsworth, University of Maryland

Awards Vice Chair

Paul Hovland, Argonne National Laboratory

Test of Time Award

Test of Time Award Chair

Elizabeth Jessup, University of Colorado, Boulder

Test of Time Award Vice Chair

David Abramson, University of Queensland

Test of Time Award Member

Martin Berzins, University of Utah

Jack Dongarra, University of Tennessee

William D. Gropp, University of Illinois

Mary Hall, University of Utah

Bruce Hendrickson, Sandia National Laboratories

Jeffrey K. Hollingsworth, University of Maryland

Dieter Kranzlmüller, Ludwig Maximilian University of Munich

Kengo Nakajima, University of Tokyo

Yves Robert, ENS Lyon, Univ. Tenn. Knoxville

Valerie Taylor, Argonne National Laboratory

Technical Papers

Technical Papers Co-Chair

Anne Benoit, ENS Lyon

Michael A. Heroux, Sandia National Laboratories, St. John's University

Algorithms

Algorithms Chair

Olivier Beaumont, French Institute for Research in Computer Science and Automation (INRIA)

Algorithms Vice Chair

Judith C. Hill, Oak Ridge National Laboratory

Algorithms Member

Emmanuel Agullo, French Institute for Research in Computer Science and Automation (INRIA)

Guillaume Aupy, French Institute for Research in Computer Science and Automation (INRIA)

Olivier Beaumont, French Institute for Research in Computer Science and Automation (INRIA)

Wesley Bland, Intel Corporation

Kate Clark, Nvidia Corporation

Karen Devine, Sandia National Laboratories

Ana Gainaru, Vanderbilt University

Abdou Guermouche, University of Bordeaux

Thomas Herault, University of Tennessee

Mark Hoemmen, Sandia National Laboratories

Mathias Jacquelin, Lawrence Berkeley National Laboratory

Kamer Kaya, Sabancı University
Christine Klymko, Lawrence Livermore National Laboratory
Sarah Knepper, Intel Corporation
X. Sherry Li, Lawrence Berkeley National Laboratory
Hatem Ltaief, King Abdullah University of Science and Technology
Loris Marchal, French National Center for Scientific Research
Anne-Cecile Orgerie, French National Center for Scientific Research
Cynthia Phillips, Sandia National Laboratories
Sanjay Rajopadhye, Colorado State University
Krzysztof Rzadca, University of Warsaw
Erik Saule, University of North Carolina, Charlotte
Miroslav Stoyanov, Oak Ridge National Laboratory
Hari Sundar, University of Utah
Harsh Vardhan, Google
Frédéric Vivien, French Institute for Research in Computer Science and Automation (INRIA)

Applications

Applications Chair

Anshu Dubey, Argonne National Laboratory, University of Chicago

Applications Vice Chair

Aydin Buluc, Lawrence Berkeley National Laboratory; University of California, Berkeley

Applications Member

Srinivas Aluru, Georgia Institute of Technology
Ariful Azad, Lawrence Berkeley National Laboratory
Michael Bader, Technical University Munich
Allison H. Baker, National Center for Atmospheric Research
Ioana Banicescu, Mississippi State University
Lorena Barba, George Washington University
Costas Bekas, IBM
John Cary, University of Colorado, Boulder
Umit Catalyurek, Georgia Institute of Technology
Aparna Chandramowlishwaran, University of California, Irvine
Ethan Coon, Oak Ridge National Laboratory
Sean M. Couch, Michigan State University
Dharshi Devendran, Pilot AI Labs
Joe Eaton, Nvidia Corporation
Sally Ellingson, University of Kentucky
Hal Finkel, Argonne National Laboratory, LLVM Foundation
Haohuan Fu, Tsinghua University; National Supercomputing Center, Wuxi (NSCC-Wuxi)
Ananth Grama, Purdue University
Boyce Griffith, University of North Carolina
George Karypis, University of Minnesota

Bronson Messer, Oak Ridge National Laboratory, University of Tennessee
Suzanne Shontz, University of Kansas
Stefan Wild, Argonne National Laboratory
Carol Woodward, Lawrence Livermore National Laboratory
Rio Yokota, Tokyo Institute of Technology
Albert-Jan Yzelman, Huawei

Architectures & Networks

Architectures & Networks Chair

Keren Bergman, Columbia University

Architectures & Networks Vice Chair

John Shalf, Lawrence Berkeley National Laboratory

Architectures & Networks Member

Jung Ho Ahn, Seoul National University
Christopher Batten, Cornell University
Jonathan Beard, ARM Ltd
Mattan Erez, University of Texas
Paolo Faraboschi, Hewlett Packard
Holger Froening, University of Heidelberg
Maya Gokhale, Lawrence Livermore National Laboratory
Bruce Jacob, University of Maryland
Steve Keckler, Nvidia Corporation
Samira Khan, University of Virginia
Michel A. Kinsy, Boston University
Moray McLaren, Google
Avi Mendelson, Technion - Israel Institute of Technology
George Michelogiannakis, Lawrence Berkeley National Laboratory
Scott Pakin, Los Alamos National Laboratory
Sébastien Rumley, Columbia University
Steve Scott, Google
Keith Underwood, Intel Corporation
Jeremiah J. Wilke, Sandia National Laboratories
Sudhakar Yalamanchili, Georgia Institute of Technology

Clouds & Distributed Computing

Clouds & Distributed Computing Chair

Kate Keahey, Argonne National Laboratory

Clouds & Distributed Computing Vice Chair

Rosa M. Badia, Barcelona Supercomputing Center

Clouds & Distributed Computing Member

Ilkay Altintas, San Diego Supercomputer Center

Amy Apon, Clemson University

Blair Bethwaite, Monash University

Kyle Chard, Argonne National Laboratory

Ewa Deelman, University of Southern California

Dick H.J. Epema, Delft University of Technology

Weicheng Huang, National Center for High-Performance Computing, Taiwan

Alexandru Iosup, Delft University of Technology

Shantenu Jha, Rutgers University

Kerstin Kleese van Dam, Brookhaven National Laboratory

Satoshi Matsuoka, Tokyo Institute of Technology, National Institute of Advanced Industrial Science and Technology

Christine Morin, French Institute for Research in Computer Science and Automation (INRIA)

Bogdan Nicolae, Huawei

Jelena Pjesivac-Grbovic, Google

Ioan Raicu, Illinois Institute of Technology

Paul Ruth, Renaissance Computing Institute

Satoshi Sekiguchi, National Institute of Advanced Industrial Science and Technology

Rich Wolski, University of California, Santa Barbara

Data Analytics, Visualization & Storage

Data Analytics, Visualization & Storage Chair

Kathryn Mohror, Lawrence Livermore National Laboratory

Data Analytics, Visualization & Storage Vice Chair

Toni Cortes, Barcelona Supercomputing Center

Data Analytics, Visualization & Storage Member

Gabriel Antoniu, French Institute for Research in Computer Science and Automation (INRIA)

Janine C. Bennett, Sandia National Laboratories

John Bent, Seagate Government Solutions

Rita Borgo, King's College London

André Brinkmann, Johannes Gutenberg University Mainz

Suren Byna, Lawrence Berkeley National Laboratory

Yong Chen, Texas Tech University

Bronis R. de Supinski, Lawrence Livermore National Laboratory

Kelly Gaither, University of Texas

Xubin He, Temple University

Dean Hildebrand, Google

Shadi Ibrahim, French Institute for Research in Computer Science and Automation (INRIA)

Scott Klasky, Oak Ridge National Laboratory

Michael Kluge, Technical University Dresden

Adrien Lèbre, French Institute for Research in Computer Science and Automation (INRIA)

Suzanne McIntosh, New York University
Tom Peterka, Argonne National Laboratory
María S. Pérez, Technical University of Madrid
Morris Riedel, Juelich Supercomputing Center
Toyotaro Suzumura, IBM
Osamu Tatebe, University of Tsukuba
Weikuan Yu, Florida State University
Zheng Zhang, Rutgers University

Performance

Performance Chair

Sameer Shende, University of Oregon

Performance Vice Chair

Lauren L. Smith, National Security Agency

Performance Member

Jean-Baptiste Besnard, ParaTools
fabien delalondre, Swiss Federal Institute of Technology in Lausanne
James Dinan, Intel Corporation
Simon Hammond, Sandia National Laboratories
Ulf Hanebutte, HPC Community Member
Katherine Isaacs, University of Arizona
Tanzima Z. Islam, Lawrence Livermore National Laboratory
Darren Kerbyson, Pacific Northwest National Laboratory
Dounia Khaldi, Intel Corporation
Matthew Knepley, Rice University
Andreas Knuepfer, Technical University Dresden
John Linford, ParaTools
Justin Luitjens, Nvidia Corporation
Allen Malony, University of Oregon
Tiffany M. Mintz, Oak Ridge National Laboratory
Mark O'Connor, ARM Ltd
Dhabaleswar K. (DK) Panda, Ohio State University
Nicholas S. Park, US Department of Defense
Olga Pearce, Lawrence Livermore National Laboratory
Marc Perache, French Alternative Energies and Atomic Energy Commission
Swaroop Pophale, Oak Ridge National Laboratory
Suzanne Rivoire, Sonoma State University
Arun Rodrigues, Sandia National Laboratories
Elliot Ronaghan, Cray Inc
Michele Weiland, University of Edinburgh
Noel Wheeler, University of Maryland
Felix Wolf, Technical University Darmstadt

Programming Systems

Programming Systems Chair

Boyana Norris, University of Oregon

Programming Systems Vice Chair

Qing Yi, University of Colorado, Colorado Springs

Programming Systems Member

George Almasi, IBM

Scott B. Baden, Lawrence Berkeley National Laboratory

Pavan Balaji, Argonne National Laboratory

Brad Chamberlain, Cray Inc

Albert Cohen, French Institute for Research in Computer Science and Automation (INRIA)

Huimin Cui, Chinese Academy of Sciences

Evelyn Duesterwald, IBM

Michael Garland, Nvidia Corporation

Jeff R. Hammond, Intel Corporation

Milind Kulkarni, Purdue University

Erwin Laure, KTH Royal Institute of Technology

Jaejin Lee, Seoul National University

Maryam Mehri Dehnavi, Rutgers University

John Mellor-Crummey, Rice University

Jiayuan Meng, Alphabet Inc.

Judy Qiu, Indiana University

Cindy Rubio-Gonzalez, University of California, Davis

Manu Shantharam, San Diego Supercomputer Center

Xipeng Shen, North Carolina State University

Shuaiwen Leon Song, Pacific Northwest National Laboratory

Didem Unat, Koc University

Richard Vuduc, Georgia Institute of Technology

Tomofumi Yuki, French Institute for Research in Computer Science and Automation (INRIA)

State of the Practice

State of the Practice Chair

Daniel S. Katz, University of Illinois; National Center for Supercomputing Applications,
University of Illinois

State of the Practice Vice Chair

Jaydeep Bardhan, Northeastern University

State of the Practice Member

Stan Ahalt, University of North Carolina

Neil Chue Hong, University of Edinburgh
Susan Coghlan, Argonne National Laboratory
Wu Feng, Virginia Tech
Sandra Gesing, University of Notre Dame
Jeff R. Hammond, Intel Corporation
Patricia Kovatch, Icahn School of Medicine at Mount Sinai
James Lin, Shanghai Jiao Tong University
Amanda Randles, Duke University
Alan Real, Durham University
Sarat Sreepathi, Oak Ridge National Laboratory
Amy Walton, National Science Foundation

System Software

System Software Chair

Sadaf R. Alam, Swiss National Supercomputing Centre

System Software Vice Chair

Larry Kaplan, Cray Inc

System Software Member

Ron Brightwell, Sandia National Laboratories
Richard Shane Canon, Lawrence Berkeley National Laboratory
Zizhong Chen, University of California, Riverside
Florina M. Ciorba, University of Basel
Dilma Da Silva, Texas A&M University
Ada Gavrilovska, Georgia Institute of Technology
Elsa J. Gonsiorowski, Lawrence Livermore National Laboratory
Bilel Hadri, King Abdullah University of Science and Technology
Atsushi Hori, RIKEN
Zbigniew Kalbarczyk, University of Illinois
Zhiling Lan, Illinois Institute of Technology
John Lange, University of Pittsburgh
Arthur Maccabe, Oak Ridge National Laboratory
Raymond Namyst, University of Bordeaux
Antonio J. Peña, Barcelona Supercomputing Center, Polytechnic University of Catalonia
Kitrick Sheets, Cray Inc
Jerome Soumagne, HDF Group
Jesper Larsson Träff, Vienna University of Technology
Eric Van Hensbergen, ARM Ltd
Peng Wu, Huawei

Tutorials

Tutorials Chair

Hank Childs, University of Oregon

Tutorials Vice Chair

Janine C. Bennett, Sandia National Laboratories

Tutorials Member

Ankit Agrawal, Northwestern University

Katie Antypas, National Energy Research Scientific Computing Center

Ray S. Chen, University of Maryland

Lydia Duncan, Cray Inc

Evan Felix, Pacific Northwest National Laboratory

Fernanda Foertter, Oak Ridge National Laboratory

Simon Hammond, Sandia National Laboratories

Kevin Huck, University of Oregon

Takeshi Iwashita, Hokkaido University

Heike Jagode, University of Tennessee

Michael Kuhn, University of Hamburg

Matthew Larsen, Lawrence Livermore National Laboratory

Stephen Leak, National Energy Research Scientific Computing Center

David Lecomber, Allinea

Samuel Li, National Center for Atmospheric Research, University of Oregon

Graham Lopez, Oak Ridge National Laboratory

Maxime Martinasso, Swiss National Supercomputing Centre

Marta Mattoso, Federal University of Rio de Janeiro

Andrey Ovsyannikov, Lawrence Berkeley National Laboratory

Hiroyuki Takizawa, Tohoku University

Douglas Thain, University of Notre Dame

Julien Tierny, French National Center for Scientific Research

Sean Treichler, Nvidia Corporation

Ana Lucia Varbanescu, University of Amsterdam

Weichung Wang, National Taiwan University

Brian J. N. Wylie, Juelich Supercomputing Center

Rio Yokota, Tokyo Institute of Technology

Panels

Panels Chair

Yves Robert, ENS Lyon, Univ. Tenn. Knoxville

Panels Vice Chair

Lois Curfman McInnes, Argonne National Laboratory

Panels Member

Gabrielle Allen, University of Illinois
Rosa M. Badia, Barcelona Supercomputing Center
Costas Bekas, IBM
David L. Brown, Lawrence Berkeley National Laboratory
Ewa Deelman, University of Southern California
Tamara Kolda, Sandia National Laboratories
Simon McIntosh-Smith, University of Bristol
Chris J. Newburn, Nvidia Corporation
Karl Rupp, Freelance Computational Scientist
Michela Taufer, University of Delaware
Bora UCAR, French National Center for Scientific Research
Ana Lucia Varbanescu, University of Amsterdam

Workshops

Workshops Chair

Almadena Chtchelkanova, National Science Foundation

Workshops Vice Chair

Luiz DeRose, Cray Inc

Workshops Member

Katie Antypas, National Energy Research Scientific Computing Center
Amy Apon, Clemson University
Pavan Balaji, Argonne National Laboratory
Ron Brightwell, Sandia National Laboratories
Franck Cappello, Argonne National Laboratory
Almadena Chtchelkanova, National Science Foundation
Giri Chukkapalli, Cavium Inc
Bronis R. de Supinski, Lawrence Livermore National Laboratory
Chen Ding, University of Rochester
Rudolf Eigenmann, National Science Foundation
Chris Gottbrath, Nvidia Corporation
Howie Huang, George Washington University
Kirk E. Jordan, IBM, Hartree Centre
Karen L. Karavanic, Portland State University
Darren Kerbyson, Pacific Northwest National Laboratory
David Lecomber, Allinea
Xiaolin Andy Li, University of Florida
Osni Marques, Lawrence Berkeley National Laboratory
Maryam Mehri Dehnavi, Rutgers University
José Moreira, IBM
Frank Mueller, North Carolina State University
Scott Pakin, Los Alamos National Laboratory

Abani Patra, University at Buffalo
Heidi Poxon, Cray Inc
Lawrence Rauchwerger, Texas A&M University
Subhash Saini, NASA Ames Research Center
Jennifer M. Schopf, Indiana University
Seetharami Seelam, IBM
Melissa C. Smith, Clemson University
Qing Yi, University of Colorado, Colorado Springs

Posters

Posters Chair

Kengo Nakajima, University of Tokyo

Posters Vice Chair

Karl W. Schulz, Intel Corporation
Osamu Tatebe, University of Tsukuba

Posters Member

Kento Aida, National Institute of Informatics, Japan
Damian Alvarez, Juelich Supercomputing Center
Achim Basermann, German Aerospace Center
James Beyer, Nvidia Corporation
Luc Bougé, Research Institute of Computer Science and Random Systems
Ron Brightwell, Sandia National Laboratories
Xing Cai, Simula Research Laboratory
Christophe Calvin, French Alternative Energies and Atomic Energy Commission
Toni Cortes, Barcelona Supercomputing Center
Alexandru Costan, French Institute for Research in Computer Science and Automation (INRIA)
Yifeng Cui, San Diego Supercomputer Center
Doug Doerfler, Lawrence Berkeley National Laboratory
Sameh Elnikety, Microsoft
Wu Feng, Virginia Tech
Niall Gaffney, Texas Advanced Computing Center, University of Texas
Todd Gamblin, Lawrence Livermore National Laboratory
Victor Gergel, Nizhny Novgorod State University
Alexander Heinecke, Intel Corporation
Weicheng Huang, National Center for High-Performance Computing, Taiwan
Florin Isaila, Charles III University of Madrid
Takeshi Iwashita, Hokkaido University
Hideyuki Kawashima, University of Tsukuba
Masaaki Kondo, University of Tokyo
Dieter Kranzlmüller, Ludwig Maximilian University of Munich
Tomohiro Kudoh, University of Tokyo
Jaejin Lee, Seoul National University

Qing Gary Liu, New Jersey Institute of Technology
Hatem Ltaief, King Abdullah University of Science and Technology
Carlos Maltzahn, University of California, Santa Cruz
Osni Marques, Lawrence Berkeley National Laboratory
Maxime Martinasso, Swiss National Supercomputing Centre
Simon McIntosh-Smith, University of Bristol
Diana Moise, Cray Inc
Hiromichi Nagao,
Hiroshi Nakashima, Kyoto University
Akira Naruse, Nvidia Corporation
Stephen L. Olivier, Sandia National Laboratories
Luke Olson, University of Illinois
Kenji Ono, Kyushu University
Manish Parashar, Rutgers University
Dirk Pleiter, Forschungszentrum Juelich, University of Regensburg
María S. Pérez, Technical University of Madrid
Judy Qiu, Indiana University
Ioan Raicu, Illinois Institute of Technology
Sivasankaran Rajamanickam, Sandia National Laboratories
Paul Renaud-Goud, Toulouse Institute of Research in Computer Science
Carlos Rosales, University of Texas
Kento Sato, Lawrence Livermore National Laboratory
Matthieu Schaller, Durham University
Takashi Shimokawabe, University of Tokyo
Hari Sundar, University of Utah
Christine Sweeney, Los Alamos National Laboratory
Keita Teranishi, Sandia National Laboratories
Didem Unat, Koc University
Rob F. Van der Wijngaart, Intel Corporation
Eric Van Hensbergen, ARM Ltd
Peng Wang, Nvidia Corporation
Nicholas J. Wright, National Energy Research Scientific Computing Center

Birds of a Feather (BOFs)

Birds of a Feather Chair

Jeffrey S. Vetter, Oak Ridge National Laboratory

Birds of a Feather Vice Chair

Naoya Maruyama, Lawrence Livermore National Laboratory

Birds of a Feather Member

Leonardo Bautista-Gomez, Barcelona Supercomputing Center

Taisuke Boku, University of Tsukuba

fabien delalondre, Swiss Federal Institute of Technology in Lausanne

Hal Finkel, Argonne National Laboratory, LLVM Foundation
Fernanda Foertter, Oak Ridge National Laboratory
Balazs Gerofi, RIKEN
Yutong Lu, National Supercomputer Center, Guangzhou
Naoya Maruyama, Lawrence Livermore National Laboratory
Kento Sato, Lawrence Livermore National Laboratory
Thomas R. W. Scogland, Lawrence Livermore National Laboratory
Jeffrey S. Vetter, Oak Ridge National Laboratory
Brian J. N. Wylie, Juelich Supercomputing Center

Emerging Technologies

Emerging Technologies Chair

Franck Cappello, Argonne National Laboratory

Emerging Technologies Vice Chair

Felix Wolf, Technical University Darmstadt

Emerging Technologies Member

André Brinkmann, Johannes Gutenberg University Mainz

Franck Cappello, Argonne National Laboratory

Almadena Chtchelkanova, National Science Foundation

Wu Feng, Virginia Tech

Kengo Nakajima, University of Tokyo

Hiroshi Nakashima, Kyoto University

Felix Wolf, Technical University Darmstadt

Scientific Visualization & Data Analytics Showcase

Scientific Visualization & Data Analytics Showcase Chair

Kelly Gaither, University of Texas

Scientific Visualization & Data Analytics Showcase Vice Chair

Vetria Byrd, Purdue University

Scientific Visualization & Data Analytics Showcase Member

Daniel Acevedo-Feliz, King Abdullah University of Science and Technology

Rita Borgo, King's College London

Amit Chourasia, University of California, San Diego

John Clyne, National Center for Atmospheric Research

Cyrus Harrison, Lawrence Livermore National Laboratory

Joseph A. Insley, Argonne National Laboratory, Northern Illinois University

Heike Lette, Kaiserslautern University

Kenneth Moreland, Sandia National Laboratories

Paul A. Navratil, University of Texas

Kristi Potter, National Renewable Energy Laboratory
David Semeraro, Texas Advanced Computing Center, University of Texas
Robert R. Sisneros, University of Illinois
Jonathan L. Woodring, Lawrence Berkeley National Laboratory

Technical Archive

Proceedings

Julien Langou, University of Colorado, Denver

Conference Artifacts

Bruce Loftis, Independent

Doctoral Showcase

Doctoral Showcase Chair

Sunita Chandrasekaran, University of Delaware

Doctoral Showcase Vice Chair

Andy Herdman, Atomic Weapons Establishment

Doctoral Showcase Member

Manisha Gajbe, Intel Corporation

Andy Herdman, Atomic Weapons Establishment

Shadi Ibrahim, French Institute for Research in Computer Science and Automation (INRIA)

Kamil Iskra, Argonne National Laboratory

Guido Juckeland, Helmholtz-Zentrum Dresden-Rossendorf

Alice Koniges, Lawrence Berkeley National Laboratory

John Leidel, Texas Tech University

Qing Gary Liu, New Jersey Institute of Technology

Kamesh Madduri, Pennsylvania State University

Boyana Norris, University of Oregon

Vivek Pallipuram, University of the Pacific

Manu Shantharam, San Diego Supercomputer Center

Eric Stahlberg, Frederick National Laboratory, National Cancer Institute

Christian Terboven, RWTH Aachen University

Early Career Program

Early Career Workshop Chair

Dorian C. Arnold, Emory University

Early Career Workshop Member

Gladys Andino, Purdue University

Elizabeth Bautista, Lawrence Berkeley National Laboratory

Scott Beamer, Lawrence Berkeley National Laboratory

Mary Leung, Sustainable Horizons Institute

Xu Liu, College of William and Mary

Suzanne McIntosh, New York University

Cindy Rubio-Gonzalez, University of California, Davis

Seetharami Seelam, IBM